Optical Remote Sensing of Local-Scale Thermospheric Dynamics Above Antarctica

Submitted by
Callum Anderson
B.Sc.(Space Science)(Hons)
La Trobe University (2005)

A thesis submitted in total fulfillment
of the requirements for the degree of
Doctor of Philosophy

School of Engineering and Mathematical Sciences
Faculty of Science, Technology and Engineering

La Trobe University,
Bundoora VIC, 3086
Australia

FEBRUARY 2011

Contents

List of Figures

List of Abbreviations

DIS Doppler imaging system
FPI Fabry-Perot interferometer
FPS Fabry-Perot spectrometer
IMF Interplanetary magnetic field
LST Local solar time
MLT Magnetic local time
SDI Scanning Doppler imager
UT Universal time

Geomagnetic and Solar Indices

AE*, *AU*, *AL Auroral electrojet indices. AL measures the maximum negative excursion of the horizontal component of the magnetic field from a number (> 10) of northern hemisphere auroral-latitude stations. Similarly, AU measures the maximum positive excursion. AE is then the difference AU-AL.

ap The 'equivalent amplitude' index, derived from Kp, providing a linear measure of magnetic field variations in 3-hour intervals.

Dst The **D**isturbance **s**torm **t**ime index. Represents the average value of the horizontal component of Earth's magnetic field measured hourly at selected low-latitude magnetic observatories. Units are nanotesla (nT).

$\mathbf{F_{10.7}}$ A measure of the energy flux radiated by the Sun at a wavelength of 10.7 cm (2800 MHz). Units are 10^{-22} W.m^{-2}.Hz^{-1}.

Kp The Kp index tracks the largest average range (lowest to highest value) of the horizontal magnetic field components measured by a number of sub-auroral magnetic observatories in 3-hour intervals. Kp indices increase on a quasi-logarithmic scale.

Abstract

The aim of this project was to test the hypothesis that small spatial scales (on the order of 500 km or less) play an important role in thermospheric dynamics. To achieve this, a new all-sky scanning, imaging Fabry-Perot spectrometer was built at La Trobe University and installed at Mawson station, Antarctica ($67°36'$S, $62°52'$E, Inv $70°30'$S, Magnetic Midnight: 2240 UT). This instrument was capable of recording independent spectra from 61 locations simultaneously across a 144° full-angle field-of-view, with a maximum time resolution of ∼2 minutes. Data from this instrument was combined with that from a narrow-field (6° full-angle), imaging Fabry-Perot spectrometer operating at Davis station ($68°35'$S, $77°58'$E, Inv $74°36'$S, Magnetic Midnight: 2200 UT).

A superposed epoch analysis of spatially-resolved wind and temperature showed predominantly pressure-gradient driven winds during quiet conditions, and convection driven winds during active conditions. A wind abatement in the pre-magnetic midnight sector was frequently observed under both levels of activity, while sunward flow was often observed equatorward of Mawson in the early magnetic morning sector.

Bistatic measurements between Mawson and Davis allowed for vertical wind measurements at three locations along the line joining the two stations in addition to the vertical wind measured routinely above each station. These vertical winds were at times correlated over horizontal scales of ∼ 160-480 km, while at other times there was little correlation over the smallest separations of ∼ 160 km. Vertical winds were positively correlated with the large-scale horizontal divergence, in agreement with the relation predicted by Burnside. However, at small scales (∼ 100 km), the Burnside relation was not a reliable predictor of vertical winds.

Strong wind shears were observed above Mawson in the presence of bright auroral arcs, and the heating rates due to these shears were found to make appreciable contributions to the energy balance. These results indicate that small spatial scales do indeed play an important role in thermospheric dynamics.

Statement of Authorship

Except where reference is made in the text of the thesis, this thesis contains no material published elsewhere or extracted in whole or in part from a thesis submitted for the award of any other degree or diploma. No other person's work has been used without due acknowledgement in the main text of the thesis. The thesis has not been submitted for the award of any degree or diploma in any other tertiary institution.

CALLUM ANDERSON

Acknowledgements

This thesis would not have been possible without the guidance and support of my supervisor Dr. Mark Conde. He has provided me with opportunities to travel to places that would otherwise have been inaccessible to me, and which have given me priceless experiences that I will never forget. His assistance has always been intelligent, insightful, and freely given, and his enthusiasm for his work has been a great motivation. Working with Dr. Conde has always been an enjoyable and enlightening experience, from which I have learned so much, and I could not have hoped for a more capable or considerate supervisor. For all of this I am truly grateful.

I am also greatly indebted to Professor Peter Dyson, whose patience, experience, and easygoing nature have ensured that my time spent at La Trobe has been enjoyable and relatively stress-free. He has always made himself available to answer questions or offer assistance to me and many other students, even into retirement, for which I am very thankful.

Mr. Theo' Davies has been a colleague and collaborator since I first started working with Fabry-Perot spectrometers during my honours year. I have worked with him in Antarctica as well as at La Trobe on both instrumental issues and data analysis. He has always been willing to help me out with any questions or problems I had, and his methodical approach to problem solving has more than once showed me the wisdom of patience. I have been very lucky to have such a helpful and knowledgeable colleague, amiable travelling companion, and most importantly a good friend. Thank you.

I have also had the privilege of working with Professor Mike Kosch, a collaborator in the Mawson SDI project. He has been a constant source of motivation, encouragement and new ideas, and through him I have been given opportunities to collaborate with many people from around the world, for which I am very grateful.

To the '07 winter expeditioners at Mawson station I will always be grateful for their assistance and most importantly their friendship. The year I spent at Mawson could not have been more enjoyable, and that is solely due to the comradeship offered me by these expeditioners. One could not hope to live and work in a cold, isolated part of the world with a finer group of people.

Thank you to Deanna Matthews for proof-reading and printing logistics, and to my parents also for their painstaking proof-reading and for supporting me in everything. Thank you.

Overview

The near-Earth space environment is tightly coupled to Earth's upper atmosphere at auroral latitudes through a dynamic system of currents, electric and magnetic fields, and precipitating magnetospheric particles. Magnetic substorms, which represent a fundamental mode of magnetospheric disturbance in response to the solar wind, can provide highly spatially and temporally localised energy inputs to the neutral upper atmosphere, which are capable of driving small-scale ($\sim$ 100 km or less) structures in the wind and temperature fields. The ultimate driver of space-weather phenomena is the Sun. An introduction to the Sun, solar wind and magnetosphere, and magnetospheric disturbances, is given in Chapter 1.

With continually improving ground-based instrumentation and in-situ observations from satellites the trend has been toward observing the upper atmosphere on ever smaller spatial scales and with higher time resolution. The evidence for small-scale structure in thermospheric winds and temperatures has thus been accumulating for many years. This observational trend has been accompanied by an increase in the resolution of complex atmospheric numerical models, which are already capable of reproducing much of the large-scale structure and dynamics of the upper atmosphere, and are beginning to focus on resolving structures at scales of 100 km or less. These numerical models take a first-principles approach to reproducing the observed properties of the atmosphere in terms of its composition, energetics and dynamics through parameterising the important driving forces and numerically solving the physical equations which govern atmospheric processes. The theoretical and experimental aspects of upper atmospheric chemistry and dynamics are reviewed in Chapter 2.

This project involved the construction and installation of a new all-sky scanning Doppler imager (SDI), a Fabry-Perot spectrometer capable of observing the weak, natural light emission (airglow) from the upper atmosphere at many tens of locations simultaneously across a wide, 144° full-angle field-of-view. The particular airglow wavelength used in this work was the 630.0 nm line, which is emitted by excited oxygen atoms with a peak emission intensity at an altitude of around 240 km, in a region of Earth's upper atmosphere known as the thermosphere. Analysis of the spectral line shapes recorded by the instrument allows temperature and two-dimensional horizontal wind fields to be inferred, with a spatial resolution of approximately 100 km and an average temporal resolution of 3-5 minutes. The instrument is also capable of recording, concurrently with spectral acquisition, a spectrally unmodulated, monochromatic image of the sky at the same wavelength as the recorded spectra (and over an identical field-of-view), which allows auroral structures in the sky to be related to horizontal gradients in the wind and temperature fields.

The aim of the project was to test the hypothesis that small spatial scales (on the order of 500 km or less) play an important role in thermospheric dynamics. To meet this aim, two broad sets of tasks were required. First, to design and write a software suite which would operate (and automate) the new instrument and perform the necessary post-processing of the observational data. Second, to install the instrument at Mawson station, Antarctica, operate it throughout the year, and analyse the resulting data, in conjunction with data from a Fabry-Perot spectrometer (FPS) already operating at Davis station, with a view to identifying small-scale structures and processes in the derived wind fields.

Initial software development required programs which could communicate with the various hardware components of the instrument, including a separation-scanned Fabry-Perot etalon, a thermoelectrically cooled electron-multiplying CCD camera, a filter wheel and two servo motors. Additional software was then required for instrument calibration, spectral acquisition and data storage, and for allowing the instrument to operate autonomously. Much of this initial software development was completed during 2006, and field-tested along with the instrument at La Trobe university. Further software routines were required for data post-processing, including inferring vector wind fields, superposed epoch analyses, combining bistatic wind measurements from the Mawson and Davis spectrometers, and various related tasks. The instrument was shipped to Antarctica during the 2006/2007 summer. Installation in the Aeronomy building at Mawson station was completed during the summer season, and useful operation of the instrument began in March 2007.

Chapter 3 presents a general introduction to Fabry-Perot spectrometer theory, while Chapters 4 and 5 provide detailed information on the theory and operation of the Mawson and Davis spectrometers which were used in this work. Three broad sets of results are then presented. In Chapter 6, the daily variability of thermospheric winds and temperatures are presented in four single-night case studies, which highlight the important changes which are observed to occur in winds and temperatures as geomagnetic activity increases. A climatological analysis is then presented, which demonstrates the average behaviour of winds and temperatures under ‘quiet’ and ‘active’ levels of geomagnetic activity. Spatially resolved maps of average wind, temperature and intensity, using the spatial resolution of the SDI measurements, are also presented.

In Chapter 7, results are presented from a bistatic experiment utilising the Mawson and Davis spectrometers and designed to investigate the spatial variability of thermospheric vertical winds at multiple locations between Mawson and Davis stations, in addition to vertical winds routinely measured above each station. This experiment allowed the vertical wind to be unambiguously calculated at three locations between the two stations, and, using the imaging capability of the Mawson spectrometer, to relate the derived vertical winds to the local auroral morphology. Three case studies of this type are presented.

Chapter 8 presents three sets of results examining the relationship between vertical winds and horizontal wind divergence, and an analysis of sheared wind flows associated with auroral arcs. Firstly, a recurrent upwelling above Davis station, identified in Chapter 6, is examined in light of the average horizontal divergence above that station, and then a more general investigation into the vertical wind/horizontal divergence relationship is presented using all available data from both Mawson and Davis. The third study investigates the relationship between the bistatic vertical winds measured during one of the case studies presented in Chapter 7 and the small-scale divergence of the horizontal wind as measured from Mawson station. The last study in Chapter 8 is an investigation into sheared wind fields which were often observed in the presence of bright auroral arcs, and a lower limit to the amount of power required to drive such wind fields is estimated. Finally, concluding remarks and recommendations for future work are presented.

Chapter 1

Sun-Earth Interaction

The near-Earth region of space is a dynamic and complex environment of energetic particles, magnetic and electric fields and high-speed plasma motions. Changes in this region occur on a large range of scales, both spatial and temporal, and are termed collectively 'space weather'. The ultimate driver of space weather is the Sun. Earth's atmosphere is heated by radiation from the Sun. Above $\sim$ 100 km altitude, the large-scale pressure gradients resulting from this heating drive a general circulation away from the region of heating to the cooler nightside. Radiation at these altitudes also creates a region of ionisation, called the *ionosphere*, which couples the neutral atmosphere to the geomagnetic field and allows the flow of currents which provide additional heating and momentum to the neutral upper atmosphere through collisions.

The Sun also emits a hot, fast-moving plasma known as the *solar wind*, which interacts with Earth's magnetic field and can drive phenomena such as substorms, which disturb the geomagnetic field and drive strong currents and auroral displays at high latitudes, which can in turn provide large and small-scale heat and momentum inputs to the neutral upper atmosphere. This section provides a brief introduction to the Sun and its interaction with Earth's magnetic field.

1.1 The Sun

1.1.1 Radiation

In Figure 1.1.1, the solar spectrum, as observed outside Earth's atmosphere, is shown for wavelengths less than 3 μm. The solar spectrum is similar to that of a black body at a temperature of around 5770 K, with a power output per unit area of approximately 1370 W.m^{-2} at Earth's average distance from the Sun of 1.496×10^8 km (this distance is equal to 1 astronomical unit, or 1 AU). Wavelengths less than about 200 nm are completely absorbed by the upper atmosphere above 80 km altitude (Roble, 1995), and therefore most directly affect it. These wavelength ranges, along with their energies and typical labels, are shown in Table 1.1.

There is a well-defined variation in the Sun's radiative output that occurs over an average period of 11 years and which is referred to as the solar cycle. This cycle is

Region	Wavelength Range	Energy Range
X-ray	0.1 - 10 nm	12.4 keV - 124 eV
Extreme Ultraviolet (EUV)	10 - 121 nm	124 eV - 10.2 eV
Ultraviolet (UV)	121 - 200 nm	10.2 eV - 6.2 eV

Table 1.1: Wavelength regions directly affecting the upper atmosphere.

a result of the differential rotation period of the outer layers of the Sun with respect to latitude, with the shortest periods at the equator. The magnetic polarity of the large-scale background solar magnetic field reverses after each 11 year cycle, and thus, taking into account magnetic field polarity, a longer 22 year solar cycle is also observed. Figure 1.1.2 shows how the number of sunspots observed on the Sun are correlated with two indices which track variations in UV (the MgII index) and 10.7 cm radio flux ($F_{10.7}$ index). The maxima and minima of the clearly visible solar cycle are referred to as solar maximum and minimum respectively. As of 2008, the solar cycle is at a minimum, and at the start of solar cycle 24. Sunspot numbers as well as radiation fluxes are seen to follow this 11-year periodicity, during which the large-scale structure of the solar corona changes from well-ordered during solar minimum to a much more complex configuration during solar maximum (Issautier, 2006). As well as the 11-year (or 22-year) cycle, variations exist on a timescale of 27-days (and higher harmonics of this period) caused by the Sun's rotation (as seen from Earth), during which active regions are brought into and out of alignment with Earth (e.g. Rottman, 1988). There are also significant variations on shorter time-scales, due to eruptive phenomena such as solar flares (e.g. Lilensten and Kretzschmar, 2006)

Heating in the upper atmosphere due to the absorption of solar radiation at wavelengths less than about 200 nm creates pressure gradients which drive a global-scale upper atmospheric circulation from the dayside to the nightside (Roble, 1983). Radiation is also largely responsible for the chemical composition of the atmosphere through the production

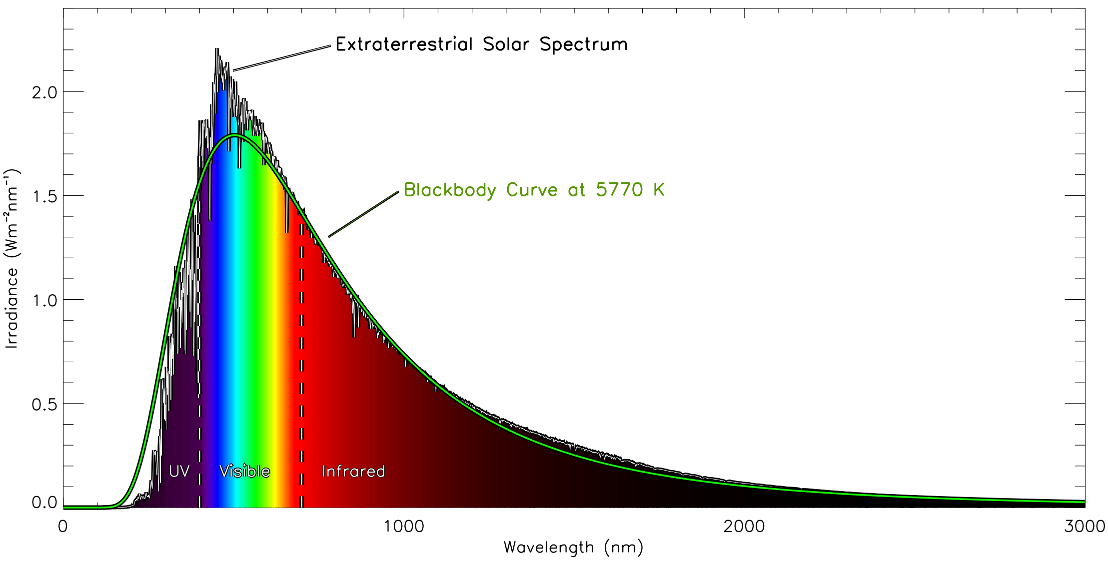


Figure 1.1.1: The solar spectrum, reproduced from Gueymard (2004).

of atomic and ionic species (Rottman, 1988). Variations in solar output can affect vertical temperature profiles, which in turn affect the altitude variation of the number densities of upper atmospheric species. Observations show that by far most of the variability in the solar spectrum is present at these short wavelengths, where the intensity can vary by up to 100% (e.g. Lilensten and Kretzschmar, 2006). This is an important fact, considering the crucial role that radiation at these wavelengths plays in the upper atmosphere.

1.1.2 The Solar Wind

The solar wind is a hot, supersonic, highly conductive plasma (consisting primarily of protons and electrons in equal numbers) which streams radially outward from the Sun (Hundhausen, 1995). This wind originates in the hot outer atmosphere of the Sun (the solar 'corona'), and is believed to be driven outwards by a combination of processes, including the pressure difference between the corona and the interstellar medium (op. cit.) and possibly magnetohydrodynamic waves (Tu and Marsch, 1997). The solar wind can be characterised by its velocity and density, both of which can vary considerably. The average wind velocity is around 400 km.s^{-1} (Gosling and Hansen, 1971), but can vary in the range of 250-800 km.s^{-1}. Likewise, the average number density at a distance of 1 AU is on the order of 5 cm^{-3}, and can vary between < 1 and 20 cm^{-3} (Rees, 1989). These characteristics are strongly dependent upon solar latitude: the solar wind emerging from the polar coronal holes, where field-lines are open, is a fast flow moving at speeds of around 800 km.s^{-1}, while the low-latitude solar wind tends to be slower (around 400 km.s^{-1}), but much more variable (Issautier, 2006).

Sudden enhancements in solar wind density and speed have been observed (Crooker et al., 2000), associated with the passage of interplanetary 'shocks', localised compressions of the solar wind which result from injections of higher-speed plasma flows into the slower background solar wind. These high-speed flows are thought to have two main sources: coronal mass ejections (CME's) and coronal holes. CME's are enormous ejections of solar wind plasma (on the order of 10^{12} kilograms), occurring more frequently around the peak of the solar cycle. These eruptions can depart the Sun with a range of speeds, and if the speed is higher than that of the ambient solar wind then a shock may form at the leading edge of the CME (Wimmer-Schweingruber, 2005).

Coronal holes are regions of low density plasma, associated with open magnetic field lines, which allow easy escape of solar wind plasma. At solar minimum, these holes are mostly confined to the Sun's polar regions, however they tend to be observed at progressively lower latitudes as the solar cycle progresses towards, and past, its maximum (Gorney, 1990). Shocks can form when high-speed streams from coronal holes collide with slower moving plasma emitted from other regions. The presence of shocks in the interplanetary medium often results in geomagnetic storms on Earth (see Section 1.2.3), which is coupled to the solar wind via the *magnetosphere* (Burns et al. (2007) and Section 1.2). Cyr et al. (2000) studied two years of SOHO (Solar and Heliospheric Observatory) satellite data and found that around 83% of geomagnetic storms (Kp ≥ 6) were associated

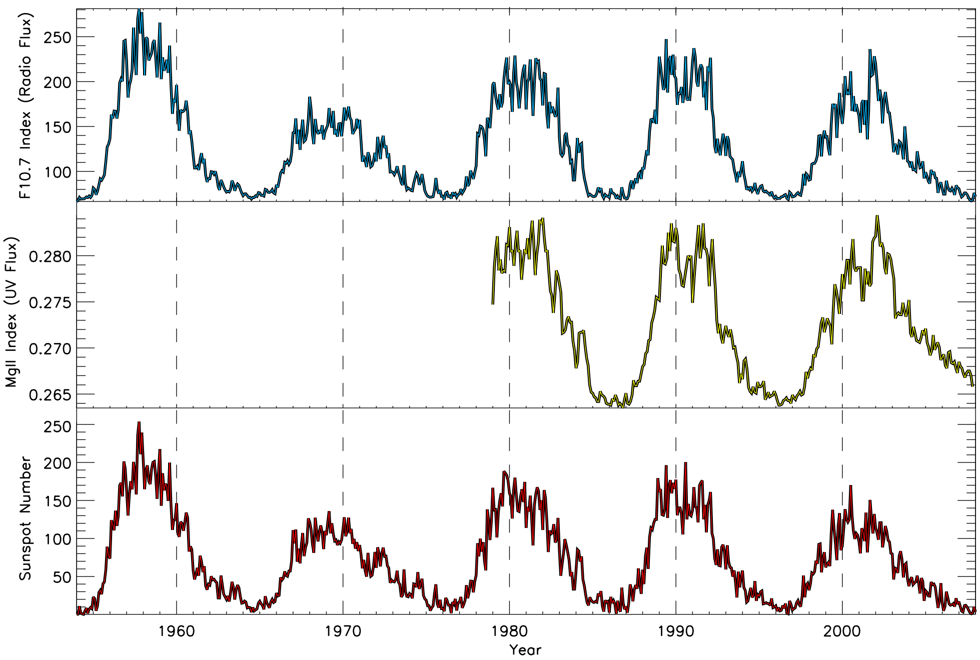


Figure 1.1.2: From top to bottom: monthly averaged 10.7 cm solar radio flux ($F_{10.7}$), MgII Index (indicating UV and EUV flux), and sunspot number. The 11-year cycle is clearly visible in all three panels.

with Earthward directed ('frontside') CME's.

The highly conductive nature of the solar wind plasma has the important consequence that magnetic field lines are effectively 'frozen-in' to the plasma, meaning that field lines embedded within a parcel of plasma are carried along with the parcel. In the strictest sense, a finite conductivity will allow field lines to diffuse through the plasma with a diffusion time-scale (τ_d) given by (Priest, 1995):

$$\tau_d \simeq \mu_0 \sigma L^2 \qquad (1.1.1)$$

where:

$$\mu_0 = 4\pi \times 10^{-7} \text{ Wb.A}^{-1}\text{.m}^{-1}$$

$$\sigma = \text{conductivity}$$

$$L = \text{characteristic scale length}$$

However in the solar wind, where the conductivity is very high, this diffusion timescale will be very long compared to the velocity of the bulk flow (Verma, 1995), and only at small scales (small L) will field-line diffusion become fast enough to be important. The net result is that magnetic field lines from the Sun are carried into interplanetary space by the solar wind; the field is then referred to as the interplanetary magnetic field (IMF). At 1 AU this field has an average magnitude in the range ~5-10 nT (Lyons, 1992), although the instantaneous field strength is highly variable. The orientation of the IMF is often described in terms of its components in the geocentric solar magnetospheric (GSM) coordinate system. In this system, the x-axis points toward the Sun from Earth, the positive z-axis is aligned along Earth's magnetic dipole axis toward the north magnetic pole, and the y-axis is perpendicular to both (positive eastward).

When charged particles encounter magnetic field lines they are caused to gyrate around them by the Lorentz force (Section 1.2.2). Time-averaged motion perpendicular to the field lines is thus severely restricted in the absence of an electric field, while motion parallel to them is unhindered. The solar wind is therefore forced to flow predominantly around Earth's magnetic field, thereby creating a cavity in the solar wind called the magnetosphere. The magnetosphere acts as the interface between the interplanetary medium (i.e. the solar wind) and Earth's upper atmosphere, coupling processes occurring in the upper atmosphere with events originating at the Sun.

1.2 The Magnetosphere

The following description of the magnetosphere is necessarily brief, and the reader is referred to Kivelson and Russell (1995) for more information on quantitative properties of the magnetosphere. The term magnetosphere refers to the cavity produced by the interaction of the solar wind with Earth's magnetic field. The presence of the fast-moving solar wind, which is unable to cross magnetic field lines owing to its very high conductivity, causes Earth's geomagnetic field to take on a distinctly non-dipole shape. Dynamic pressure, exerted on the dayside[1] geomagnetic field by the solar wind, results in a compression of closed field lines on the dayside, while 'open' field lines (those closing with the IMF) and those on the nightside are swept many Earth radii downstream. Figure 1.2.1 shows an idealised cross-section of the magnetosphere through the noon-midnight meridian. A 'bow-shock' is formed on the dayside where the solar wind encounters the geomagnetic field and is forced to flow around the magnetospheric cavity. Earthward of this bow-shock is a region called the *magnetopause*, which forms the boundary between the magnetosphere and the solar wind. The location of the magnetopause is variable, depending on solar wind pressure and the rate of field line reconnection (see below), but is typically around 8-11 Earth radii (R_E). In response to discontinuities in solar wind speed, the magnetopause location can move at speeds of up to 600 km.s^{-1} (Kallenrode, 2004, p. 294).

On the nightside, field lines are stretched many Earth radii in the antisunward direction, forming the *magnetotail*, which consists of north and south 'lobes' containing bundles of oppositely directed magnetic field lines (pointing Earthward in the north lobe and tailward in the south lobe). The plasma density in these lobes is very low, on the order of 10^{-2} cm^{-3}. At the equatorial boundary between these two lobes is a thin sheet of plasma typically around 6 R_E thick near midnight, increasing to around twice this size on the dawn and dusk flanks (Hill, 1974), and varying with distance downstream. This plasma sheet, which is around 10-100 times more dense than the tail lobes, is the location toward which charged particles in the magnetotail drift due to the combined action of the electric

[1] The terms dayside and nightside are often used to differentiate between the sunlit and non-sunlit sides of Earth. Dawn and dusk are used in a similar way to mean respectively the hemisphere which will move into daylight, and that which will move into darkness, due to Earth's rotation.

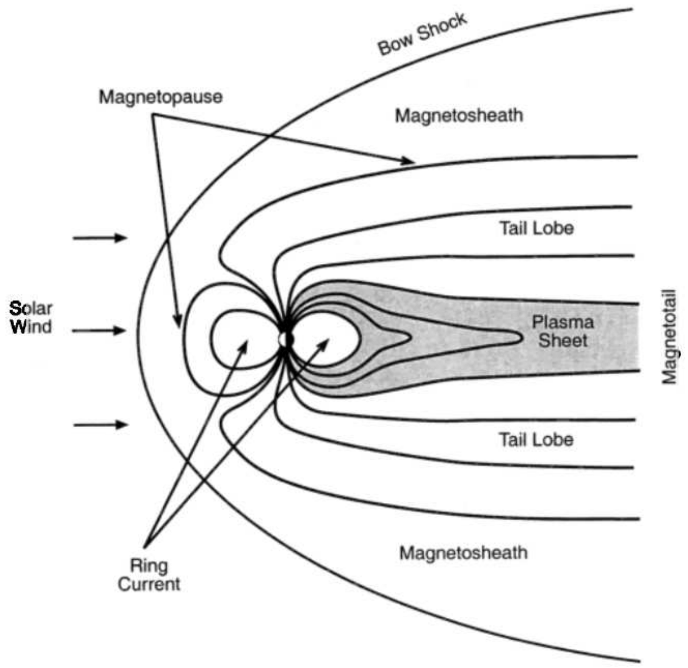


Figure 1.2.1: Magnetospheric cross-section in the noon-midnight meridian, showing the regions of interest. From Lui (2001).

and magnetic field ($\mathbf{E} \times \mathbf{B}$ forcing). A strong, dawn-dusk oriented electric current flows along this plasma sheet, and closes on the boundary of the magnetotail (see Figure 1.2.7).

The plasma sheet extends Earthward toward regions of plasma trapped by the approximately dipolar magnetic field which exists in the near-Earth region. Known as *radiation belts*, these regions are more like a torus extending longitudinally around the Earth. At high-latitudes, the convergence of the geomagnetic field lines acts as a magnetic 'mirror', causing particles reaching these latitudes to be deflected back toward the opposite hemisphere, effectively trapping them. This trapping is not perfect, however, as collisions between particles can scatter them into the 'loss-cone', the small range of angles (relative to the local magnetic field direction) at which the charged particles will not be deflected by the mirror force before they penetrate to low enough altitudes where collisions with neutral atmospheric particles are inevitable. Thus these particles can escape their magnetic confinement and enter the upper atmosphere.

In addition to the north-south back-and-forth motion maintained by magnetic mirroring, radiation belt particles also engage in a longitudinal drift around the planet. This so called gradient-curvature drift is also a result of magnetic field topology, but is charge dependent, sending positively charged particles to the west and negatively charged particles to the east. This differential motion of positively and negatively charged particles constitutes the equatorial 'ring current'. This westward current (with its associated magnetic field directed opposite to that of the main geomagnetic field) produces magnetic field perturbations at Earth's surface. During magnetic disturbances (such as geomagnetic storms and substorms), plasma sheet particles are injected into the ring current, increasing its strength and consequently the magnitude of these perturbations. A geophysical index

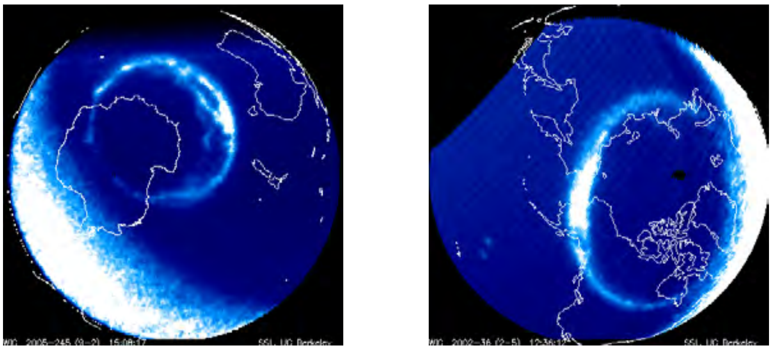


Figure 1.2.2: Wideband (false-colour) images recorded by the Far UltraViolet (FUV) instrument aboard the IMAGE satellite, showing the location of the southern (left) and northern (right) hemisphere auroral ovals during two separate events.

commonly used to track storm progress is derived from these perturbations arising from the ring current magnetic field, and is known as the Disturbance Storm Time (Dst) index. This index is derived from magnetic field measurements at low-latitude magnetometer stations, and tracks deviations of the horizontal component of the geomagnetic field.

At high magnetic latitudes, 'cusp' regions are formed between the highest-latitude closed field lines on the dayside and the lowest-latitude 'open' field lines that map to the distant nightside magnetotail, where solar wind particles can freely enter the magnetosphere and hence Earth's upper atmosphere. Aurora, a direct result of energetic particle entry into the upper atmosphere, often occur within the region between 'open' and closed field lines, along a band (in each hemisphere, as in Figure 1.2.2) known as the 'auroral oval', which on average is centred on approximately 67° magnetic latitude at magnetic midnight and at about 78° near magnetic noon (Brekke, 1997). The size of this oval is variable however, and can expand to lower latitudes during magnetic storms (Meng, 1984) and substorms (see Section 1.2.3), and contract towards the magnetic poles in association with variations in the IMF (Nakai et al., 1986). At ionospheric altitudes, particle precipitation in the auroral region is a source of both heat and enhanced conductivity.

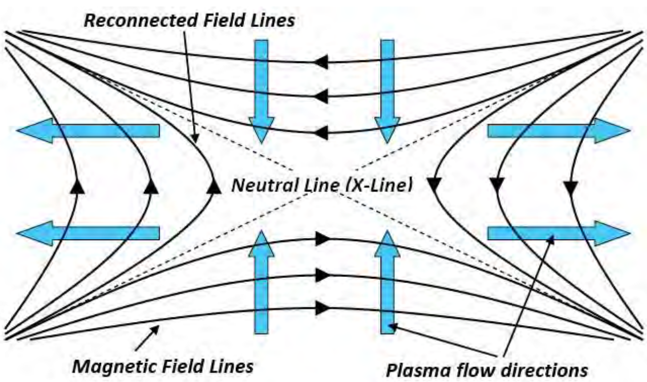


Figure 1.2.3: Magnetic reconnection. Field-lines lying in different magnetic domains (lines at the top and bottom of the figure) 'link-up', forming new lines (at left and right) which move away from the reconnection region. A neutral line is formed in the centre of this region (which in the diagram is extended in the direction normal to the page). Motion of the plasma is indicated by the blue arrows.

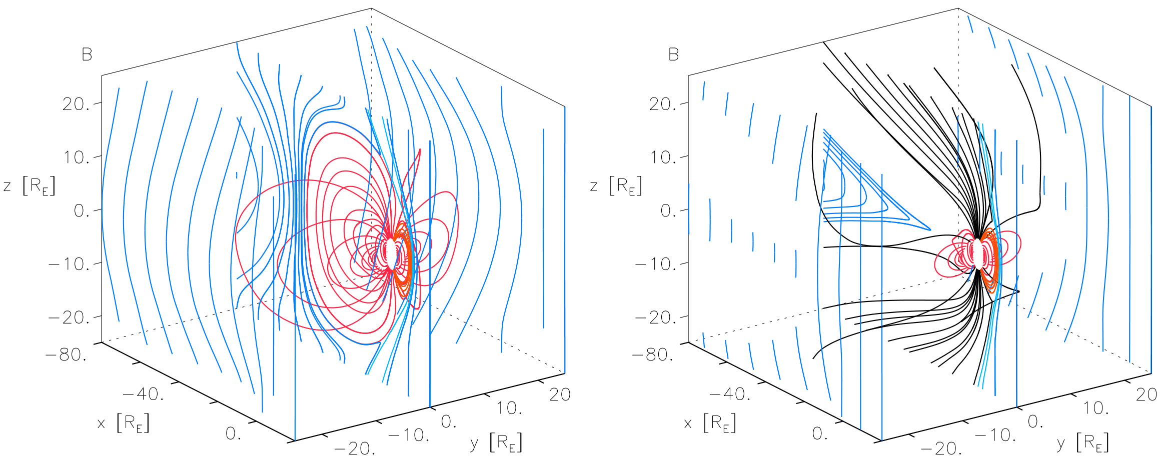


Figure 1.2.4: Results of two BATS-R-US (Block-Adaptive-Tree-Solarwind-Roe-Upwind-Scheme) model runs simulating northward (left panel) and southward (right panel) IMF B_z conditions. Magnetic field lines are colour-coded as 'Closed' - both ends anchored in the terrestrial field (red/orange), 'Open' - one end embedded in the IMF (black) and purely IMF (blue).

The configuration of the geomagnetic field is dependent upon the speed and density of the solar wind, as well as the strength and direction of the IMF. Of the latter, it is the north-south (B_z) component of the IMF that is the dominant factor affecting the magnetic field configuration, as it is principally this component which controls the level of coupling between the solar wind plasma and the magnetosphere. This coupling is the result of a process called *magnetic reconnection* or field line *merging*, whereby magnetic field lines lying in different magnetic domains can 'link-up' (see Figure 1.2.3), effectively creating new field lines (with a different topology) which move away from each other in opposite directions.

In terms of the magnetosphere, southward directed IMF field lines, as they encounter the geomagnetic field, are thought to reconnect or merge with the northward directed geomagnetic field lines at low or mid latitudes, forming new field lines which have one end anchored in the geomagnetic field and the other embedded within the IMF. Reconnection couples the magnetosphere to the solar wind firstly by allowing direct entry of solar wind plasma into the magnetic cavity along the newly opened field lines, and secondly through the transfer of momentum and energy as the opened field lines are swept antisunward by magnetic tension and the motion of the solar wind. As field lines move tailward and become tail lobe field lines, plasma trapped along these lines is transported toward the tail plasma sheet. This situation of predominantly southward-directed IMF is known as the 'open' magnetosphere, and produces the classic magnetospheric cavity shown in cross-section in Figures 1.2.1 and 1.2.4 (right panel).

During periods of northward IMF, dayside reconnection does not cease altogether, but is thought instead to shift to higher magnetic latitudes, where the new field configuration produces the greatest shear between terrestrial and interplanetary magnetic field lines (Amm et al., 2005). The rate of dayside reconnection will however be reduced, along with the level of coupling between the solar wind and the magnetosphere. In this configuration the magnetosphere is termed 'closed', since entry of solar wind plasma is more difficult under these conditions. The left panel in Figure 1.2.4 is an example of Earth's modelled magnetosphere during such northward IMF conditions.

1.2.1 Convection

Field lines which have been opened on the dayside and carried along with the solar wind can reconnect once again in the magnetotail, at the boundary between the oppositely directed northern and southern hemisphere field lines. After being reconnected in the tail, field lines Earthward of the reconnection region will be transported sunward, returning magnetic flux to the dayside and completing the cycle of magnetospheric 'convection', in which magnetic flux is transported from the dayside to the nightside over the polar cap, with a return flow at lower magnetic latitudes. The term convection applies to bulk motion of the plasma, invoking similarities to thermal convection. Magnetospheric convection is mapped along field lines down to the ionosphere (see Section 2.4.1), where it manifests as a twin-cell circulation of plasma, antisunward over the polar cap and sunward at lower latitudes. An ideal magnetised plasma with bulk flow velocity **v** must satisfy the

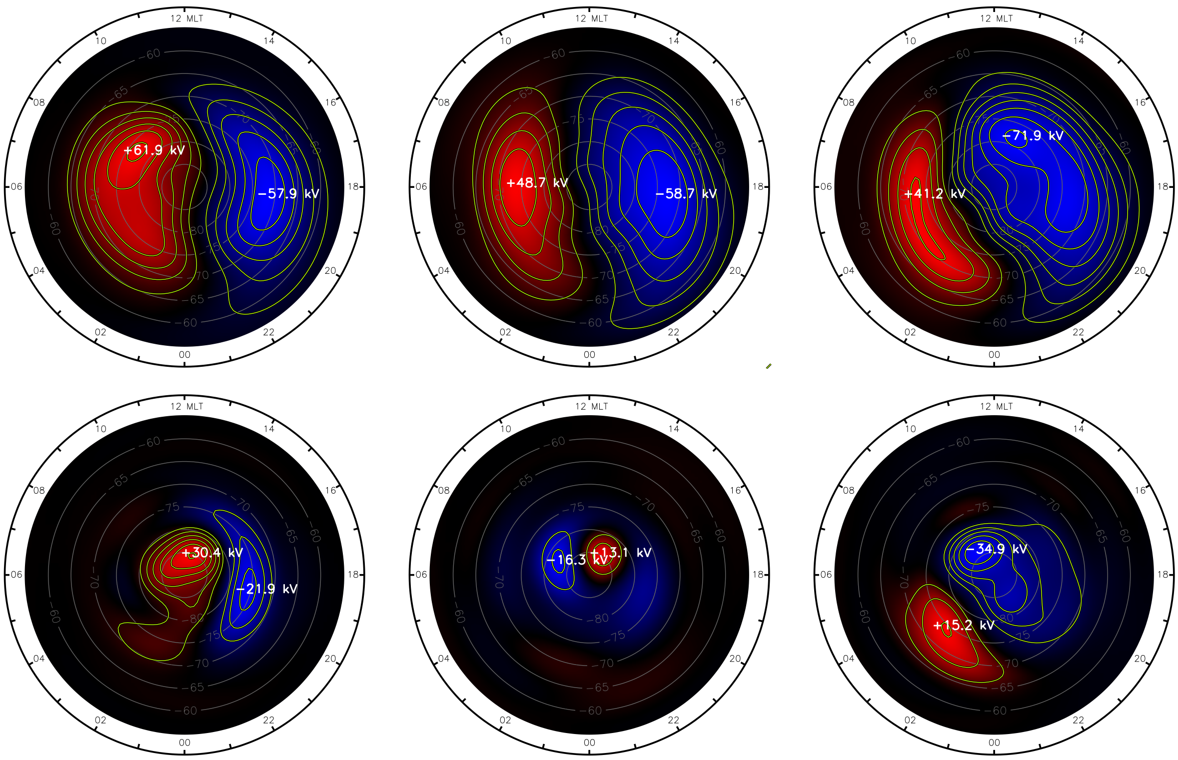


Figure 1.2.5: High-latitude electric potential maps under various IMF conditions, from the Weimer96 model (Weimer, 1996). Convection patterns are ordered, from left to right, according to the magnitude of the IMF B_y component (-5, 0 and +5 nT). The top row corresponds to southward IMF (negative B_z), the bottom row northward IMF (positive B_z). Maximum and minimum electric potentials in kilovolts are shown, with the polar cap electric field directed from the red to blue regions (ie. in the dawn-to-dusk direction during southward IMF). Magnetic local time is indicated around the outer circle, and the Sun is to the top of the figure.

magnetohydrodynamic (MHD) condition (Alfvén, 1950):

$$\mathbf{E} = -\mathbf{v} \times \mathbf{B} \tag{1.2.1}$$

Thus the pattern of convective magnetospheric flow is associated with an electric field, which is directed from dawn-to-dusk over the polar cap and from dusk-to-dawn in the auroral oval. High-latitude magnetic field lines, being (essentially) equipotential, map this magnetospheric (convection) electric field down to the ionosphere, where it combines with the (approximately) perpendicular magnetic field to drive a similar convection of the ionospheric plasma via $\mathbf{E}\times\mathbf{B}$ drift. Assuming a stationary state ($\frac{\partial \mathbf{B}}{\partial t} = 0$), then $\mathbf{E} = -\nabla\Phi$, and Equation 1.2.1 gives:

$$\begin{aligned}\mathbf{v} \cdot \nabla\Phi &= -\mathbf{v} \cdot \mathbf{E} \\ &= -\mathbf{v} \cdot (\mathbf{v} \times \mathbf{B}) \\ &= 0 \end{aligned} \tag{1.2.2}$$

Hence the streamlines of ionospheric plasma (at F-region altitudes, at least) form contours of constant electric potential. Observed convection patterns are often described in terms of the two-dimensional electric potential which would produce the observed flow. Typical polar-cap potential drops ($\Delta\Phi$) are on the order of 50 kV, but values up to 240 kV have been observed (Banks, 1975). Boyle et al. (1997) derived an empirical relation between the steady-state polar-cap potential (Φ_A) and solar wind-IMF parameters, given by:

$$\Phi_A = A_0 v^2 + A_1 B \sin^3\left(\frac{\theta}{2}\right) \quad \text{kV} \tag{1.2.3}$$

where:

$$A_0 = 10^{-4}\ \text{kV.s}^2\text{.km}^{-2}$$
$$A_1 = 11.7\ \text{kV.nT}^{-1}$$

where v is the solar wind speed in kilometres per second, B is the magnitude of the IMF in nanotesla, and $\theta = \arccos(B_z/B)$ in GSM coordinates. Example southern hemisphere electric potential maps under various IMF conditions are shown in Figure 1.2.5, derived from the Weimer 1996 ionospheric electric potential model (refer to Section 2.7). F-region plasma flow is in the $\mathbf{E} \times \mathbf{B}$ direction, following the electric potential contours, and thus for southward IMF is antisunward over the polar cap and sunward at auroral latitudes.

Convection is most stable during southward IMF, and displays asymmetries with respect to the direction of the east-west (B_y) component (see Section 2.4.1). Convection flows during northward IMF are less well defined, often reflecting a multicellular convection pattern instead of the usual twin-cell circulation. This process of convection is dependent upon the respective rates of dayside and nightside reconnection. If reconnection occurs faster on the dayside then more magnetic flux (and hence energy) will be carried into the tail than is being returned by nightside reconnection. This energy cannot be stored indefinitely, and a critical point will be reached at which this stored energy must be released, often impulsively, in the form of a magnetospheric substorm (see for example: Rostoker, 1972; Vasyliunas and Wolf, 1973; McPherron, 1979, and the discussion of Section 1.2.3).

1.2.2 Currents

When charged particles encounter a magnetic field they are caused to gyrate, via the Lorentz force, in a direction dependent on the sign of the charge. For example, a magnetic field line directed up out of the page, when viewed from above the page, will force an electron to gyrate anticlockwise, and a proton (or positive ion) clockwise. The radius and frequency of this gyration are critical parameters governing the behaviour of a plasma within a magnetic field, and are given by:

$$\text{Radius:} \quad r_{gyro} = \frac{mv}{|q|B} \qquad\qquad \text{Frequency:} \quad \omega_{gyro} = \frac{|q|B}{m} \tag{1.2.4}$$

When charged particles of the solar wind first encounter the geomagnetic field they gyrate, completing approximately one-half of an orbit before they are returned to the solar wind stream. During this half-orbit, however, positively charged particles will be forced to the right (when viewed along the solar wind stream, from the Sun towards Earth), while negatively charged particles will be effectively pushed to the left. A current is thus established, flowing in the dawn-to-dusk direction along the dayside magnetopause. The direction of this current is such that the magnetic field produced by it acts to strengthen the magnetic field inside the magnetosphere, and to cancel it outside, thus confining the

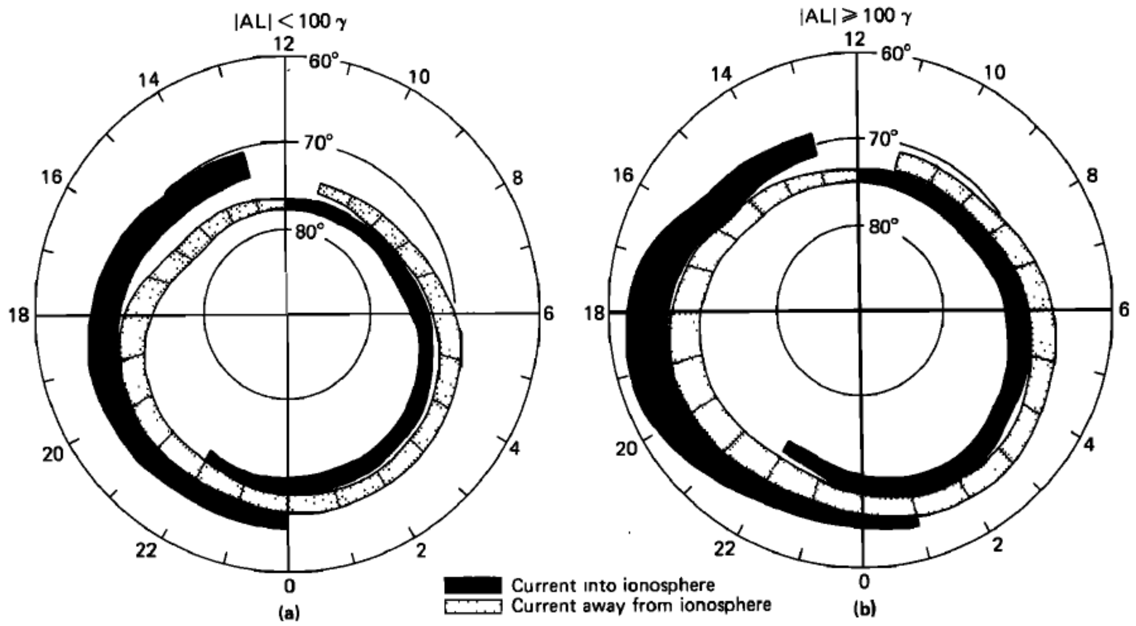


Figure 1.2.6: The statistical distribution of large-scale field-aligned currents, from Iijima and Potemra (1978). On the left are shown the distribution for weakly disturbed conditions (|AL| index $< 100\ \gamma$), with active conditions (|AL| index $\geq 100\ \gamma$) on the right.

geomagnetic field inside the magnetospheric cavity. This magnetopause current system is also known as the Chapmann-Ferraro current system (see Figure 1.2.7).

In the tail, the presence of oppositely directed magnetic field lines near the plasma sheet also implies a current flowing in the dawn-to-dusk direction, and closing along the boundary of the magnetotail. This cross-tail current is seen to intensify and move Earthward during the substorm growth phase (Kaufmann (1987), and Section 1.2.3). The ring current, mentioned in Section 1.2, can also experience an intensification during substorms as plasma is injected into it.

A well known ionospheric current system is the *auroral electrojet*, which is a direct result of magnetospheric convection (Section 1.2.1 and Figure 1.2.5) and the differential motion of positive ions and electrons brought about by collisions of the former with neutral atmospheric species. Electrons are free to move along the electric potential contours while the ions are retarded by collisions with the neutral species. A strong current is established due to the enhanced conductivity of the auroral zone, flowing westward on the dawnside and eastward on the duskside. While not in itself a magnetospheric current system, the ends of the electrojet currents are believed to partially close via currents flowing along magnetic field lines and joining with the ring current (Anderson and Vondrak, 1975). Such field-aligned currents are sometimes known as Birkeland currents, after the early work of Birkeland (1908). Birkeland currents (or more appropriately current *sheets*) are observed to flow into and out of the ionosphere all the way along the auroral oval (Iijima and Potemra, 1976, 1978). These sheets are aligned along the oval (see Figure 1.2.6), both poleward ('region 1' currents) and equatorward ('region 2' currents) of it, with dawnside current flow into the ionosphere along region 1 currents and out of the ionosphere along region 2 currents (flow directions of these current systems are reversed on the duskside). Most of this current closes across the auroral oval (hence flowing perpendicular to it), with a small amount of current either closing over the polar cap or around the auroral oval (Stern, 1996).

A more intermittent current system associated with the disruption of the cross-tail current during substorms is the substorm current wedge, SCW (Section 1.2.3). This system is another example of a field-aligned current, diverting the near-Earth cross-tail current along magnetic field lines into the ionosphere on the dawnside, and flowing back into the plasma sheet along duskside field lines. The current closes along the westward electrojet, producing the intensification of this ionospheric current observed during substorms. In both of the cases outlined (Birkeland currents and the SCW), the ionosphere acts as a resistive load to the magnetosphere, into which energy is lost mainly through Ohmic (Joule) heating.

Ionospheric currents (more correctly current *densities*) are often treated in terms of their components in the plane perpendicular to the local magnetic field direction **B**. These are the components along the local electric field direction (the Pedersen current, J_P) and perpendicular to it (the Hall current, J_H). Under the assumption of equipotential magnetic field lines (directed vertically upwards), the ionosphere can be approximated as a thin shell and the height-integrated Pedersen and Hall currents can be defined:

$$\mathbf{I}_P = \int_{z_0}^{z} J_P \; dz = \Sigma_P \mathbf{E} \tag{1.2.5}$$

$$\mathbf{I}_H = \int_{z_0}^{z} J_H \; dz = \Sigma_H (\mathbf{E} \times \mathbf{B}/B) \tag{1.2.6}$$

where Σ_P and Σ_H are respectively the height-integrated Pedersen and Hall conductivities ($\int_{z_0}^{z} \sigma_P \; dz$ and $\int_{z_0}^{z} \sigma_H \; dz$). The requirement of current continuity then leads to:

$$\begin{aligned} J_{\parallel} &= -\nabla_{\perp} \cdot \mathbf{I} \\ &= -\nabla_{\perp} \cdot \mathbf{I}_P - \nabla_{\perp} \cdot \mathbf{I}_H \end{aligned} \tag{1.2.7}$$

Lyons (1992) used typical ionospheric parameters to estimate the magnitude of these divergent currents:

$$\begin{aligned} \nabla_{\perp} \cdot \mathbf{I}_P &\simeq 10^{-5} \text{ Am}^{-2} \\ \nabla_{\perp} \cdot \mathbf{I}_H &\simeq 2 \times 10^{-8} \text{ Am}^{-2} \end{aligned} \tag{1.2.8}$$

The contribution from the Hall current is thus negligible, and therefore the field-aligned current $J_{\parallel}$ is due almost entirely to the divergence of the Pedersen current.

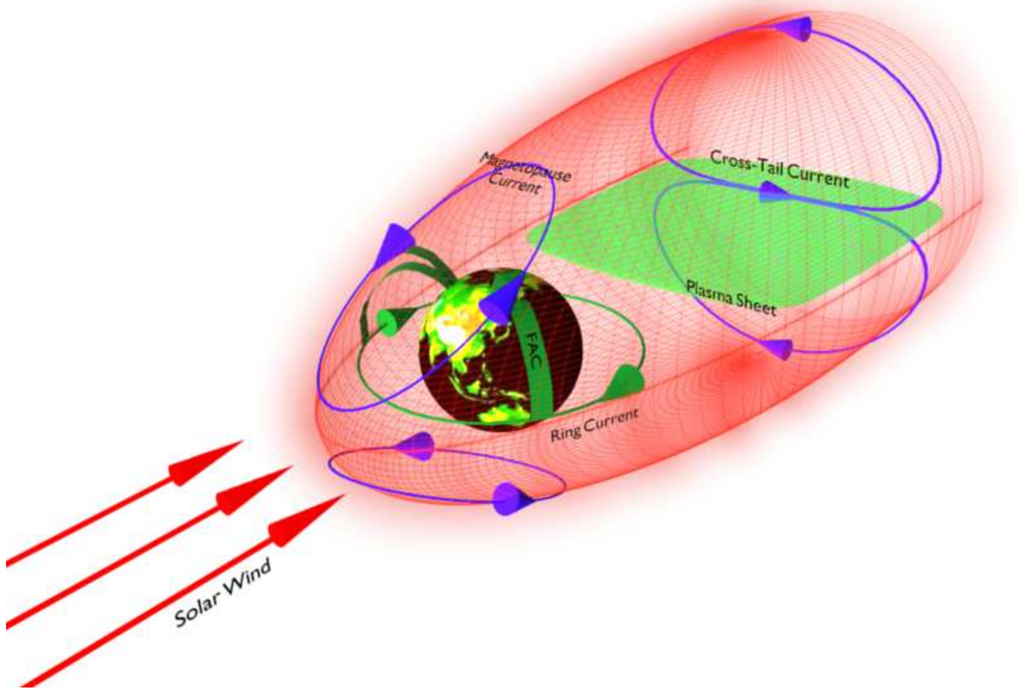


Figure 1.2.7: Three-dimensional schematic showing the main magnetospheric current systems.

1.2.3 Magnetospheric Disturbances

The magnetosphere owes its existence to the highly variable solar wind, and as such is a dynamic environment which is continually responding to changes originating in the Sun. Two main forms of magnetospheric disturbance are observed: *substorms*, the effects of which are predominantly observed at auroral latitudes and which last on average for perhaps 1 to 3 hours, and *storms*, which are a global phenomena that can last for days. These two modes of disturbance are described below.

Substorms

Syun-Ichi Akasofu, one of the pioneers of substorm research, gives this definition of a magnetospheric substorm (Akasofu, 2004):

> "The magnetospheric substorm is a fundamental mode of magnetospheric disturbance. It reflects the response of the magnetosphere to a specific solar wind IMF variation that results in an increase of energy transfer from the solar wind to the magnetosphere. During a magnetospheric substorm, a great variety of disturbances occur throughout the magnetosphere. Auroral substorms and polar magnetic substorms are manifestations of magnetospheric substorms in the polar upper atmosphere. When intense substorms occur frequently, a magnetospheric storm develops as their non-linear consequence."

While the precise mechanisms which trigger substorms are still not universally agreed upon, there is a general agreement about substorm energisation and subsequent morphology. This usually entails breaking the substorm process up into a number of stages which are roughly ordered in time. These stages are often referred to as the *growth* (after McPherron et al. (1973)), *expansion* and *recovery* phases respectively.

Growth: As the name implies, the first phase in substorm progression is one of growth, during which time the magnetosphere is loaded with excess energy above the magnetospheric 'ground state'. Rostoker et al. (1980) defines this as the state which the magnetosphere reaches after a prolonged period of northward IMF. Excess magnetospheric loading can result from a southward turning of the IMF, which more effectively couples the magnetosphere to the solar wind via the process of reconnection outlined in Section 1.2. This coupling results in a transfer of magnetic flux from the dayside magnetosphere into the magnetotail, where a second site of reconnection allows tailward-driven flux to return to the dayside. The growth phase is thus an example of what is termed a 'directly driven' process, driven as it is by a direct input of energy from the solar wind (Elphinstone et al., 1996). The beginning of the growth phase is often coincident with a southward turning of the IMF. Note here that this need not be an absolute change of IMF B_z to a negative value; a change of B_z from +5 nT to +1 nT still constitutes a southward turning, and can result in a weak substorm (Akasofu, 2004).

If the rate of dayside reconnection is higher than that of the nightside, then the dayside magnetopause is 'eroded', as the low rate of returning flux is unable to balance the tailward flux transport, pushing the magnetopause boundary closer to Earth (Holzer and Slavin, 1978). As a consequence, magnetic flux builds up in the magnetotail, increasing the lobe field strength (McPherron, 1979). At the same time, the auroral oval is observed to expand equatorward (Grocott and Yeoman, 2006), as is the boundary between energetic ($\simeq 45$ keV) particle precipitation and the softer precipitation poleward of it. The cross-tail current is often enhanced and its inner edge seen to move Earthward (Kaufmann, 1987), accompanied by a thinning of the plasma sheet. This results in a reconfiguration of the near-Earth magnetic field from the quiet-time dipolar configuration to a more stretched, 'tail-like' configuration.

Expansion: The magnetotail cannot support this energy loading indefinitely, and a point will be reached at which this energy must be released. The expansion phase is characterised by a sudden brightening of the equatorward edge of the discrete aurora (Friedrich et al., 2001). This so-called auroral breakup is statistically most likely to occur around 22:30 magnetic local time (Liou et al., 2001). Following the definition of Rostoker et al. (1980), this phase will often exhibit multiple intensifications of the westward auroral electrojet, each associated with a westward travelling surge (a large auroral form resembling a spiral moving westward along a pre-existing arc) and a Pi2 magnetic pulsation (Lee, 1998). Auroral activity expands poleward and westward behind the surge, forming an auroral 'bulge' (Fukunishi, 1975). The electrojet intensifications are also associated with a disruption of the cross-tail current sheet, as some portion of this current is diverted to flow along magnetic field lines (the substorm current wedge), closing in the ionosphere as the westward electrojet current. This clearly entails a corresponding intensification of the region 1 and 2 Birkeland currents which form the bridge between the ionosphere and magnetosphere (Birn and Hesse, 1991; Frank et al., 2002). Precipitation of energetic particles into the auroral zone is generally increased, and plasma is injected into the ring current (Frank et al., 2002).

In the near-Earth region, the previously tail-like field is rapidly returned to a more dipolar configuration in a process called *dipolarisation*. The cause of this dipolarisation is under debate. Cross-tail current disruption is one possibility, and Akasofu (2004) suggests that the westward electrojet is actually closed by an eastward-directed ('counter-cross-tail') current within the plasma sheet, which would produce the required dipolar field configuration Earthward of the current, and could also explain cases of over-dipolarisation by requiring that the magnitude of the counter-cross-tail current exceed that of the cross-tail current. Reconnection is another process often invoked to explain the rapid return to a dipole field configuration, as this mechanism (by definition) links the oppositely directed lobe fields into the required dipole configuration. Around 20-30% of the open flux in the tail lobes is estimated to be reconnected in this manner (Baker et al., 1999). Associated with this linkage is an Earthward (and tailward) motion of reconnected field lines on opposite sides of the reconnection region due to the magnetic tension stored within them.

Further out in the tail, the lobe field strength decreases, and the radius of the tail shrinks (McPherron, 1979). Rapid thinning of the plasma sheet has been observed beyond about 15 R_E, with some studies also reporting rapid thinning near the dipolarisation region closer to Earth (Lui et al., 2008; Sergeev et al., 2008). Tailward moving plasma flows with velocities of $\sim$ 300-800 km.s^{-1}, called *plasmoids*, have also been observed (Slavin et al., 1998, 2002), associated with southward magnetic fields, suggesting magnetic reconnection in the tail.

Substorm onset, the point in time at which the expansion phase is considered to have begun, thus coincides with the onset of what are known as 'loading-unloading' events. Phenomena observed during the expansion phase are likely to be a result of both directly driven (DD) and loading-unloading (UL) events. The DD component varies slowly compared to the impulsive UL component, and is directly related to the level of solar wind-magnetosphere coupling, responding approximately linearly to the convection electric field (Akasofu, 2004). For example, enhanced ionospheric convection is associated with the DD component. The UL component, on the other hand, depends mainly on energy stored within the magnetotail, and as such is not tied directly to the level of solar wind-magnetosphere coupling as is the DD component. UL events are intensive and unpredictable. The intensification of the westward electrojet (likely as a result of the substorm current wedge development), for example, is a UL event (Kamide et al., 1996), as is the intensification of the ring current due to plasma injection. Baker et al. (1997) used a non-linear dynamical model of global substorm processes to model the AL(12) index during substorms, and found that both DD and UL components were necessary to generate realistic behaviour of this index. Tanskanen et al. (2002), while agreeing with the need for energy loading during the substorm growth phase, concluded that substorm size was mostly governed by the direct energy input during the expansion phase. In a more recent study of IMAGE-SI12 imager data, Blockx et al. (2009) examined 256 substorms between June 2000 and December 2002 and concluded that loading-unloading was the statistically dominant process.

While the exact mechanism responsible for triggering substorm expansion continues to be investigated, two leading scenarios have emerged to explain this process in terms of internal triggering. The first is the Near-Earth Neutral Line model (NENL), in which magnetic reconnection takes place in the mid-tail between around 20-30 R_E, accelerating plasma sheet particles both toward and away from Earth. The Earthward accelerated plasma is slowed by the near-Earth magnetic field, creating an eastward current opposing the cross-tail current, and generating the substorm current wedge (Lui, 2001). Baker et al. (2002), in a case study of a single isolated substorm in August 2001, were able to construct a time-line consistent with magnetic reconnection around 18 R_E downstream of Earth approximately 7 minutes before the expansion phase was registered from ultraviolet auroral images. Similarly, Angelopoulos et al. (2008) used data from the "Time History of Events and Macroscale Interactions during Substorms" (THEMIS) mission to construct a time-line of events in one case study which supported the idea of tail reconnection as the likely substorm trigger for that event. Ohtani et al. (2002) compared Geotail data from two extreme events, and found reconnection signatures (very fast plasma flow with a strongly southward magnetic field) without significant magnetic disturbances in the auroral zone, suggesting that lobe reconnection does not always lead to a global substorm.

The second popular expansion phase trigger is known as the Current Disruption (CD) model (e.g. Lui, 1996). Under this scheme, plasma instabilities originating in the near-Earth region (6-10 R_E) disrupt the cross-tail current and are responsible for initiating the substorm current wedge (Amm et al., 2005). The current disruption expands tailward, where it can create the right conditions for reconnection to occur in the mid-tail region. Observations of a single substorm reported by Lopez et al. (1993), for example, were interpreted as being more consistent with the CD model than with the NENL model. Friedrich et al. (2001) studied eight isolated substorms, and concluded that expansion phase onset occurred some 1-5 minutes prior to lobe reconnection, also consistent with the current disruption model.

As well as the above internal triggering mechanisms, which are caused by an instability within the magnetosphere, evidence of external triggers has been found in the close association between IMF conditions and substorm onsets. Kamide et al. (1977) found that when the 1-hour-averaged B_z component of the IMF was less than -5 nT there was a 100% probability of substorm occurrence. Lyons et al. (1997) developed criteria for identifying a substorm trigger in IMF data which corresponded to a northward turning of the IMF. They found that the chances of this trigger being associated with substorm onset purely by chance was insignificant. Hsu and McPherron (2003) concluded that about 60% of substorm onsets were associated with a northward turning of IMF B_z, and suggested that substorm onset is a consequence of an internal magnetospheric instability that is highly sensitive to changes in magnetospheric convection (induced by sudden changes in IMF), but that these changes are not always necessary. Morley and Freeman (2007), using an internally triggered, minimal substorm model, concluded that a northward turning of the IMF was not a necessary condition for substorm onset, but that an internal magnetospheric instability, following an interval of solar wind energy input into the magnetosphere (due to a southward IMF), was a necessary and sufficient condition for substorm onset.

Recovery: The recovery phase is the period during which the magnetosphere (and ionosphere to which it is coupled) return to their pre-substorm configurations. The auroral bulge halts its poleward expansion and moves equatorward, while the visual auroral displays become dimmer. The strength of the westward electrojet decreases, along with the associated field-aligned currents (McPherron, 1979), while the plasma sheet returns to its original size. The auroral oval contracts poleward, and pulsating auroral patches are commonly observed (Elphinstone et al., 1996).

Timescales: While substorms vary greatly in both intensity and duration, it is useful to examine the mean timescales of each of the above substorm phases. It is difficult to arrive at a typical duration of the growth phase, since there is no universal way of defining when the growth phase begins. The subsequent storm intensity will depend upon the duration of the growth phase as well as on the rate of energy input from the solar wind, since these factors determine the amount of energy stored in the magnetotail and available for release in the form of a substorm. As a general guide, Lyons et al. (1997), based on their triggering probabilities, estimated a median growth phase duration of $\sim$ 64-72 minutes.

Chua et al. (2004) used the value of hemispheric power derived from northern hemisphere Polar UVI auroral images to define substorm expansion and recovery phases, and applied this method to over 300 substorms between 1996 and 2001. Onset was determined by the time at which the hemispheric power increased above one standard deviation of the power in the preceding 30 minutes. The expansion phase was then defined as the time between substorm onset and the time at which hemispheric power peaked, while the recovery phase was quantified by the value of τ, the time taken for the hemispheric power to drop to $1/e$ of its peak value (the e-folding time). Typically, the time taken for hemispheric power to drop to pre-storm levels was found to be slightly greater than 4τ (for $B_z < 0$). Under these conditions of southward IMF, mean expansion phase duration was found to be 36.1 min, with a standard deviation of 27.6 min. The e-folding time τ showed a similar duration, with a mean of 32.8 min and a standard deviation of 19.1 min. Thus the mean duration of the recovery phase was on the order of 2 hours or more. Both expansion and recovery phases (in terms of τ) were found to be longer during intervals of southward IMF when compared with substorms occurring during northward IMF, which showed timescales 40% and 12% shorter respectively. These authors did however note that their data were biased toward southward IMF conditions.

Storms

Substorms are predominantly a high-latitude phenomenon, in that their effects are observed most commonly at auroral latitudes. They are seen as a fundamental mode of magnetospheric activity, with isolated substorms lasting on the order of a few hours. Magnetospheric storms on the other hand are a global phenomenon, which typically last much longer. Storms produce extraordinary fluctuations in Earth's magnetic field, due to an enhancement of the trapped magnetospheric particles which form the ring current (Gonzalez et al., 1994). Figure 1.2.8 shows the Dst signature of the March 1989 superstorm, during which time the Dst index almost reached -600 nT. Storms can be characterised by this sudden decrease in Dst, which is called the *main phase*, followed by a gradual recovery

Storm Level	Dst (nT)	B_z (nT)	ΔT (hours)
Intense	-100	-10	3
Moderate	-50	-5	2
Small (typical substorm)	-30	-3	1

Table 1.2: Classification of storm intensities and IMF prerequisites at an 80% occurrence level, from Gonzalez et al. (1994). Dst is the peak Dst reached during the storm main phase, B_z is the required strength of the IMF north-south component, and ΔT is the length of time for which this IMF component must be sustained for an 80% chance of a storm occurring at the given level of intensity.

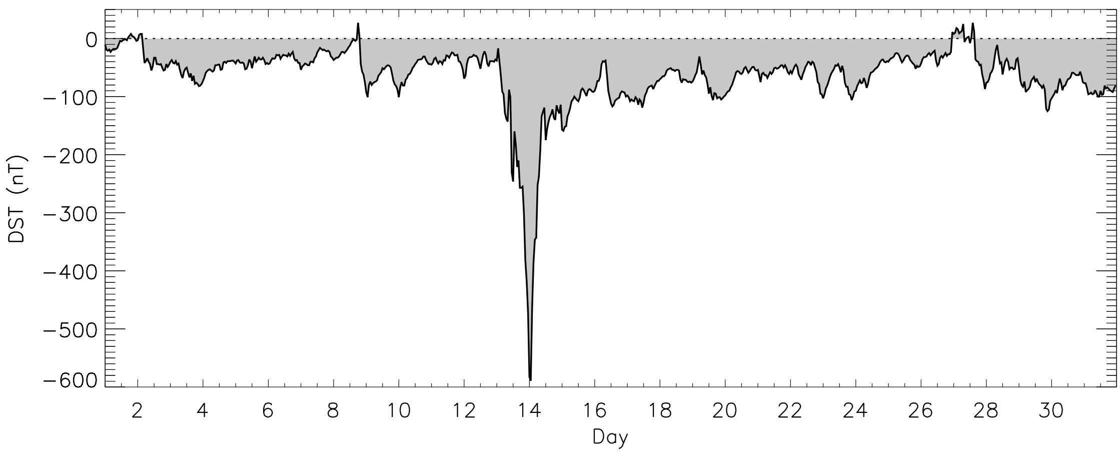


Figure 1.2.8: The March 1989 superstorm.

of Dst back to pre-storm levels. The main phase typically lasts 10-20 hours, while the recovery phase can last from 1 to 2 days (Meloni et al., 2005). Some storms also exhibit an initial increase in Dst (called a storm sudden commencement, or SSC) which is due to a sudden compression of the dayside magnetosphere due to an increase in the solar wind ram pressure (Kamide et al., 1998). Gonzalez et al. (1994) provides Table 1.2 as a classification of storm intensities based on the peak level of Dst reached during the main phase, and the IMF conditions which have been found to initiate them (at an occurrence level of 80%). The strong, long duration southward fields required for magnetic storms can be provided by coronal mass ejections, which are the main cause of non-recurrent magnetic storms during solar maximum (see Section 1.1.2). Recurrent magnetic storms are caused primarily by high-speed solar wind streams from coronal holes, and are therefore modulated by the 27-day rotation period of the Sun (Meloni et al., 2005).

A long standing issue is that of the relationship between storms and substorms. Substorms are known to occur in the absence of a storm, however no storms have been observed in the absence of substorms (Gonzalez et al., 1994). From the discussion of substorms in Section 1.2.3, it was noted that when the north-south component of the IMF was less than -5 nT for a period of an hour then there was a 100% probability of a substorm occurring. From Table 1.2, this condition is included in that for a moderate magnetic storm, which will occur with an 80% probability if IMF B_z is less than -5 nT for 2 hours or more. Thus substorms are almost guaranteed during a storm of this intensity.

It is sometimes assumed that magnetic storms are the result of the frequent occurrence of substorms (indeed this is why they were termed *sub*storms), and that ring current injections (which are responsible for the decrease in Dst during the main phase) are simply a superposition of substorm injections. However Reeves and Henderson (2001) found that storm-time injections (meaning those that occur during the storm main phase) are more effective than injections during isolated substorms. They noted that during the first hour of the storm there was continual injection of particles into the inner magnetosphere, which is generally not present during isolated injections. In addition to this, it is possible to model with reasonable accuracy the growth and decay of storm-time Dst using solar wind variables alone, indicating that knowledge of substorm activity is not required to predict Dst evolution during storms (Kamide et al., 1998). Both storms and substorms are processes which redistribute solar wind energy within the magnetosphere-ionosphere system, and are therefore likely to produce a coupled and complex response to enhanced solar wind energy input, such as occurs during long periods of southward IMF. The question of whether a storm is a fundamental magnetospheric response, and not merely a superposition of frequent and intense substorms, continues to be debated.

Chapter 2

Introduction to Earth's Upper Atmosphere

2.1 Overview

In Chapter 1 the role of the Sun as the driving force for space weather was discussed, as were the processes by which the energy from the solar wind is coupled to Earth's upper atmosphere via the magnetosphere. The focus of this chapter is the region of Earth's upper atmosphere between approximately 100 and 300 km, a region known as the *thermosphere*. The temperature profile shown in Figure 2.1.1 can be used to define several regions of Earth's atmosphere. The thermosphere refers to the region lying above the temperature minimum near 90 km, called the mesopause. Above the mesopause, temperature increases with increasing altitude, until a height is reached at which the thermal conductivity of the atmosphere redistributes heat so efficiently that the vertical temperature gradient is forced to be essentially zero. The temperatures reached at this altitude are highly dependent on solar activity, and can range from around 600 K during a quiet solar period to 1500 K or higher with an active sun. The lower thermosphere encompasses most of the ionosphere (which also extends below the mesopause) and extends upwards until the vertical temperature gradient becomes negligible, at which point the exosphere is usually considered to begin.

The 'ionosphere' refers to the plasma component embedded within the neutral upper atmosphere. This ionized region plays a major role in the composition, energetics and dynamics of the thermosphere, and must therefore be treated along with the neutral component. At high latitudes the configuration of the geomagnetic field allows for the entry into the upper atmosphere of particles from the solar wind and plasmasphere. The collisional interaction of these energetic *auroral* particles with species in the upper atmosphere results in heat deposition, ionisation, excitation and the consequent emission of visible and infrared radiation. High-latitude electric fields mapped down to ionospheric altitudes from the magnetosphere drive a circulation of the plasma component, which, as well as depositing heat to the neutral gas through friction, drags the neutral atmosphere along with it. The behaviour of the upper atmosphere can be described in terms of three main characteristics: composition, energetics and dynamics. It must be kept in mind however that these areas are tightly coupled; each affects and is affected by the other two.

2.2 Composition

The pressure and density of the atmosphere decrease approximately exponentially with increasing altitude. Under the assumption of hydrostatic equilibrium, for which the buoyancy force and gravity are balanced, the rate of change of pressure p (kg.m^{-1}.s^{-2}) with altitude z (m) in an isothermal atmosphere is given by:

$$\begin{aligned} \frac{dp}{dz} &= -\rho g \\ &= -nmg \end{aligned} \tag{2.2.1}$$

where ρ is the mean mass density (kg.m^{-3}), n the particle number density, m the mean mass (kg) and g the gravitational acceleration (m.s^{-2}). Particle number densities on the order of 10^{13} m^{-3} ensure a sufficient number of collisions occur on time scales of minutes and over distances of several cubic metres (Hargreaves, 1992, p.112), such that the quantities pressure, temperature and density are well defined, and under the assumption

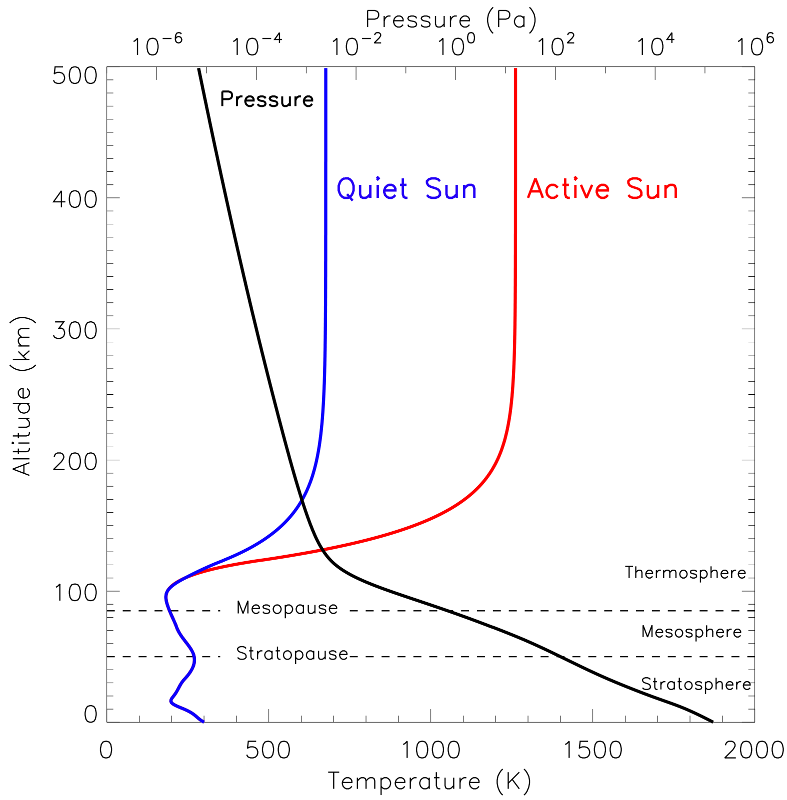


Figure 2.1.1: Altitude profiles of atmospheric temperature and pressure. Data were generated using the NRLMSISE-00 atmospheric model (described by Picone et al., 2002).

of an ideal gas the equation of state which relates these quantities is the ideal gas law:

$$
\begin{aligned}
pV &= NkT \\
\Rightarrow p &= nkT
\end{aligned}
\tag{2.2.2}
$$

where N is the number of particles, T is the temperature (K), and k is the Boltzmann constant ($\simeq 1.38 \times 10^{-23}$ J.K^{-1}). The pressure is then obtained as a function of altitude by substituting 2.2.2 into 2.2.1, and integrating:

$$
\begin{aligned}
\frac{1}{p}\frac{dp}{dz} &= -\frac{mg}{kT} \\
\Rightarrow \int_{p_0}^{p} \frac{dp}{p} &= -\frac{mg}{kT}\int_{z_0}^{z} dz \\
\Rightarrow p &= p_0 \exp\left(-\frac{z - z_0}{H}\right)
\end{aligned}
\tag{2.2.3}
$$

where:

$$
H = \frac{kT}{mg} \tag{2.2.4}
$$

In Equation 2.2.3, p_0 is the pressure at altitude z_0. The quantity H is called the 'scale height', and is equal to the height range over which the pressure (and density) changes by a factor of $1/e$, assuming T and m are constant with height.

Below around 100-110 km the gross composition of the atmosphere remains relatively constant, both spatially and temporally, due to turbulent mixing (however minor species may still occur at preferred heights). This region is known as the turbosphere (or homosphere), and is separated from the regions above it by the turbopause (homopause). Above the turbopause, however, turbulent mixing rapidly ceases, and in its absence the composition of the atmosphere becomes much more structured. The composition is governed by the need to conserve species number densities. In a given volume V (with closed surface S), the concentration n_j of species j must obey the relation:

$$
\iiint\limits_V \frac{\partial n_j}{\partial t}\,\mathrm{d}V + \iint\limits_S n_j(\mathbf{u}_j \cdot \hat{\mathbf{a}})\,\mathrm{d}S = \iiint\limits_V (P_j - L_j)\,\mathrm{d}V \tag{2.2.5}
$$

where $\mathbf{u}_j$, P_j and L_j are the altitude-dependent velocity, production rate and loss rate respectively for the jth species, and $\hat{\mathbf{a}}$ is the unit vector normal to the surface (directed outwards from the volume). The two terms on the left-hand-side represent the time rate of change of number density and the total inflow/outflow of number density through the surface of the volume (i.e. the number *flux* into or out of the volume). Applying Gauss's theorem allows this surface integral to be replaced by the volume integral of the divergence of $n_j\mathbf{u}_j$, and thus Equation 2.2.5 can be expressed in point form:

$$
\frac{\partial n_j}{\partial t} + \nabla \cdot (n_j\mathbf{u}_j) = P_j - L_j \tag{2.2.6}
$$

Although horizontal variations in density and composition are observed, they are generally small compared to variations with altitude. As such, Equation 2.2.6 is often reduced to the one-dimensional (altitude dependent) case, where the total flux is replaced by the flux in the z (vertical) direction only. To a first approximation the vertical density and composition of the major thermospheric species is controlled by the flux, which, in the absence of large vertical winds, is dominated by diffusive separation. This separation refers to the differential diffusion of atmospheric species with different molecular weights according to their individual scale heights (Equation 2.2.4).

The neutral composition of the thermosphere (see Figure 2.2.1) is dominated by molecular nitrogen and atomic oxygen, in contrast to the lower atmosphere where oxygen exists almost entirely in molecular form. Molecular and atomic oxygen exist in similar amounts at around 110 km, but above this altitude the O:O_2 ratio rapidly increases due to the dissociation of O_2, principally by solar ultra-violet radiation in the Schumann-Runge continuum ($130.0 \leq \lambda \leq 175.0$ nm). Photodissociation of O_2 by radiation within this range leaves one of the dissociated atoms in the electronically excited (^1D) state:

$$O_2 + h\nu(\lambda \leq 174.8 \text{ nm}) \rightarrow O + O(^1D) \tag{2.2.7}$$

The most efficient loss mechanisms for atomic oxygen are:

$$O + O + M \rightarrow O_2 + M \tag{2.2.8}$$

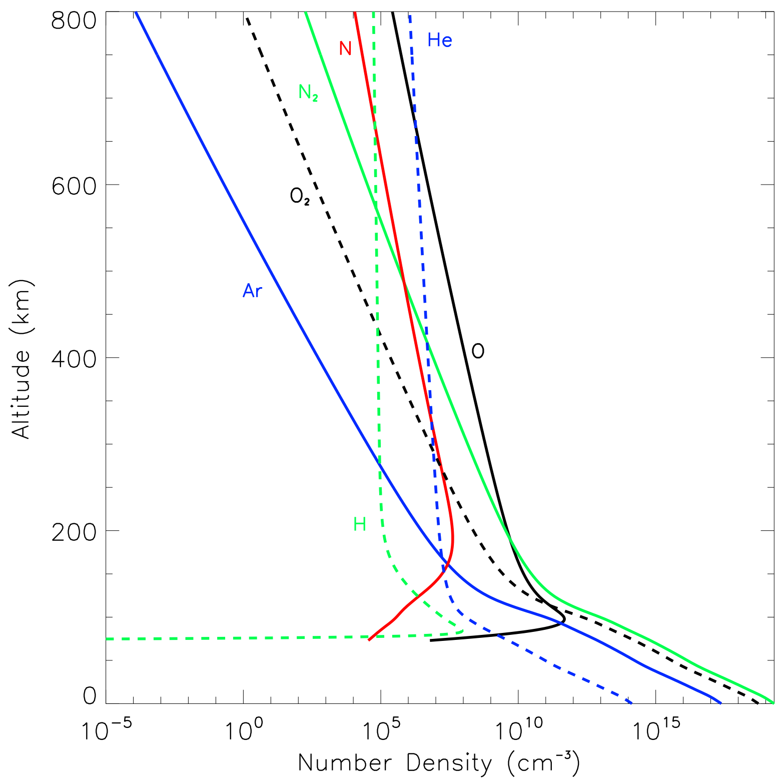


Figure 2.2.1: Vertical number density profiles of the principal thermospheric species. Data were generated using the NRLMSISE-00 atmospheric model (described by Picone et al., 2002).

and the pair of reactions:

$$O_2 + O + M \rightarrow O_3 + M \tag{2.2.9}$$
$$O_3 + O \rightarrow 2O_2 \tag{2.2.10}$$

where M represents a third species (typically N_2 at thermospheric altitudes) required in the three-body reactions to carry away excess energy. At an altitude of 100 km the O-O-M collision frequency is so low that the lifetime of an oxygen atom is on the order of a month, increasing rapidly with height, whereas lower down in the mesosphere lifetimes are much shorter. Atomic oxygen must therefore diffuse downward before it can recombine to form molecular oxygen or ozone.

Due to it's very small dissociation cross-section, nitrogen remains predominantly in its molecular state below around 500 km. At exospheric heights (above around 600 km) He and H become the dominant constituents due to their small molecular mass and correspondingly large individual scale heights. Not shown in Figure 2.2.1, but playing an important role in the radiative cooling of the lower thermosphere, is nitric oxide. NO concentration typically peaks around 100-110 km, with greater concentrations at high-latitudes and during periods of high geomagnetic activity (Hedin, 1979). At a fixed alti-

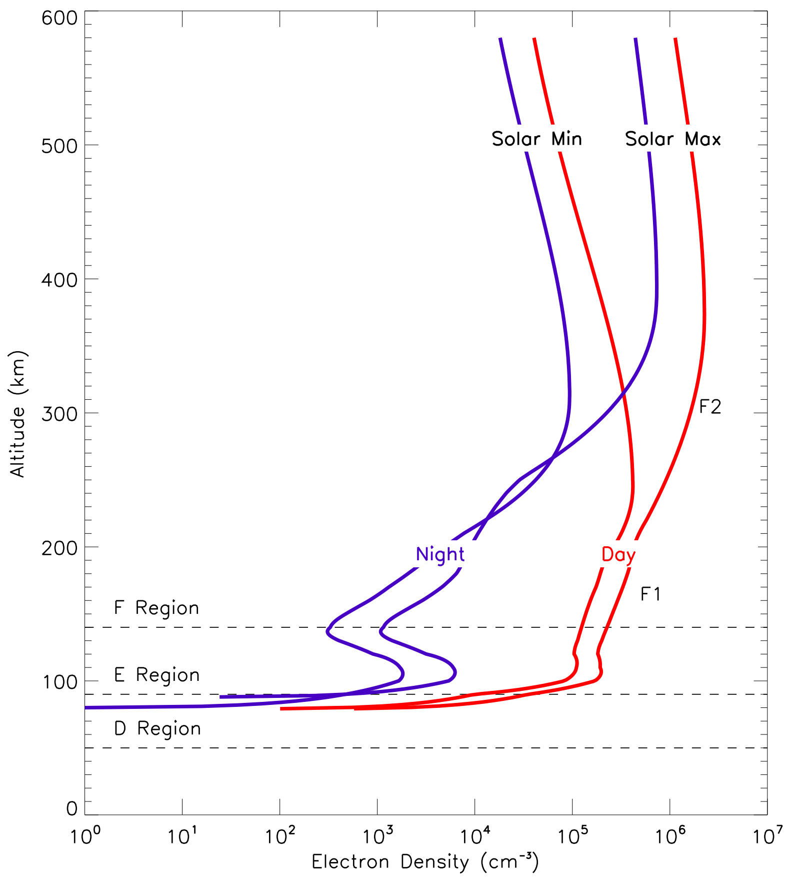


Figure 2.2.2: Electron density variation with altitude. Data were generated using the International Reference Ionosphere (2001) model (see Section 2.7).

tude, the summer NO concentration can be 3-4 times as high as in winter (Wells et al., 1997).

The ionosphere is produced primarily during the day by photoionisation of the major thermospheric species, O, N_2 and O_2, by solar radiation in the X-ray and extreme ultraviolet bands ($\lambda < 102.6$ nm). The ionosphere is separated into regions based on the altitude variation of electron density, which, by charge neutrality, is assumed equal to the sum of the positive ion densities (negative ions being present only in trace amounts). Figure 2.2.2 shows model electron density profiles for day and night, solar minimum and maximum conditions, and the naming scheme applied to the various plasma density peaks.

Photoionisation proceeds via:

$$\mathrm{N_2} + h\nu(\lambda \leq 79.6\mathrm{nm}) \rightarrow \mathrm{N_2^+} + e \tag{2.2.11}$$

$$\mathrm{O} + h\nu(\lambda \leq 91.1\mathrm{nm}) \rightarrow \mathrm{O^+} + e \tag{2.2.12}$$

$$\mathrm{O_2} + h\nu(\lambda \leq 102.6\mathrm{nm}) \rightarrow \mathrm{O_2^+} + e \tag{2.2.13}$$

Photons of sufficient energy are able to both dissociate *and* ionise molecular species:

$$\mathrm{N_2} + h\nu(\lambda \leq 51.0\mathrm{nm}) \rightarrow \mathrm{N^+} + \mathrm{N} + e \tag{2.2.14}$$

$$\mathrm{O_2} + h\nu(\lambda \leq 66.2\mathrm{nm}) \rightarrow \mathrm{O^+} + \mathrm{O} + e \tag{2.2.15}$$

Collisional sources are also important:

$$\mathrm{N_2} + e_p \rightarrow \mathrm{N_2^+} + e_p + e_s \tag{2.2.16}$$

$$\mathrm{N_2} + e_p \rightarrow \mathrm{N^+} + \mathrm{N} + e_p + e_s \tag{2.2.17}$$

$$\mathrm{O_2} + e_p \rightarrow \mathrm{O_2^+} + e_p + e_s \tag{2.2.18}$$

$$\mathrm{O_2} + e_p \rightarrow \mathrm{O^+} + \mathrm{O} + e_p + e_s \tag{2.2.19}$$

$$\mathrm{O_2} + e_p \rightarrow \mathrm{O} + \mathrm{O^*} + e_p \tag{2.2.20}$$

$$\mathrm{N_2} + e_p \rightarrow \mathrm{N} + \mathrm{N^*} + e_p \tag{2.2.21}$$

where e_p is a photoelectron or primary auroral electron, and e_s is an ejected (or secondary) electron. An asterisk superscript denotes a species in an excited state. The fate of such a species depends on the frequency of collisions with other species (including electrons) and the radiative lifetime of the state. For excited states which decay via a 'forbidden' transition (the so called 'metastable' states), radiative lifetimes can be long enough that collisional deactivation (or *quenching*) occurs:

$$\mathrm{M^*} + \mathrm{X} \rightarrow \mathrm{M} + \mathrm{X^*} \tag{2.2.22}$$

where X can be an atom, molecule or electron. Because metastable species are more likely to undergo quenching, these species play a major role in the thermal structure of the upper atmosphere, as some fraction of the excitation energy is redistributed as heat to the ambient gas (Torr and Torr, 1982). If the radiative lifetime is short, or the collision frequency sufficiently low, the species may return to its ground configuration through the emission of radiation (called 'airglow', see Section 2.6):

$$\mathrm{M}^* \rightarrow \mathrm{M} + h\nu \tag{2.2.23}$$

The $\mathrm{O}(^1D)$ produced in Equation 2.2.7 is an example of a metastable state of oxygen. This state experiences severe quenching below the F-region (Rees, 1989), but above this height can radiate by emitting a photon at 630.0 nm.

As well as the sources discussed above, there are many more indirect sources of atomic, molecular and ionic species resulting from chemical reactions. While these reactions do not greatly affect the concentrations of the major neutral species, they are very important for the formation of the ionosphere since they operate continuously, allowing the ionosphere to be maintained at night, albeit at lower plasma densities. These reactions are also the only source of thermospheric NO, which, although only a minor neutral species, is a major species in its ionic form (NO^+) at E-region heights. Production of NO^+ relies principally on chemical reactions since the concentration of the neutral molecule is too low for photoionisation or electron impact to be effective (Rees, 1989)

2.2.1 Compositional Variations

There are many processes which can perturb the density and composition of the thermosphere on a range of temporal and spatial scales. There is for example a diurnal variation (or 'tide') in constituent densities due to solar insolation, as shown in Figure 2.2.3. A fraction of the incoming solar radiation goes into heating of the atmospheric gas. This heating results in an expansion on the dayside, raising the density height profiles of the thermospheric constituents. The pressure gradient set up by this heating in turn drives a global-scale circulation, which results in a *wind-induced diffusion* whereby (primarily)

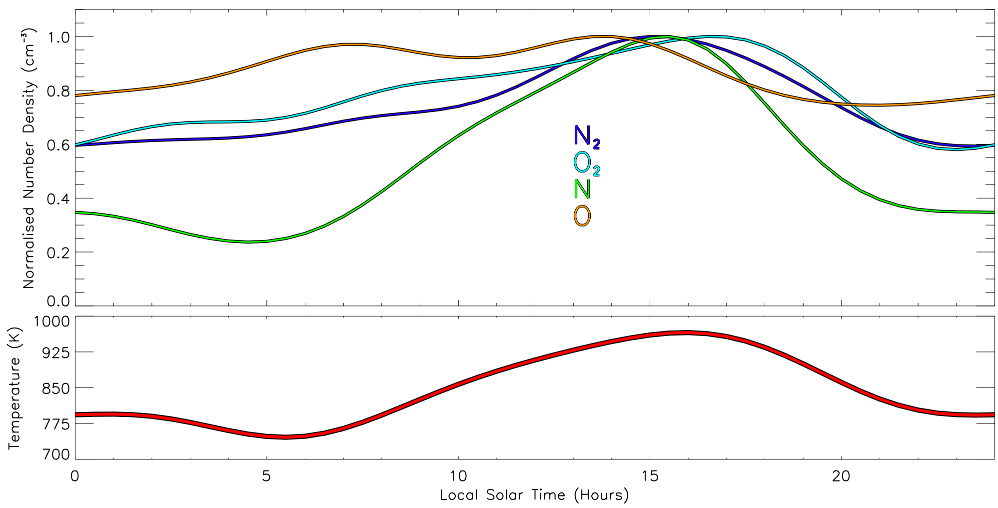


Figure 2.2.3: Diurnal density variations for the major thermospheric species, at an altitude of 240 km, from NRLMSISE-00. Number densities of each species are normalised to the peak number density of that species over 24 hours.

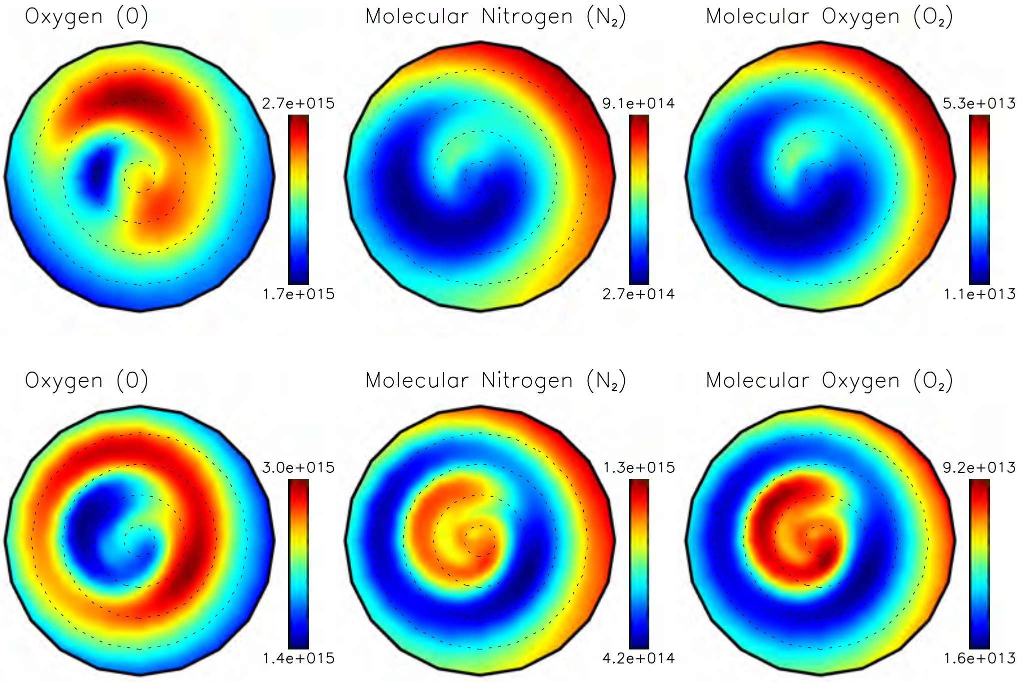


Figure 2.2.4: Southern hemisphere number densities (m^{-3}) of the three major thermospheric species (O, N_2 and O_2) at 240 km as output by the Coupled Thermosphere/Ionosphere Plasmasphere (CTIP) Model. Local time increases clockwise in each panel, with local noon at the top of the page. The highest (geographic) latitude circle is at 80°, and latitudes are indicated in 20° increments. Top row corresponds to day number 110, with quiet geomagnetic conditions (Hemispheric Power Index = 3.8), while the bottom row is for day number 143, under active geomagnetic conditions (Hemispheric Power Index = 8.1). For both rows the solar $F_{10.7}$ index was ~ 70.

the lighter species (such as He, O, and Ar) are transported to the nightside (Mayr et al., 1978), offsetting the dayside density bulge.

A parameter often used to study thermospheric composition is the O/N_2 ratio. In a modelling study Rishbeth and Müller-Wodarg (1999) found that the global circulation of the atmosphere, with its summer-to-winter flow of air, results in regions of elevated O/N_2 ratio at F-region heights (near the daytime F2-layer peak), coincident with regions of downwelling air. These large O/N_2 ratios give rise to larger F2-Layer electron densities in the daytime winter hemisphere, just equatorward of the auroral oval, referred to as the F2-Layer winter (or seasonal) anomaly. In a later study, Rishbeth et al. (2004) examined variations of the p-parameter (similar to the O/N_2 ratio, but applicable over a range of altitudes) and found that at high-latitudes, magnetic activity results in a lower p-parameter (and thus a more molecular atmosphere) in the summer, presumably due to upwelling caused by Joule and particle heating in the auroral zones (Section 2.3). The summer-to-winter difference in solar heating is found to affect the thermospheric composition more than the day-to-night cycle of heating (and the consequent up/down atmospheric motion), leading to a less molecular thermosphere near winter solstice, and magnetic quiet-day variations in the F2-layer that are well explained by the global thermospheric circulation (Rishbeth et al., 2000).

Longer-period variations also occur. Since temperature can vary by almost a factor of 2 between solar minimum and maximum, scale heights are much reduced at solar minimum, and the molecular component of the thermosphere sits well below the F-region of the ionosphere, while at solar maximum molecular constituents are present in much greater concentrations in the F-region (Smith et al., 1988).

Schlegel et al. (2005) reported observations from the Challenging Minisatellite Payload (CHAMP) satellite which showed neutral density structures (density maxima and minima) in the upper thermosphere (at altitudes near 400 km). These structures had widths of between a few hundred and 2000 km, and amplitudes reaching 50% of the ambient neutral density. Maxima were observed to cluster around $-75°$ geomagnetic latitude, with minima closer to the poles. Similar observations from the CHAMP satellite were reported by Lühr et al. (2004), who observed density enhancements of almost a factor of 2 associated with intense fine-scale field-aligned currents in the dayside cusp region. Joule heating in the E-region was proposed as the driving force for upwelling in this region, resulting in the observed density enhancements at higher altitudes. Crowley et al. (1996) obtained diurnally reproducible simulations using the National Center for Atmospheric Research's Thermosphere Ionosphere Electrodynamics General Circulation Model (NCAR-TIEGCM) for a range of magnetic activity levels which demonstrated the existence of such structures (called 'cells' by these authors) at high (magnetic) latitudes. These cells, which were present from about 120 km extending into the upper thermosphere, had diameters of 1000 to 2000 km, with density enhancements or depletions of up to 30% of the hemispheric average. Similar cells were reported by Schoendorf et al. (1996b,a). At high levels of magnetic activity, for example, four cells were observed, low density at (magnetic) dawn and dusk, and high density near noon and midnight.

Figure 2.2.4 shows the southern hemisphere spatial density distribution of the major thermospheric species modelled by the Coupled Thermosphere-Ionosphere Plasmasphere (CTIP) model. The top row corresponds to quiet geomagnetic conditions (day number 110) and the bottom row to active geomagnetic conditions (on day number 143). Both model runs are for low solar activity ($F_{10.7} \sim 70$). Under active conditions, the effect of the high-latitude circulation (Section 2.4.1) on N_2 and O_2 number densities is particularly clear in the model.

On a smaller scale, local vertical winds can transport species upwards from lower regions where number densities are higher, affecting the local composition. This in turn can affect chemical reaction rates and plasma densities. As well as increasing the rates of collisional reactions (Equations 2.2.16 - 2.2.21), precipitating auroral particles deposit heat into the atmosphere, which can again lead to vertical winds. Variations in high-latitude composition may be transported to lower-latitudes by thermospheric meridional winds. Pallamraju et al. (2004) explained an anomalously large increase in airglow emissions at a low geomagnetic latitude site as resulting from increased neutral densities brought in by travelling atmospheric disturbances (TAD's) originating at high latitudes during a magnetic storm.

2.3 Energetics

In a global, time-averaged sense the temperature profile of the thermosphere (Figure 2.1.1) is the result of a balance between the absorption of solar radiation and the downward molecular conduction of heat (Killeen, 1987). At high-latitudes Earth's magnetic field acts to couple the ionosphere/thermosphere system to processes occurring within the magnetosphere. Electric fields mapped down from the magnetosphere drive current systems which can exchange energy and momentum with the neutral atmosphere through collisions. Energetic magnetospheric particles accelerated along magnetic field lines likewise deposit energy as they are decelerated by collisions with neutral species. The total power input to the global ionosphere/thermosphere system through these mechanisms is in the range of tens to hundreds of gigawatts (GW).

Using three empirical models, Knipp et al. (2004) estimated the relative contributions to the global heating budget from solar EUV, Joule and (electron) particle heating. They found that on average, solar heating provided 464 GW, while Joule and particle heating delivered 95 GW and 36 GW respectively. These numbers provide a good reference for what follows. However, it should be kept in mind that during the top 15 heating events studied (op. cit.), power from geomagnetic sources accounted for two-thirds of the total power budget, and that during each of these events Joule heating alone exceeded that due to solar heating.

2.3.1 Solar Heating

Solar radiation represents the dominant, large-scale, time-averaged heat source for the thermosphere. The thermal energy deposited per unit volume at altitude z is described by (Whitten and Poppoff, 1971):

$$q_T(z) = \sum_i \int F_\lambda(\lambda, z)\varepsilon_i(z)\sigma_i(\lambda)n_i(z)\, d\lambda \tag{2.3.1}$$

where $F(\lambda, z)$ is the radiation flux per unit wavelength at height z, n_i is the height-dependent concentration of the ith species, $\sigma_i(\lambda)$ is the absorption cross-section of species i at wavelength λ, and ε_i is the fraction of the absorbed energy which is eventually released as heat (the heating *efficiency*). The primary wavelength regions which are absorbed in the thermosphere include the X-ray, EUV and UV bands (see Section 2.2).

The processes responsible for converting this energy into heat depend strongly on altitude (Torr et al., 1980; Torr and Torr, 1982). The volume heating rate due to absorption of radiation in the Schumann-Runge continuum maximises around 120 km (Gordiets et al., 1982). This energy is initially channelled into the dissociation of O_2 as in Equation 2.2.7. As well as imparting kinetic energy to the products, the excess energy of the reaction goes into exciting one of the dissociated atoms to the (^1D) state, which at these lower altitudes is turned into heat via collisional deactivation (with an efficiency of around 33% (Killeen, 1987)). The dissociation energy (5.1 eV) is lost until the atomic oxygen recombines below 100 km (Section 2.2). Roble et al. (1987) found that this process of O_2 dissociation is the dominant source of heat in the altitude range 100-150 km. From around 150-250 km, heating is mainly due to exothermic neutral-neutral and ion-neutral chemical reactions.

Solar heating depends of course on the energy output of the Sun, and hence on the solar cycle. Hernandez and Roble (1995) examined thermospheric winds and temperatures observed above Fritz Peak Observatory (39.9°N, 105.5°W) with a Fabry-Perot spectrometer (FPS) under both solar minimum and solar maximum conditions. For the solar maximum data they presented, average solar flux at 10.7 cm (as measured by the $F_{10.7}$ index) was 235×10^{-22} W.m^{-2}.Hz^{-1}. Temperatures were observed to decrease from around 1300-1400 K at sunset to 1000-1100 K in the early morning. For solar minimum conditions, average $F_{10.7}$ was down to 80×10^{-22} W.m^{-2}.Hz^{-1}. Temperatures on these nights were in the range 700-900 K at sunset, and showed only a small decrease (approximately 100 K) by the early morning. Temperatures observed during solar maximum were therefore not only higher in an absolute sense, by around 500 K, but also showed a greater diurnal variation than during solar minimum.

2.3.2 Joule Heating

Joule heating, otherwise known as frictional or Ohmic heating, is a high-latitude heat source caused by a differential velocity between ions and neutrals (Cole, 1962). Velocity differences arise due to the different nature of the forces acting on the plasma and the neutral gas. The ionospheric plasma responds to electric fields mapped down from the magnetosphere along the equipotential magnetic field lines, causing the ions and electrons to drift in the $\mathbf{E}\times\mathbf{B}$ direction. When the IMF B_z component is directed southward, this drift resembles a twin-cell convection pattern (Section 1.2.1), with antisunward flow over the polar caps and a sunward return flow at auroral latitudes. While the F-region ionosphere can respond very quickly (on time-scales of 2-10 minutes) to changes in IMF (Etemadi et al., 1988; Todd et al., 1988; Khan and Cowley, 1999; Ruohoniemi et al., 2002), the large-scale neutral wind response time (due to collisions between the charged and neutral species) to the this new forcing is generally expected to be in the range of $1/2$ - $6 1/2$ hours (Kosch et al., 2001). In addition, field-aligned currents directed into and out of the ionosphere are at least partially closed by Pedersen currents flowing in the auroral oval (perpendicular to the oval), which dissipate electric field energy through collisions. At F-region altitudes there are additional forces acting on the neutral species in addition to momentum transfer from the ions, such as the pressure-gradient force due to solar heating, and the Coriolis force (see Section 2.4). Thus the equilibrium velocity of the charged and neutral species can be quite different, and there will be a transfer of momentum and heat resulting from collisions between them. Convection electric fields and neutral winds are a constant feature of the high latitude thermosphere, and thus Joule heating should generally always be present (Rees et al., 1983b).

For a two species atmosphere (a single ion, i, and neutral species, n), and assuming the temperature of each species to be equal, the (neutral) volume heating rate per unit mass due to collisions is given by (Brekke, 1997):

$$\begin{aligned}\frac{\partial q_J}{\partial t} &= \frac{\nu_{ni}}{m_n + m_i} m_n (\mathbf{u}_i - \mathbf{u}_n)^2 \\ &= \frac{1}{2}\nu_{ni}(\mathbf{u}_i - \mathbf{u}_n)^2\end{aligned} \tag{2.3.2}$$

where $\mathbf{u}$ is the velocity, ν_{ni} is the ion-neutral collision frequency, and it has been assumed that the neutral mass is equal to the ion mass ($m_n = m_i$). The height-integrated rate of Joule heat dissipation is also often expressed in terms of the electric field and current density vectors (Brekke, 1997):

$$\frac{\partial Q_J}{\partial t} = \int_z \mathbf{j} \cdot \mathbf{E}\ dz \tag{2.3.3}$$

where $\mathbf{j}$ is the current density vector and $\mathbf{E}$ is the electric field. Above an altitude of around 100 km, the electric field component parallel to the magnetic field $\mathbf{B}$ is 'shorted out' by the extremely high conductivity along the magnetic field lines (since these field lines are effectively equipotential). Thus the electric field lies almost completely in the plane perpendicular to the magnetic field, and is labelled $\mathbf{E}_\perp$. By introducing σ_P, the Pedersen conductivity parallel to $\mathbf{E}_\perp$, and taking into account the neutral wind flow, Equation (2.3.3) becomes:

$$\begin{aligned}\frac{\partial Q_J}{\partial t} &= \int_z \sigma_P(z)(\mathbf{E}'_\perp)^2\ dz \\ &= \int_z \sigma_P(z)(\mathbf{E}_\perp + \mathbf{u}_n \times \mathbf{B})^2\ dz\end{aligned} \tag{2.3.4}$$

where the primed quantity is in the frame of reference of the neutral wind, and $\mathbf{u}_n$ is the (altitude-dependent) neutral wind vector. As Vasyliunas and Song (2005) point out, this equation is more correctly described as representing the frictional heating due to relative motion between the plasma and the neutrals, since it is not Ohmic heating in the strictest sense. A simplification which is often made when calculating the Joule heating rate is to set the neutral wind vector equal to zero (for all altitudes), in which case Equation 2.3.4 simplifies to:

$$\begin{aligned}\frac{\partial Q_J}{\partial t} &= \int_z \sigma_P(z)\mathbf{E}_\perp^2\ dz \\ &= \Sigma_P \mathbf{E}_\perp^2\end{aligned} \tag{2.3.5}$$

where Σ_P is the height-integrated Pedersen conductivity. Energy deposition per unit volume due to Joule heating maximises around 120 km altitude, near the maximum of the current density $\mathbf{j}$ (but possibly as much as 18 km above it, see for example Brekke and Rino (1978)), however the neutral gas heating rate (in units of K.s^{-1}) is maximum at around 400 km (Deng and Ridley, 2007), due to the approximately exponential decrease in neutral density with altitude. Thayer (1998) found that including the neutral wind in the calculation of Joule heating could at times increase the heating rate by 400%, or reduce it by 40%, relative to the simplified calculation of Equation 2.3.5. Often the effect of the neutral wind was to confine the region of Joule heating to a narrower altitude range. Aruliah et al. (2005), using a tristatic arrangement of Fabry-Perot interferometers collocated with tri-static radar measurements, reported a 320% increase in the rate of Joule heating when a 1-min average velocity was used in the calculation compared with a 15-min average, and on one night the contribution of the neutral wind dynamo to Joule heating was around 29%, further highlighting the important role that the neutral atmosphere can play in determining the rate of Joule heating. The heating efficiency of Joule energy deposition is nearly 100% (Thayer and Semeter, 2004), in comparison with a solar EUV heating efficiency of approximately 50%. Thus, while solar energy deposition is dominant in a time and hemispherically averaged sense, the higher heating efficiency of Joule dissipation can make it the hemispherically dominant source of heating during periods of elevated geomagnetic activity (Knipp et al., 2004).

In a comprehensive analysis of around 25,000 orbits of the Atmosphere Explorer C (AE-C) satellite, Foster et al. (1983) calculated hemispheric Joule heating rates of 35 GW during undisturbed periods, rising to 80-100 GW for disturbed periods ($3 < \mathrm{Kp} < 6$). Heating was found to be most intense along the dayside auroral dawn and dusk sectors, resembling an inverted horseshoe with reduced heating in the (magnetic) midnight sector. Approximately 25 GW was found to be associated with the dayside cleft region. Olsson et al. (2004) used six months of data from the Astrid-2 satellite to perform a similar study to that of Foster et al. (1983), of which selected results are shown in Figure 2.3.1. Despite the difference in solar cycle between the two studies (Foster et al. (1983) during solar minimum, Olsson et al. (2004) during solar maximum), there was good agreement between the two studies. Olsson et al. (2004) found that approximately $1/3$ of the total Joule heating occurred in the nominal nightside auroral oval (65°-75° invariant latitude) region, and that under increased magnetic activity ($\mathrm{Kp} > 2$) the fraction was closer to $1/4$, likely due to the expansion of the auroral oval to lower latitudes under such conditions.

Østgaard et al. (2002a) found that during substorms Joule heating accounted for 56% of the total time-integrated energy dissipation, with the remaining energy being channelled into the ring current and particle heating. Tanskanen et al. (2002) examined substorm data from two years, 1997 and 1999, from which they derived Joule heating rates. During 1997, which was a more active year than 1999, Joule heating accounted for the dissipation of approximately one third of solar wind energy input. During 1999, this value decreased to around one quarter. These authors also found that Joule heat dissipation was relatively more important during isolated substorms, as compared with storm-time substorms (Section 1.2.3).

Codrescu et al. (1995) suggested that many studies actually underestimated Joule heating by not properly accounting for small-scale spatial or temporal variations in the electric field strength. In a simplified argument they showed that including small-scale variability in the electric field could increase derived Joule heating rates by around 33%. Rodger et al. (2001) estimated Joule heating from EISCAT Incoherent Scatter Radar data and found that using hourly-averaged electric field data led to an underestimation of the

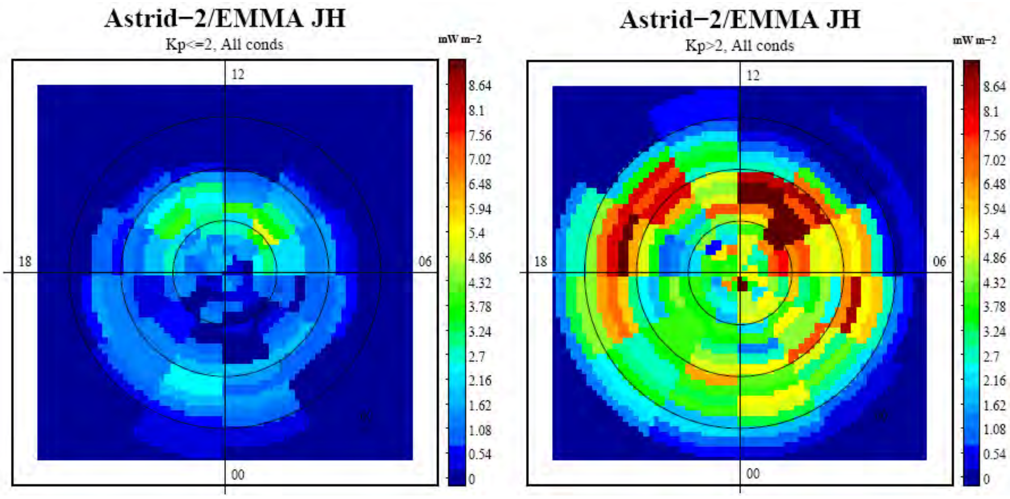


Figure 2.3.1: Average global Joule heating derived from Astrid-2 satellite measurements of the magnetic and electric field, under conditions of Kp ≤ 2 (left panel) and Kp > 2 (right panel). The central circle marks 80° invariant latitude (ILAT), with outer circles at 10° increments. From Olsson et al. (2004).

heating rate by approximately 20%, when compared to values derived using 6-minute data. Deng and Ridley (2007) reported that increasing the latitudinal resolution of the General Ionosphere Thermosphere Model (GITM) from 5° to 1.25° increased the derived neutral gas heating rate at 200 km altitude by 20%, further highlighting the importance of small-scale variability.

Studies have shown that a nearly linear relationship exists between hemispherically integrated Joule heating and the auroral electrojet (AE) index. Proportionality constants ranging from 0.05 to 0.54 GW.nT^{-1} (Østgaard et al., 2002a) have been derived. From a study of three substorm events, Aksnes et al. (2004) gave proportionality factors of between 0.13 and 0.23 GW.nT^{-1}, after the effects of energetic electrons were taken into account. As well as AE, studies have related Joule heating to the Kp index (Foster et al., 1983; Kosch and Nielsen, 1995; Cosgrove and Thayer, 2006; Olsson et al., 2004), Polar cap index (a proxy for the polar cap electric field) (Chun et al., 1999), epsilon (ϵ) parameter[1] and solar wind kinetic energy flux (Olsson et al., 2004). In relation to the IMF, Kosch and Nielsen (1995) showed that the average rate of Joule heating increased with increasingly negative B_z.

2.3.3 Particle Heating

Magnetospheric particles moving along magnetic field lines can enter the upper atmosphere at auroral latitudes. These energetic particles (electrons, protons, and to a lesser extent, oxygen ions) deposit their energy in this region through collisions with neutral species.

[1]The epsilon parameter is given by (Akasofu, 1981):

$$\epsilon = \frac{4\pi}{\mu_0} v B^2 L^2 \sin^4\left(\frac{\theta}{2}\right)$$

where v is solar wind speed, B is the IMF magnitude, L is an empirically determined length-scale of the magnetosphere and θ is the IMF clock angle.

Between 100 and 200 km (where most of the auroral energy is deposited) approximately 50% of this energy goes into heating the neutral gas, however this efficiency decreases rapidly above 200 km altitude to less than 10% by 400 km (Rees et al., 1983b). Particle heating is much more variable and spatially localised than, for example, Joule heating, and may be the principal cause of sudden composition and temperature changes in the non-sunlit polar thermosphere (Rees et al., 1983b). By ionizing neutral thermospheric species, precipitating particles can influence the rate of Joule heating by modifying the conductivity (Palmroth et al., 2006).

Baker et al. (2004) combined Global Ultraviolet Imager (GUVI) data with Super Dual Auroral Radar Network (SuperDARN) electric field measurements to produce height-integrated, night-side Joule and particle heating rates under quiet to moderately active geomagnetic conditions. They found a complimentary relationship between the two heat sources due to the often observed anti-correlation between electric field strength $|\mathbf{E}|$ and the height-integrated Pedersen conductance (which is a function of auroral brightness). Joule heating was found to be strongest in the evening sector auroral zone, while particle heating was seen to be dominant in the morning, collocated with bright aurora.

Particle heating is often considered to be a less important heat source than Joule heating at high-latitudes, however Aksnes et al. (2004) have shown that by considering the effects of energetic electrons (with energies from 20 keV up to approximately 100 keV), the ratio of Joule to particle heating can be reduced by more than 25%. During a severe storm Joule heating was found to contribute 1.67 times the hemispheric integrated energy flux due to precipitating particles, while at other times the ratio was 1.04 and 0.955, suggesting a much more important contribution by particle precipitation during less active periods than is often assumed (Aksnes et al., 2004). This conclusion is supported by Wilson et al. (2006), who found that although Joule heating is clearly dominant during large storms, during smaller storms the energy input is approximately equally divided between Joule and particle heating. Chun et al. (2002) studied the relationship between Joule heating rates derived by the Assimilative Mapping of Ionospheric Electrodynamics (AMIE) technique and the polar cap index (as an indicator of geomagnetic activity in the polar cap) and concluded that during active times (positive polar cap index) Joule heating was approximately 4 times greater (globally) than that due to energetic particles, while at quiet times (near-zero polar cap index) the two heat sources were equal contributors.

Østgaard et al. (2002b) also suggested that many studies underestimate the energy deposition due to precipitating electrons, and proposed a non-linear relationship between the rate of particle energy dissipation U_A (GW) and the (real-time, or quick-look, QL) auroral AL_{QL} and AE_{QL} indices, of the form:

$$U_A = 4.6\ \mathrm{GW}\ (\mathrm{nT})^{-\frac{1}{2}} AE_{QL}^{\frac{1}{2}} - 23\ \mathrm{GW}$$
$$U_A = 4.4\ \mathrm{GW}\ (\mathrm{nT})^{-\frac{1}{2}} AL_{QL}^{\frac{1}{2}} - 7.6\ \mathrm{GW}$$

By time-integrating these relations during 4 days which showed substorm activity they were able to estimate the energy deposition rate to within ±20% of the energy derived from satellite data (PIXIE and Polar-UVI).

2.3.4 Dynamical Heating

Tides: The diurnal variation in atmospheric heating due to solar radiation gives rise to atmospheric (thermal) 'tides', periodic global scale oscillations of atmospheric pressure, temperature, density and wind, at harmonics of the solar day (Müller-Wodarg and Aylward, 1998). Heating induced by absorption of Schumann-Runge continuum radiation by O_2 at heights in the range 100-150 km (Section 2.3.1) forces the thermosphere's in-situ diurnal tide, which drives the most basic wind circulation at that altitude.

Here the focus is on tides generated at lower altitudes and which can propagate upward, dissipating their energy at altitudes where molecular viscosity and thermal conduction become appreciable. These tides originate predominantly below about 70 km due to the absorption of solar radiation by O_3 and H_2O (Forbes and Garrett, 1979) and latent heat release of water vapor (Forbes et al., 1997). To an observer on the ground, these oscillations appear to move westward, following the apparent motion of the Sun, and are therefore known as 'migrating' tides. As these tidal oscillations propagate upwards, their amplitudes increase in order to conserve energy in a medium whose density decreases approximately exponentially with height. This exponential increase in amplitude is limited by diffusive dissipation, and energy is then deposited at altitudes of 100-170 km (Forbes and Garrett, 1979).

The strongest tides observed in the upper atmosphere are the diurnal and semidiurnal tides. Groves and Forbes (1984) estimated that the (globally averaged) specific heating rate due to dissipation of solar diurnal and semidiurnal tides in the altitude range 100-400 km to be on the order of 50-100 mW.kg^{-1}. At high latitudes this heating rate is comparable to that due to solar heating (Volland, 1988). Heating per unit volume maximises at altitudes around 108 km at these high latitudes, and is due mainly to the semidiurnal component; at low latitudes, heating is due to the diurnal component, and maximises around 80 km (Groves and Forbes, 1984).

Gravity Waves: Another important energy source in the upper atmosphere is due to the dissipation of gravity waves (see Section 2.4.4). Viscosity, thermal conduction, radiative heat exchange and ion-drag are all processes which can remove energy from gravity waves (Yeh and Liu, 1974), and thus convert wave energy to heat, although the rate of local heating can be offset by turbulent cooling associated with wave breaking (Gavrilov and Roble, 1994). Some upward-propagating gravity waves can reach F-region altitudes (Vadas and Fritts, 2005), where a heating rate on the order of 10^{-10} W.m^{-3} has been inferred from observations of ionospheric disturbances (Testud, 1970). Klostermeyer (1973) derived similar values for the heat input per unit volume using a model atmosphere and satellite derived neutral density perturbations, which for the modelled wave properties (period of 0.5 hours and horizontal wavelength 600 km) resulted in altitude-dependent (irreversible) temperature increases of between 20-50 K after 2 hours. The energy deposition due to gravity wave dissipation therefore represents a significant contribution to the thermospheric heat budget at thermospheric altitudes, considering that the solar heating rate is on the order of 4.5×10^{-10} W.m^{-3} (Yeh and Liu, 1974).

In addition to energy deposited at thermospheric altitudes by gravity waves propagating upwards from lower altitudes, predominantly horizontally propagating gravity waves generated in-situ can redistribute energy and momentum within the thermosphere. Thermospheric sources of gravity waves are most likely to be auroral processes, such as auroral electrojet perturbations and atmospheric responses to particle precipitation (Hunsucker, 1982), and possibly moving sources (such as periodic motion of the auroral oval, see for example Innis et al. (1996)). These types of waves are important, since at times of increased geomagnetic activity (for example during substorms) the energy deposited into the auroral zone by Joule and particle heating is often greater than the hemispherically averaged solar heating rate, and gravity waves are one mechanism by which this energy can be transported to lower latitudes (another mechanism is meridional circulation).

2.3.5 Cooling

Above approximately 180 km, downward molecular conduction is the primary mechanism that balances heat deposition (Roble and Emery, 1983). The flux of thermal energy per unit time and unit mass is given by:

$$\Phi = -\frac{\kappa}{\rho}\nabla T \tag{2.3.6}$$

where κ is called the thermal conductivity and ∇T is the temperature gradient. The rate of local cooling/heating due to thermal conduction is governed by the divergence/convergence of this heat flux, and can thus be expressed (for spatially uniform ρ and κ) as:

$$\begin{aligned}\frac{\partial Q_C}{\partial t} &= \nabla \cdot \Phi \\ &= -\frac{\kappa}{\rho}\nabla^2 T\end{aligned} \tag{2.3.7}$$

In the altitude range 90-180 km, radiative cooling in the infrared is the main energy loss mechanism (Gordiets et al., 1982). Of this, most loss is due to the 15-μ band of CO_2 (90-120 km) and the 5.3-μ band of NO (130-180 km). Of these, radiation in the 5.3-μ band is relatively more important at auroral latitudes as there is a well established link between increased NO concentration and auroral activity (Baker et al., 2001; Barth et al., 2003; Dobbin et al., 2006a,b). Maeda et al. (1989) modelled the zonally averaged thermospheric response to auroral activity, and concluded that the enhancement of NO concentration due to auroral activity was able to significantly damp the aurorally driven temperature increase and shorten the temperature relaxation time. Similarly, during the recovery phase of a geomagnetic storm, cooling in the lower thermosphere was found to be dominated by NO (Maeda et al., 1992).

2.4 Dynamics

2.4.1 Horizontal Winds

Winds in the upper atmosphere are driven primarily by the pressure gradient established by solar heating, centred on the subsolar point. The 'pressure-bulge', which maximises at around 16:00 hours local solar time (LST), drives a circulation away from the dayside to the nightside, westward in the morning (the dawn sector) and eastward in the afternoon (the dusk sector), and directed away from the Sun across the polar caps at higher latitudes. At low and mid latitudes, the presence of ions which are confined to move along the geomagnetic field lines acts to oppose this pressure-driven motion of the neutral gas predominantly in the east-west direction. Figure 2.4.1 shows this global scale flow superimposed over a temperature map output by the NRLMSISE-00 atmospheric model. At thermospheric heights, the Coriolis force becomes less important, and winds tend to flow perpendicular to the pressure contours, unlike winds in the lower atmosphere, where Coriolis forcing causes air parcels to circulate in opposite directions around high and low pressure regions, following the isobars. The term diurnal tide is used to refer to the daily variation in meridional (north-south, positive northward) and zonal (east-west, positive eastward) wind observed by a station moving underneath this global-scale circulation of the neutral gas.

Emmert et al. (2006a,b) analysed Fabry-Perot interferometer observations of thermospheric neutral winds from seven stations covering geographic latitudes from 90°S to 76.5°N. Under quiet geomagnetic conditions, they observed higher meridional wind speeds at high latitudes, maximising at local midnight (with wind flow directed toward the equator) and midday (directed toward the polar cap). At lower latitudes the diurnal variation in meridional wind had a smaller amplitude. Zonal wind speeds were seen to exhibit less

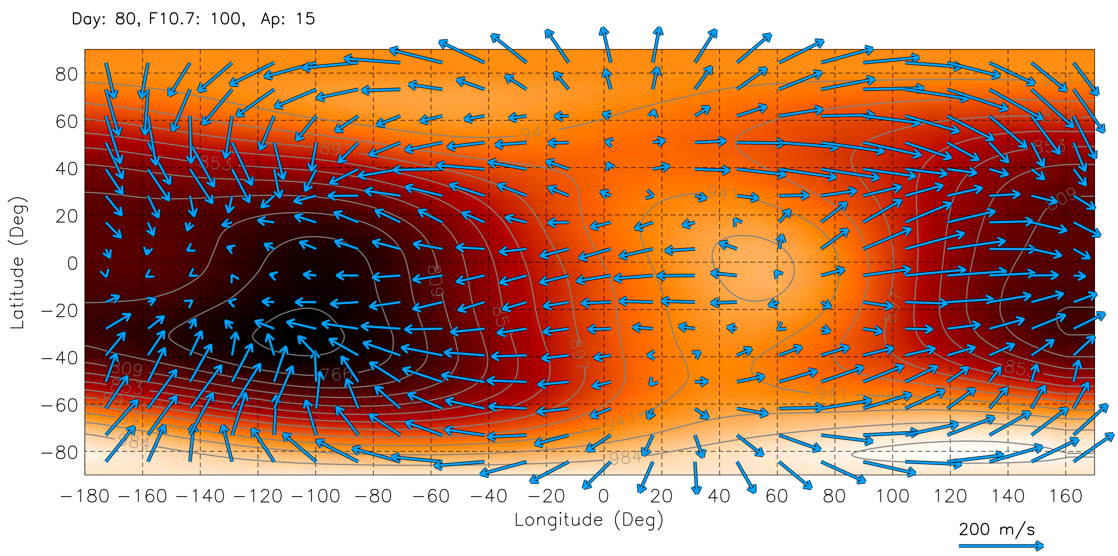


Figure 2.4.1: NRLMSISE-00 global temperatures and isothermal contours, with wind vectors predicted by HWM93 superimposed at an altitude of 240 km. Important model inputs are listed at top-left, vector wind scale is shown at bottom-right.

latitudinal variability than the meridional winds (see Figure 2.4.2). Since the primary driver of the global-scale circulation is solar heating, wind magnitudes would be expected to increase with increasing solar EUV irradiance. This effect was found to saturate for values of $F_{10.7} \simeq 150 \times 10^{-22}$ W.m^{-2}.Hz^{-1}, which at low latitudes may be explained by the associated increase in ion-drag as plasma densities increased with the higher levels of solar irradiance (Emmert et al., 2006a).

At high-latitudes the solar pressure-gradient driven wind flow can be significantly modified by the effects of ion-drag (Killeen and Roble, 1988), as charged particles drift under the combined effect of the convection electric field (mapped down from the magnetosphere) and the geomagnetic field. Neutral particles acquire momentum through collisions with the drifting ions, and as a result the neutral wind pattern begins to follow that of the ions drifting in the $\mathbf{E} \times \mathbf{B}$ direction parallel to the contours of constant electric potential. As mentioned in Section 2.3.2, the time taken for the neutral wind speed to approach that of the drifting ions (the 'e-folding time') is generally between 1/2 - 6 1/2 hours (Kosch et al., 2001). Once established, the large inertia of the neutral atmosphere can help to maintain this convection pattern in the absence of magnetospheric/ionospheric driving, a mechanism called the "flywheel" effect (Lyons et al., 1985).

It is to be expected therefore that factors affecting the high-latitude ionospheric con-

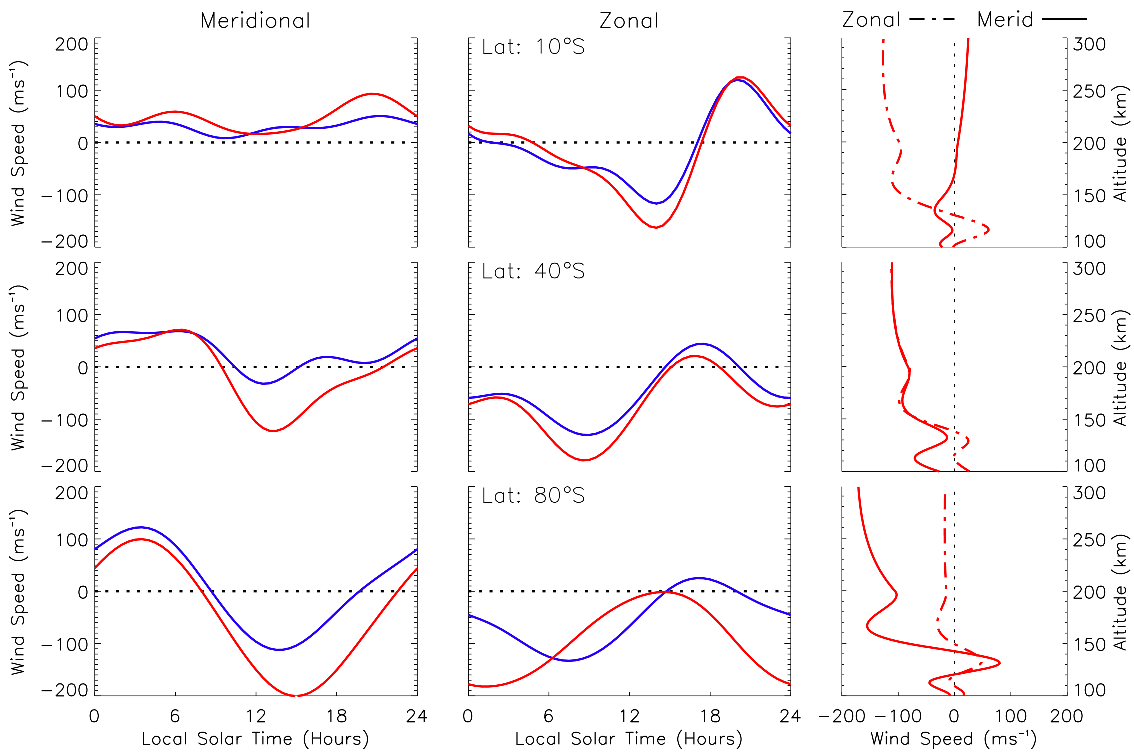


Figure 2.4.2: First two columns from the left show the meridional and zonal components of the HWM93 daily wind variation at an altitude of 240 km and for three latitudes, for low activity (ap = 10, $F_{10.7}$ = 90 - blue line) and high activity (ap = 60, $F_{10.7}$ = 200 - red line). The right-hand column shows the altitude variation of the zonal (dash-dot line) and meridional (solid line) components of the active-period winds. Each row corresponds to a different geographic latitude, indicated in the middle column.

vection will be correlated with the neutral wind response. Under conditions of southward-directed interplanetary magnetic field (IMF), the ionospheric flow resembles a twin-cell circulation (see Figure 2.4.3), referred to as a convection pattern (see for example Khan and Cowley (1999); Baker et al. (2007); Haaland et al. (2007)). In the dayside cleft region, the pressure-gradient and ion-drag forces both act to accelerate neutral air parcels in the antisunward direction, over the polar cap (sometimes referred to as the cross-polar 'jet'). A similar situation occurs around magnetic midnight, where the forces combine to drive air parcels toward the magnetic equator. A return flow of ions at auroral latitudes opposes the pressure-gradient, driving the neutral atmosphere eastward (westward) in the dawn (dusk) sector. The dusk sector sunward flow is more commonly observed than the dawn sector (McCormac and Smith, 1984; Killeen et al., 1988; Lühr et al., 2007), a result which has been attributed to the Coriolis force opposing the cyclonic vorticity on the dawn side while reinforcing the anticyclonic flow on the dusk side (Gundlach et al., 1988; Thayer and Killeen, 1993). Under conditions of strong northward IMF B_z, a multicellular convection pattern consisting of three or even four cells can emerge (Förster et al., 2008a), which can lead to a weakening or even a reversal of the antisunward flow across the polar cap. Under very strongly northward IMF ($B_z > 5$ nT), Niciejewski et al. (1994) observed such a flow reversal, with sunward neutral winds at the highest latitudes near local solar noon.

Many studies have demonstrated the effect of the B_y (east-west) component of the IMF on convection pattern morphology (Jones et al., 1987; Cannon et al., 1991; Shue and Weimer, 1994; Killeen et al., 1995). McCormac et al. (1985), using northern hemisphere neutral wind measurements from the Dynamics Explorer-2 (DE-2) satellite, found that the region of most rapid antisunward flow over the polar cap shifted from the dawn side to the dusk side as the IMF B_y component changed from positive to negative. Under conditions of B_y negative, this study found that the region of neutral gas entry into the polar cap became biased towards the dusk side of the noon-midnight (magnetic) meridian, and that the velocities associated with the dawn side convection cell were enhanced.

Förster et al. (2008b) compared accelerometer data from around 5600 transpolar overflights of the Challenging Minisatellite Payload (CHAMP) satellite during 2003 with electron drifts inferred from the Electron Drift Instruments (EDI) aboard the Cluster satellite during the same period (see Figure 2.4.3). They found for example that ion-neutral momentum coupling is asymmetric between the northern and southern hemispheres with respect to IMF B_y. Greater cross-polar cap winds were observed during B_y- (negative) conditions in the northern hemisphere, and B_y+ (positive) in the southern hemisphere. Wind magnitudes in both hemispheres were found to be higher during B_z southward (negative) conditions, presumably due to the greater coupling between the magnetosphere and solar wind under this IMF configuration. A further asymmetry exists between convection flows derived in the northern and summer hemispheres due to the greater offset of the south geomagnetic pole relative to the south geographic pole (Lühr et al., 2007). This offset is almost twice as large as that in the northern hemisphere (Smith et al., 1988).

Conde and Dyson (1995a) examined 103 nights of Fabry-Perot spectrometer data recorded at Mawson station, Antarctica. Wind flows characteristic of both pressure-

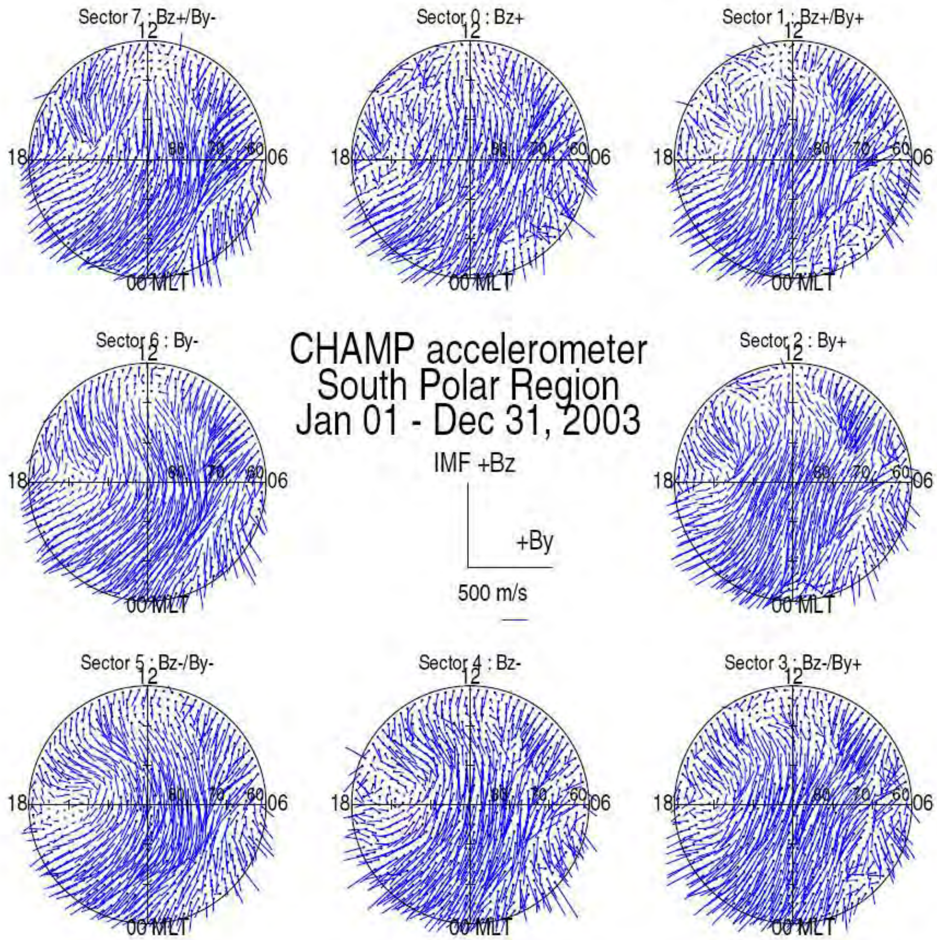


Figure 2.4.3: Southern hemisphere wind patterns for various IMF configurations relative to the GSM y-z plane. The outer circle of each plot is at 60° magnetic latitude. Plots are shown for an observer located above the north pole, looking through a transparent Earth. From Förster et al. (2008b).

gradient and convection forcing were observed, with most nights displaying signatures of both driving forces. Smith et al. (1998), comparing observations from Fabry-Perot spectrometers located at three Antarctic stations (including Mawson), found signatures of dawn and dusk circulation cells, and the cross-polar jet which slowed down rapidly in the equatorward direction.

Thayer and Killeen (1993) performed a kinematic analysis of the average neutral wind circulation at high-latitudes, by decomposing the vector wind fields into their divergent and non-divergent components. They found that under both active ($3+ \leq \mathrm{Kp} \leq 6$) and quiet ($\mathrm{Kp} \leq 3$) conditions, the non-divergent component, driven primarily by the ion-drag and Coriolis forces, made up a large percentage of the total wind field. The divergent field, driven by solar heating, Coriolis and nonconvective-ion drag forces (such as Joule heating), was found to compliment the antisunward flow over the polar cap, while inhibiting the nondivergent flow in the dawn and dusk sectors. A near cancellation of the divergent and nondivergent components of the wind field in the dawn sector explained the often poor development of sunward winds in that sector relative to the dusk sector. In an earlier work (Thayer and Killeen, 1991), these authors noted that the ratio of rotational (non-divergent) to divergent flow for active conditions was around 2:1, while for quiet conditions this ratio was more like 4:1. This may be explained by the dependence of the forcing terms on the strength of the electric field $|\mathbf{E}|$. Momentum forcing due to convecting ions (which drives a rotational flow) is proportional to $|\mathbf{E}|$, while Joule heating (driving divergent flow) is proportional to $|\mathbf{E}|^2$ (Forbes, 2007).

It is well known (Yokoyama et al., 1998; Ahn et al., 2005; Aikio et al., 2006; Wang et al., 2008a) that the auroral oval expands equatorward during magnetic substorms. Winds observed from stations situated at the transitional region between mid and polar latitudes are therefore very sensitive to geomagnetic activity, as the predominantly solar driven, antisunward flow at low levels of activity changes to that of a convection-driven flow (characteristic of auroral latitudes) at higher levels of activity (Niciejewski et al., 1996). Wang et al. (2008a) for example found that the centers of the eastward and westward electrojets can expand to 55° magnetic latitude during intense storms. However, observations of high-velocity (1400 m.s^{-1}) convection-driven east-west ion flows have been reported at magnetic latitudes as low as 30°, accompanied by neutral thermospheric winds in the range 300-640 m.s^{-1}, during two severe magnetic storms (Reddy and Mayr, 1998).

2.4.2 Vertical Winds

Vertical winds play an important role in the dynamics of the thermosphere, as these motions can oppose diffusive separation, transporting species to new altitudes at which they may no longer be in chemical equilibrium. These compositional changes will in general affect the production and loss rates of ionospheric species, through changes in the numbers of atomic and molecular species available for ionization and recombination, and hence result in variations of the peak heights of ionospheric layers. Vertical winds are also linked to horizontal motions through the need to conserve mass, and as such can drive or be driven by divergent and convergent horizontal wind flow.

On a global scale, vertical wind motions are assumed to be relatively minor, driven by solar heating of the sunlit atmosphere, which results in a diurnal expansion and contraction of the entire atmosphere (Smith, 1998). At high latitudes there is additional heating due to geomagnetic sources (Joule and particle heating), and as such a diurnal variation may be introduced through the daily movement of a high-latitude station under geomagnetically-heated regions. The magnitude of the vertical wind driven by these mechanisms should be at most a few m.s^{-1} (Smith, 1998). Conde and Dyson (1995b), for example, observed a diurnal vertical wind variation of -2.6 to 3.0 m.s^{-1} above Mawson station, Antarctica, during low levels of geomagnetic activity. A larger amplitude diurnal variation (average amplitude 40 m.s^{-1}) was observed by Smith and Hernandez (1995) at South Pole station, during the geomagnetic storm period of June 1991.

However much larger vertical winds have been reported. Spencer et al. (1982) analysed data from the Wind and Temperature Spectrometer (WATS) on board DE-2, which included in-situ measurements of vertical wind across large latitude ranges which included the north pole. Vertical winds with speeds of 100-250 m.s^{-1} were observed at auroral latitudes, at altitudes of approximately 300 km. These considerable upwellings were associated with large exospheric temperature enhancements, suggesting that localised heating at lower altitudes was a likely driver. Crickmore et al. (1991) measured strong (50 m.s^{-1}) downward winds with a Fabry-Perot spectrometer located at the Antarctic station Halley (75.5°S). These downwellings were frequently observed when the station was at the equatorward edge of the auroral oval. Innis et al. (1996) observed many instances of strong (50-100 m.s^{-1}) upwellings on the poleward side of the auroral oval at Mawson station. In a later study, Innis et al. (1999) focused on a single event, during which an upward wind of $\geq$ 100 m.s^{-1} was first observed at Davis station, and then seen above Mawson station (4° magnetically north of Davis) approximately one hour later. This upward wind was coincident with the poleward edge of the auroral oval. These authors estimated that the power required to drive this vertical wind was 3-7% of the geomagnetic power input to the southern hemisphere during the interval.

In a study of mean vertical winds at both Mawson and Davis stations, Greet et al. (2002) found a similar relationship between the measured vertical wind and a station's location relative to the auroral oval, with downward winds on the equatorward boundary of the oval, while on the poleward boundary the average vertical wind was downward in the early magnetic evening and upward around magnetic midnight (see Figure 2.4.4). Larger vertical wind amplitudes were observed on the poleward edge of the oval compared to those within the oval. Ishii (2005), in a single-day case study, reported similar observations from Poker Flat, Alaska. This study focused on the relationship between thermospheric vertical winds and the location of the auroral electrojet. Downward vertical winds were observed equatorward of the electrojet after about 00:30 MLT, while upward winds were observed poleward of it. However before 00:30 MLT the reverse situation was observed: upward winds were seen equatorward of the electrojet and downward winds observed poleward of

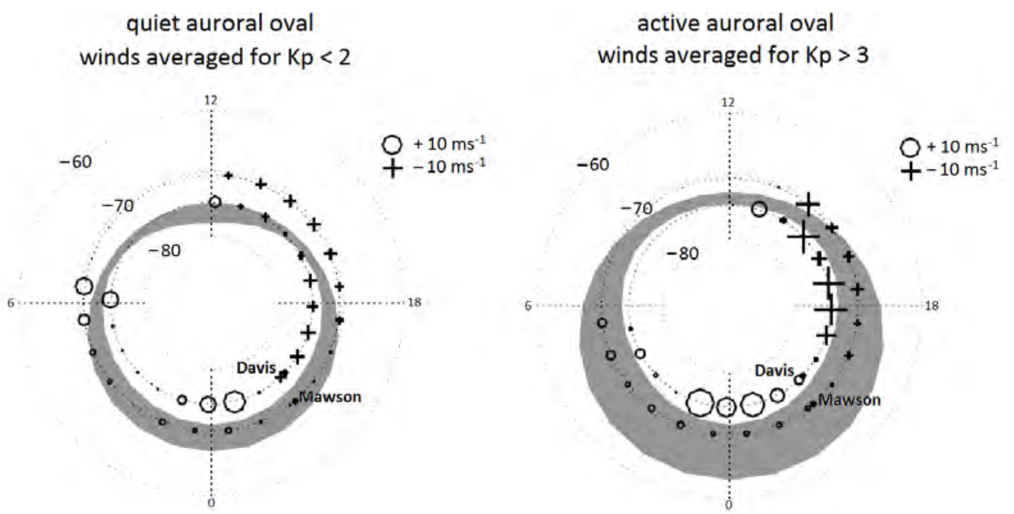


Figure 2.4.4: Locations of positive and negative mean vertical winds at Mawson and Davis stations, relative to the approximate location of the auroral oval, and for two ranges of magnetic activity. From Greet et al. (2002).

it.

Vertical winds at mid-latitudes were observed during a bistatic campaign between the Laurel Ridge and Millstone Hill observatories (Sipler et al., 1995). Average data suggested that quiet-time winds were on average downward throughout the night with speeds on the order of 10 m.s^{-1}. Although the database was small, higher speeds and greater variability was found on geomagnetically active nights. The data of Crickmore et al. (1991) were also suggestive of net downward mean vertical winds at Halley.

Simultaneous upwellings at two different altitudes have been observed by Price et al. (1995) at Poker Flat, Alaska. Observations of both E- and F-region emissions (557.7 and 630.0 nm respectively) were made on 21 nights. Upwelling events lasting between 15 and 25 minutes were observed to occur simultaneously at both altitudes on 2 nights, with peak velocities of 42 m.s^{-1} (at 138 km altitude) and 138 m.s^{-1} (240 km altitude). These upwellings were located on the poleward side of the auroral oval, in agreement with previous observations, and appeared to correlate with temperature increases of approximately 200 K measured at the lower altitude. Ishii et al. (2001), in a similar study also carried out at Poker Flat, reached the same conclusions regarding both the type (upward or downward) of vertical wind in relation to the aurora, and the simultaneous occurrence of wind events at E- and F-region altitudes. A vertically extended region of heating was postulated to explain the simultaneous vertical wind response observed at both altitudes, since in many cases these events coincided with increased auroral luminosities. These authors did however note a phase-difference in the vertical wind response at the two altitudes which was coincident with a thin, bright auroral arc passing over the observatory. In contrast Kosch et al. (2000) observed only weak coupling between E- and F-region vertical winds above Ramfjord and Skibotn, in Norway. The coupling was however observed to increase in response to greater geomagnetic activity.

In the same study Kosch et al. (2000) observed essentially no correlation between the vertical winds observed above each station at either altitude. The two stations were separated longitudinally by only 45 km. In the presence of aurora, Ishii et al. (2004) observed very good correlation between lower thermosphere vertical winds recorded by stations at similar geomagnetic latitudes which were separated by approximately 300 km, suggesting that lower thermospheric vertical winds were uniform along the arc over at least this distance. However the inter-station correlation between vertical winds in the upper thermosphere were generally low.

Innis and Conde (2002) analysed vertical winds measured by the WATS instrument on-board the DE-2 satellite, in the altitude range $\sim$ 250 to 650 km. A measure of vertical wind activity was obtained by calculating the standard deviation of the vertical wind measured within a sliding window equivalent to an along-track distance of $\sim$ 900 km. Results were separated by altitude, AE index and solar zenith angle. The clear result from this study was that the greatest vertical wind activity was observed at high invariant latitudes, largely confined to the region bounded by the nominal auroral oval. Increased activity was found at higher altitudes (above 450 km) and during periods of increased AE index, while solar zenith angle had no significant influence. The magnetic midnight-dawn sector consistently showed increased activity relative to other magnetic time sectors, in both hemispheres. Increased activity levels were interpreted as signatures of atmospheric gravity waves, since wave-like fluctuations were often observed in the WATS vertical wind time series. These authors then estimated what the gravity wave production rate would be under different levels of geomagnetic activity (as measured by the AE index). For all confidence limits presented, the probability of observing wave activity increased with increasing AE.

Vertical and horizontal motions are linked by the need to conserve mass. In their study of thermospheric winds above South Pole station (Invariant latitude 75°S), Smith and Hernandez (1995) investigated the relationship between the divergence of the horizontal wind field (estimated from the horizontal winds measured in off-zenith look directions) and the measured vertical wind. Their motivation for this was the point made by Burnside et al. (1981) that at least for small vertical wind speeds (on the order of 5 m.s^{-1}) the horizontal divergence should be approximately related to the vertical wind through the equation:

$$w = H\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right) \tag{2.4.1}$$

where H is the scale height and w, u and v are the vertical, zonal and meridional components of the wind field respectively, where the latter two components are chosen to be aligned along the x (zonal) and y (meridional) axes. This relationship would be expected to hold (in an atmosphere that is isothermal above the level of the airglow) if the time-scale for changes in pressure at the altitude of interest are long compared with the time-scales for vertical motion, the horizontal velocity components are constant with height, and the horizontal gradients in atmospheric density can be ignored. The term in brackets is simply the total horizontal divergence of the wind field.

Smith and Hernandez (1995) plotted vertical wind as a function of divergence, and obtained a linear relationship of the opposite sign to that expected by the Burnside relation (Equation 2.4.1), i.e. upward vertical winds corresponded to negative divergence and downward vertical winds corresponded to positive divergence. These authors interpreted this result as showing that divergence was the driving force for the vertical winds, and not the other way around as was proposed by Burnside et al. (1981), however recent results from a local-scale, three-dimensional neutral atmospheric model (Cooper et al., 2009) suggest that another possible reason for the discrepancy with respect to Equation 2.4.1 was due to an energy deposition height above the level of the observation height (the altitude of peak airglow emission, $\sim$ 240 km). From the slope of a regression line fitted to the upper thermospheric vertical wind/divergence data these authors calculated H to be -78.8 km, approximately twice the scale height which would be expected at the assumed altitude of 240 km.

Guo and McEwen (2003) examined data from five winters of Fabry-Perot Interferometer data recorded within the central polar cap in the northern hemisphere, and also observed an approximately linear relationship between vertical wind and divergence, however in this case the sense of the data suggested that vertical winds were driving horizontal divergence, as would be expected from the Burnside relation. The magnitude of the observed divergence was found to be much more variable at higher vertical wind speeds ($|w| > 50$ m.s^{-1}). These data also showed a marked asymmetry between divergence observed during upward and downward wind events, with a saturation of divergence associated with strong upward wind events, in contrast to the approximately linear relationship between downward winds and negative divergence.

Ishii et al. (1999) found evidence for a more complex relationship between horizontal divergence and vertical wind, from observations made at Ramfjord, Norway, by both a narrow-field and an all-sky imaging Fabry-Perot Interferometer. These authors examined the way in which the correlation between vertical wind and divergence varied as a function of local solar time (LST) and also of time-delay, at both lower and upper thermospheric heights. In the upper thermosphere, the correlation was strongly negative between approximately 19:30 and 21:30 LST, associated with a time-delay which decreased from around 20 minutes to 0 minutes throughout the period. Beginning around 20:30 LST, strong positive correlation was also observed at time-delays of approximately 30 minutes, decreasing to around 10 minutes by 22:00 LST. In the lower thermosphere, strong negative correlation was observed around zero time-delay between approximately 19:40 and 21:40 LST. Positive correlation was limited mostly to the period between 20:00 and 21:00 LST, with a time-delay of $\backsimeq$ 20 minutes. A time-delay between vertical wind events and horizontal divergence was also by observed by Ishii et al. (2001) above Poker Flat, Alaska.

Ground-based spectroscopic observations of upper atmospheric winds often rely on the assumption that, over the period of observation (usually one night), the mean vertical wind speed will be zero, or in other words that during a night's observation there will be no net upward or downward motion of the atmosphere. This assumption is used to obtain a zero Doppler-shift reference in the absence of a convenient laboratory source of the line under investigation (in this case at 630.0 nm). Aruliah and Rees (1995) argued that in the absence of strong localised heating, and with a full 24 hour period of observation, this assumption may be valid. However, most spectrometers observe during the hours of darkness, which at high-latitudes implies a dependence of the length of the daily observation period on the day of year, maximising at mid-winter. From 1242 nights of Fabry-Perot spectrometer data covering the period from November 1981 through to April 1990, Aruliah and Rees (1995) found that during periods of moderate magnetic activity ($2 < \mathrm{Kp} < 5$), mean vertical winds above Kiruna (68°N) showed a systematic variation during the night, with consistently upward winds not exceeding 10 m.s^{-1} observed between 20 UT and 02 UT, coinciding with times when the auroral oval was overhead. These authors concluded that during the mid-winter period (when observing periods are long) and under conditions of low geomagnetic activity, the zero-velocity assumption was likely valid. However, under geomagnetically active conditions, or at times when observing periods are short, a systematic error in the zero Doppler-shift baseline equivalent to 10-20 m.s^{-1} may result.

2.4.3 Small Scale Structure

An assumption often made when dealing with upper thermospheric dynamics is that the large viscosity of this region smooths out gradients (both vertical and horizontal) in neutral winds. Under this assumption wind flows should be uniform across the field-of-view of a typical spectrometer ($\sim$ 1000 km). Wang et al. (2008b) investigated vertical profiles of upper thermospheric neutral winds using the Coupled Magnetosphere-Ionosphere-Thermosphere Model (CMIT). They found that at low and mid latitudes, wind flow during periods of low magnetospheric forcing (geomagnetically quiet times) indeed showed minimal variation with altitude. However there were noticeable vertical gradients in the horizontal winds at high latitudes which were present even during quiet times. During storm times, significant vertical shears were observed globally at both high and low latitudes. These authors concluded that during both quiet and storm conditions viscosity was not a major forcing term for the neutral winds at most heights, and that faster processes such as pressure gradient forcing and ion-drag were able to set up significant wind shears which viscosity, being a relatively slow process, could not effectively oppose.

Observational evidence for large vertical gradients in horizontal winds have been reported (Larsen et al., 1995; Larsen and Odom, 1997; Larsen, 2002), which used chemical releases from rockets to determine the altitude profiles of zonal and meridional winds. Chemical release experiments provide unambiguous measurements of the wind with excellent altitude resolution. These results often showed strong shears on the bottomside of the E-region wind peak, for example reversals in wind direction (requiring a change in wind speed of more than 100 m.s^{-1}) over vertical scales as small as 10-15 km (Larsen and Odom, 1997). In a survey of chemical release experiments over four decades, Larsen (2002) presented superposed vertical profiles of wind shear, which indicated a region of maximum wind shear between approximately 100 and 110 km, with shear magnitudes up to 100 m.s^{-1}.km^{-1}. The shears were distributed evenly in both the meridional and zonal wind components.

A prototype Doppler imaging system (DIS) was used by Rees et al. (1984a) to observe the 630.0 nm airglow emission near Kiruna, Sweden (67.8°N, 20.4°E). This prototype instrument was a field-widened (80° full-angle field-of-view) Fabry-Perot spectrometer utilising an imaging photon detector to record 6 interference fringes (see Chapter 3) at a wavelength of 630.0 nm. A total of 144 independent spectra were derived from 24 azimuthal sectors around each fringe. Gaussian fits to these spectra resulted in estimates of line-of-sight wind speed, temperature and intensity. A horizontal vector wind-flow pattern was then derived through a least entropy fit to the observed (radial) line-of-sight winds. Rees et al. (1984a) reported observations from this instrument during a geomagnetic disturbance which followed a storm sudden commencement (see Section 1.2.3) on December 17, 1982, during which time a second narrow-field FPS was also in operation. By combining the horizontal vector wind fields derived from both instruments, the authors were able to estimate the horizontal wind flow over an area of the sky some 800 km in diameter. Strong westward flow was observed, with average speeds around 750 m.s^{-1}. The inferred flow patterns showed strong gradients perpendicular to the mean flow, with speed changes of ± 150 m.s^{-1} over spatial scales of 100-200 km. Approximately one hour after this flow pattern was observed, a slower (~ 650 m.s^{-1}) flow was observed, with smaller gradients, which reflected a decreased level of geomagnetic activity.

A later study (Batten and Rees, 1990) using the same DIS reported frequent observations of small-scale structure in the neutral thermospheric wind field, such as strong meridional and zonal gradients, rapid rotations in wind direction (45° rotation in under 20 minutes, for example), as well as oscillations in the direction of wind flow and, frequently, convergent or divergent winds. Of the latter, some periods of strong divergence/convergence were associated with rapid changes of local auroral brightness, however on other occasions the wind fields were uniform despite the presence of bright, highly structured aurora within the field-of-view (Batten and Rees, 1990). With regard to the auroral oval, abrupt wind speed reductions were frequently observed over a region of approximately 100 km on the equatorward edge of the oval.

Killeen et al. (1988) combined ultraviolet images of the auroral oval with thermospheric neutral wind measurements derived from Dynamics Explorer (1 and 2) data. Spatially narrow reversals were often observed to be correlated with the location of the auroral oval. Modelling studies (Fuller-Rowell, 1985; Walterscheid and Lyons, 1992; Keskinen and Satyanarayana, 1993) have predicted that strong neutral wind shears can develop in the vicinity of auroral arcs on spatial scales of tens to hundreds of kilometers. Conde et al. (2001), using an all-sky imaging Fabry-Perot spectrometer (a scanning Doppler imager, or SDI), compared F-region horizontal vector wind fields with auroral images and magnetometer-derived electrojet currents to investigate the spatial relationship between the visual aurora, ionospheric currents and neutral winds. Strong latitudinal neutral wind shears were observed, associated with auroral boundaries and shears in the ionospheric plasma convection. Ion-drag from the inferred F-region plasma convection was observed to drive a latitudinally confined channel of geomagnetically westward wind flow before midnight, which was bounded on both its poleward and equatorward edges by latitudinal shears in the neutral wind.

Using the same instrument, Conde and Smith (1998) investigated the spatial structure of thermospheric neutral winds above Poker Flat, Alaska. These authors observed spatially structured winds within the approximately 1000 km diameter field-of-view of the instrument, structure which appeared to arise in response to ion-drag forcing. For example the cross-polar jet exhibited a sharp boundary on the evening side, and was often preceded by a period of weak and variable wind flow. There was also a considerable contribution to the evening sunward flow from lower latitudes. These features were not reproduced in a generic run of NCAR's TIEGCM model.

Many investigators have reported observations of horizontal divergence in upper thermospheric winds. Biondi (1984) observed mid-latitude thermospheric neutral winds using a pressure-scanned Fabry-Perot spectrometer that was field-widened by a multiaperture exit plate (described by Sipler et al., 1983). By observing the 630.0 nm airglow line at three different elevation angles, these authors inferred a divergence in the horizontal meridional wind of $\simeq 8 \times 10^{-5}$ s^{-1} in a particular 10 minute time window. Upward vertical winds were observed to be associated with divergent horizontal flow (in both the meridional and zonal directions), and downward vertical winds were observed when the horizontal flow was convergent. Wardill and Jacka (1987) observed zonal divergence of 1.2×10^{-4} s^{-1} at Mawson, which they attributed to momentum transferred to the neutral atmosphere by dissipating atmospheric gravity waves, while Crickmore et al. (1991) reported an average divergence in the zonal wind of about 9×10^{-5} s^{-1} from 4 nights of observations at Halley, Antarctica. By combining measurements from Fabry-Perot spectrometers located at Mawson and Davis stations, Greet et al. (1999) inferred the thermospheric horizontal vector wind field above both sites. Both meridional and zonal gradients in the inferred wind fields were observed, the meridional gradients associated with auroral substorm activity.

Rees et al. (1998) described two Doppler imaging systems that were installed at Kiruna, Sweden, and at Longyearbyen, Svalbard. These instruments were more advanced versions of the prototype system described by Rees et al. (1984a). Data were presented from both instruments, which showed considerable ($\sim$ 100 m.s^{-1}) differences in line-of-sight wind speeds estimated in opposite cardinal look directions - separated by a distance of approximately 400 km - in both the zonal and meridional directions. Aruliah and Griffin (2001) also observed a significant gradient in the *average* neutral zonal wind near Kiruna, over a horizontal scale of approximately 800 km, indicating that the divergence was a recurrent feature. In addition, these authors showed a 150 m.s^{-1} wind speed change coincident with a sharp spike in the neutral temperature of 70 K magnitude, both of which events lasted less than 40 minutes. Using an all-sky imaging instrument based near Longyearbyen, Svalbard (78.2°N,15.8°E), Griffin et al. (2008) reported observations of the 630.0 nm airglow emission which showed significantly different line-of-sight neutral wind speeds (by around 100 m.s^{-1}) in atmospheric volumes separated by only $\sim$ 220 km, during a period of Joule heating.

Ford et al. (2007) presented results of very high time resolution measurements of thermospheric neutral wind, temperature and intensity using a narrow-field (1° field-of-view) FPS near Kiruna, Sweden. The best time resolution achieved by the instrument was 15 seconds between consecutive measurements, with resulting wind uncertainties of 14-18 m.s^{-1} on average. By viewing the wind speed, intensity and temperature time-series on progressively smaller time-scales, these authors estimated that the minimum time resolution needed to fully resolve the short time-scale variability was around 1 minute. This is a very short time considering that the e-folding time for the neutrals is generally between 1/2 - 6 1/2 hours (Kosch et al., 2001; Heelis et al., 2002).

These studies indicate that significant structure can exist in upper thermospheric neutral winds on small spatial and temporal scales. Many of the early results suggesting small-scale structure (such as horizontal wind divergence) were limited by a trade-off between spatial and temporal resolution, introduced by narrow-field spectrometers cycling through small atmospheric observing volumes. This introduced temporal ambiguities when comparing wind measurements made in opposite look-directions (for example in the geographic north and south) to look for divergence, since the measurements were not concurrent. The all-sky imaging system of Rees et al. (1984a) overcame this ambiguity by deriving multiple, independent wavelength spectra from fringe images mapping to a large field-of-view on the sky simultaneously (see also Sekar et al., 1993; Biondi et al., 1995), and this concept has been further advanced by Conde and Smith (1997), who implemented a separation-scanning etalon, removing the need for spatial scanning and the distortions introduced by angular variation of the source brightness distribution, while minimising distortions due to temporal variation in the source brightness by co-adding spectra from multiple short-duration etalon scans. The principal instrument used in the present study is of the type described by Conde and Smith (1997). It allows for simultaneous optical measurements of the airglow from many tens of locations across a large (144° full-angle) field-of-view, at a spatial resolution on the order of 100 km (at the peak height of the 630.0 nm emission), and with integration times as low as 1-2 minutes during very bright aurora (and 3-5 minutes during normal viewing conditions). The major strength of the instrument is thus its ability to resolve spatial gradients in horizontal winds and temperatures, while maintaining a reasonable time resolution. It is thus perfectly suited to the investigation of small-scale thermospheric dynamics.

2.4.4 Waves

Gravity waves were mentioned briefly in Section 2.3.4 in the context of thermal balance in the thermosphere. The atmosphere is capable of supporting various types of wave propagation. Gravity waves (see Nappo (2002) for a detailed discussion of gravity waves) are those oscillations for which the restoring force is buoyancy, with frequencies less than (or equal to) the Brunt-Väisälä (or buoyancy) frequency ω_b, given by:

$$w_b = \sqrt{\frac{g}{\theta}\frac{\partial\theta}{\partial z}} \tag{2.4.2}$$

The quantity θ is called the *potential temperature* and is defined by:

$$\theta = T\left(\frac{p_0}{p}\right)^{\frac{\gamma-1}{\gamma}} \tag{2.4.3}$$

where $\gamma = \frac{c_p}{c_v}$ is the ratio of the specific heat at constant pressure to that at constant volume ($\simeq 1.4$), and θ represents the temperature that an air parcel would have when brought to a standard reference pressure p_0. The buoyancy period ($\frac{2\pi}{w_b}$) varies with altitude between approximately 5 minutes at 100 km and up to 25 minutes in the upper thermosphere (Yeh and Liu, 1974).

The general treatment of gravity waves involves separating the equations describing atmospheric properties such as wind components, pressure, density, and temperature (Section 2.5) into "background" (or slowly-varying) and perturbation components. The perturbation components are then solved for by assuming harmonic wave solutions of the form:

$$\psi(t, \mathbf{x}) = A \exp[i(\omega t - \mathbf{k} \cdot \mathbf{x})] \tag{2.4.4}$$

where ψ represents an atmospheric property (for example pressure), A is the amplitude, ω is the angular frequency, and $\mathbf{k}$ is a complex wave vector, having (complex) components in three spatial dimensions ($\mathbf{k} = (k_x, k_y, k_z)$). The solutions ψ are functions of both time t and position $\mathbf{x}$, and are constrained by a dispersion relation (describing the relation between wave properties such as frequency and propagation direction/speed), which for two-dimensional propagation (horizontally in the x direction and vertically in the z direction) in an isothermal, non-dissipative atmosphere is given by (Brekke, 1997):

$$\omega^4 - ((k_x^2 + k_z^2)c^2 + w_a^2)w^2 + k_x^2 c^2 w_b^2 = 0 \tag{2.4.5}$$

where:

$$c = \sqrt{\gamma g H} \qquad \text{(sound speed)} \tag{2.4.6}$$

$$w_a = \frac{g\sqrt{\gamma}}{c} \qquad \text{(acoustic cutoff frequency)} \tag{2.4.7}$$

The values of w, k_x and k_z that satisfy Equation 2.4.5 are shown in Figure 2.4.5 (left panel), for an assumed sound speed of 800 m.s^{-1} (see Yeh and Liu (1974)). There are two separate branches to the dispersion surface: a high-frequency branch, the waves of which are termed "acoustic" waves, and a low-frequency branch, corresponding to gravity waves. The main distinction between these two types of waves is the restoring force, which is pressure in the case of acoustic waves and buoyancy in the case of gravity waves. For the acoustic branch, as $k \to 0$ the frequency $\to w_a$, hence w_a is termed the acoustic cutoff frequency, below which acoustic waves cannot propagate. For gravity waves, as k becomes very large, $\omega \to w_b$. Between the two allowed branches is a region where no internal wave propagation is possible. Equation 2.4.5 can be rearranged to show the relationship between the vertical and horizontal wave numbers:

$$k_z^2 = k_x^2 \left(\frac{w_b^2 - w^2}{w^2} \right) - \left(\frac{w_a^2 - w^2}{c^2} \right) \tag{2.4.8}$$

the form of which can be seen in the right panel of Figure 2.4.5. In this figure, which is a contour plot of the dispersion surface shown in the left panel, solid lines trace out points of constant frequency in the (k_x, k_z) plane (red lines represent acoustic waves, blue lines gravity waves). The wave phase velocity (the speed of a point of constant phase in the direction of wave propagation, $\mathbf{k}$) is given by:

$$\mathbf{v}_p = \frac{\omega \hat{\mathbf{k}}}{\sqrt{k_x^2 + k_z^2}} \tag{2.4.9}$$

for the two-dimensional propagation of waves having a frequency ω (i.e. a single frequency component). However the group velocity, at which the energy of the wave propagates, is given by:

$$\mathbf{v}_g = \left(\frac{\partial \omega}{\partial k_x}, \frac{\partial \omega}{\partial k_z} \right) \tag{2.4.10}$$

which in the case of gravity waves (in a dispersive medium) is directed perpendicular to lines of constant ω (the contours of Figure 2.4.5) and with a vertical component of the opposite sign to that of the phase velocity (Yeh and Liu, 1974). Thus, for gravity waves with upward phase progression, energy propagation is downwards, and vice versa.

As mentioned in Section 2.3.4, waves propagating upward from below the thermosphere can provide net heating/cooling and deposit momentum, while waves which originate in the thermosphere (for example from auroral sources) can redistribute energy and momentum within the thermosphere, by propagating large distances away from their source regions before being dissipated. Travelling ionospheric disturbances (TID's), the ionospheric signature of gravity waves, have been studied extensively (see for example Hunsucker (1982); Williams et al. (1993); Hocke and Schlegel (1996); MacDougall et al. (2001)). These waves are present on a range of scales, from small-scale, with periods approaching the buoyancy frequency, up to large-scale waves, with periods of 3 hours and wavelengths on the order of 1000 km (Hunsucker, 1982).

Optical observations of gravity wave signatures are less numerous. de Deuge et al. (1994) used a three-field photometer at Mawson station, Antarctica, to observe intensity variations in the 630.0 and 557.7 nm airglow emissions of atomic oxygen (at peak emission heights of approximately 240 and 120 km respectively, see Section 2.6). Seven years of data from 1982-1989 were analysed for the presence of gravity waves. At the upper altitude,

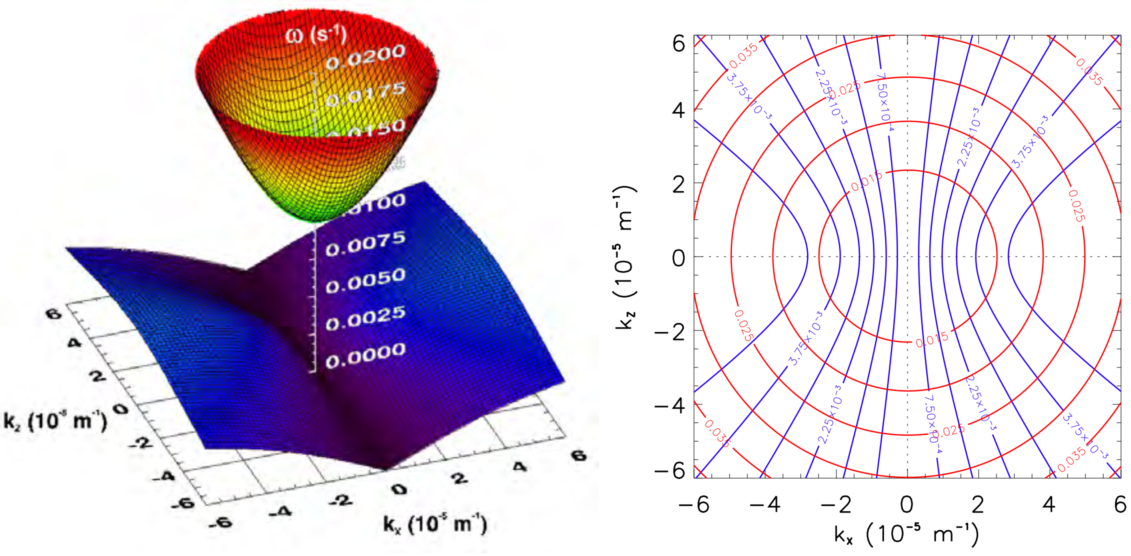


Figure 2.4.5: Left: the acoustic-gravity wave dispersion surface, defined by Equation 2.4.5, calculated for a sound speed of 800 m.s^{-1}, showing the acoustic wave branch (upper surface, coloured green to red) and the gravity/buoyancy wave branch (lower surface, coloured black to blue). Right: contours of the dispersion surface shown on the left. Lines of constant ω are plotted in the (k_x, k_z) plane, with red lines corresponding to acoustic waves, and blue lines corresponding to gravity waves.

waves were observed predominantly with periods of 30-60 minutes, and with horizontal trace speeds of 100-300 m.s^{-1} (the trace speed is the speed of a point of constant phase along a given direction). The propagation direction at this altitude, throughout the study interval of 7 years, was dominantly geographic northwest, which at Mawson translates to the geomagnetic northward direction, consistent with the assumption of an auroral electrojet source (de Deuge et al., 1994). In 1999, Innis et al. (2001) used a narrow-field FPS observing only in the zenith direction, in conjunction with the same three-field photometer used by de Deuge et al. (1994) to search for gravity wave signatures above Davis station, Antarctica ($\sim$ 635 km east of Mawson). Of 21 cloud-free nights, 3 were selected as suitable for analysis, where a common wave-like signal was observed from both instruments. Derived horizontal phase velocities were in the range 230-340 m.s^{-1}, periods between 15 and 24 minutes and horizontal wavelengths in the range 220-450 km.

Fagundes et al. (1995) used an imaging FPS at Kiruna (67.8°N, 20.4°E) observing the 630.0 nm airglow line to measure the meridional and zonal components of the horizontal wind and the emission intensity during geomagnetic storms. A low band-pass filtering technique was used to separate the short-period fluctuations from the background winds, and a cross-correlation analysis used to infer wave propagation velocities. These authors noted that large wind perturbations were associated with intensity enhancements and/or close proximity to the auroral electrojet, and that the meridional speed perturbations were larger than the zonal wind perturbations. On two nights wave propagation was southward, with horizontal trace speeds of 890 and 530 m.s^{-1}, and a northward propagating wave with a trace speed of 380 m.s^{-1} was observed on another night. Joule heating from the auroral electrojet was proposed as the source of these waves.

Ford et al. (2008) conducted a statistical analysis of data from three Fabry-Perot spectrometers in Sweden, Finland and Norway. No dependence on geomagnetic activity was found, nor was a preferred wave frequency. Fewer short-period waves were observed by the FPS at the higher latitude site, which these authors attributed to energy dissipation by ion-drag, and which supported the idea of the auroral oval as the primary source region for the waves. In their review of gravity waves and TID's, Hocke and Schlegel (1996) reviewed a number of studies into gravity wave generation and propagation. Key points from this review were:

- A strong correlation was found between auroral electrojet variability and the time-delayed occurrence of large-scale TID's at mid-latitudes. Auroral zone Joule heating and Lorentz forcing, which often showed a quasi-periodic time structure, frequently heralded the arrival of large-scale TID's at lower latitudes, waves which possessed a similar periodicity to the assumed auroral source.

- The average speed of equatorward-propagating large-scale TID's was 500 m.s^{-1}.

- Medium-scale TID's appeared often during periods of quiet and moderate geomagnetic activity, and their occurrence did not seem to increase with increasing activity.

- These medium-scale TID's were mostly equatorward propagating.

- TID's observed during the day had smaller amplitudes than those observed at night, presumably due to higher rates of wave dissipation during the day from ion-drag.

At Arecibo observatory (18°N, 67°W), Djuth et al. (2004) reported observations suggesting that a continuum of gravity waves is routinely present above that site at thermospheric altitudes. Ford et al. (2007) observed gravity waves with periods down to 14 minutes in FPS data on several nights during a high time-resolution campaign, suggesting that short period gravity waves were present reasonably often. As mentioned in Section 2.4.2, Innis and Conde (2002) produced maps of vertical wind activity using data from the Dynamics Explorer 2 satellite, which showed that the greatest vertical wind variability was largely confined to the region bounded by the auroral oval (Figure 7.4.2 in Chapter 7 shows one of these maps). These authors noted that wave-like structures were often observed in the DE-2 vertical wind data, and associated the increased vertical wind activity at high-latitudes with the presence gravity waves. The observational evidence thus supports the idea that auroral zone heating and dynamics are an important source of thermospheric gravity waves on a range of scales, which can propagate to much lower latitudes and thereby redistribute magnetospheric energy deposited in the high-latitude thermosphere, particularly during geomagnetic storms and substorms. These waves, the climatology of which are only poorly understood (Hocke and Schlegel, 1996), play a very important role in thermospheric dynamics on both the small and large scale (Smith, 2000).

2.5 Mathematical Formulation

The mathematical formulation of the behaviour of the atmospheric gas mixture, in terms of its composition, energetics and dynamics, follows from considering an air parcel element of volume dV, mass density ρ, and bulk velocity $\mathbf{u}$. The atmosphere is assumed to behave as an ideal gas, and for such a gas the equation relating pressure, density and temperature is the ideal gas law, which was introduced previously (Equation 2.2.2). This relationship is reproduced here:

$$p = \rho R_s T \tag{2.5.1}$$

where R_s is the specific gas constant, equal to the ratio between the universal gas constant $R = 8.314$ J.mol^{-1}.K^{-1} and the mean molar mass. The governing equations then follow from the requirement for conservation of mass, momentum and energy, which give rise to the equations controlling composition, velocity and temperature respectively. While the expressions derived below are valid for a single species atmosphere, the formulation is often extended to apply to a multi-component atmosphere in global-scale numerical models (Section 2.7).

Mass Continuity: The equation controlling composition has been encountered already (Equation 2.2.6). In terms of the mass density ρ, it is seen that the time rate of change of this quantity will depend solely on the divergence of the mass flux, since there can be no in-situ production or loss of mass. Thus the terms on the right hand side of Equation 2.2.6 vanish, and the continuity equation becomes:

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{u}) = 0 \tag{2.5.2}$$

This form of the continuity equation holds for an observer fixed relative to Earth's surface, and states that the time-rate of change of mass density within the volume is due solely to the divergence of the mass flux.

Momentum Conservation: Conservation of momentum, expressed by Newton's second law of motion, leads to an expression for the time rate of change of velocity due to the sum of the forces (per unit mass) acting on the air parcel:

$$\mathbf{a} = \frac{d\mathbf{u}}{dt} = \sum \frac{\mathbf{F}}{m} \tag{2.5.3}$$

The Lagrangian time derivative operator ($\frac{d}{dt}$) in Equation 2.5.3 holds in an inertial frame moving with the air parcel (with velocity $\mathbf{u}$), however often the most convenient formulation is that in a frame of reference which is fixed relative to Earth's surface, which is a rotating (and hence non-inertial) frame of reference. Equation 2.5.3 can be recast relative to this frame firstly by replacing the Lagrangian derivative with the Eulerian[2] derivative by including the effects of advection:

$$\frac{d}{dt} = \frac{\partial}{\partial t} + (\mathbf{u} \cdot \nabla) \tag{2.5.4}$$

and secondly by the inclusion of the acceleration due to the Coriolis and centrifugal pseudo-forces (see Rees (1989, pp. 204)):

$$\mathbf{a}_{\text{cor}} = -2\mathbf{\Omega} \times \mathbf{u} \tag{2.5.5}$$

$$\mathbf{a}_{\text{cen}} = |\mathbf{\Omega}|^2 \mathbf{R} \tag{2.5.6}$$

where $\mathbf{\Omega}$ is Earth's angular velocity vector, and $\mathbf{R}$ is the displacement vector from Earth's rotation axis to the air parcel, in a direction perpendicular to that axis. Equation 2.5.3 then becomes:

$$\frac{\partial \mathbf{u}}{\partial t} = \sum \frac{\mathbf{F}}{m} - (\mathbf{u} \cdot \nabla)\mathbf{u} - (2\mathbf{\Omega} \times \mathbf{u}) + |\mathbf{\Omega}|^2 \mathbf{R} \tag{2.5.7}$$

Note that Equation 2.5.7 cannot be considered a true vector equation (one which is invariant under a change of coordinates) unless the ∇ operator of Equation 2.5.4 is interpreted as the covariant derivative when applied to the vector field $\mathbf{u}$. However, in Cartesian coordinates, the form of Equation 2.5.7 can be considered a convenient shorthand for expressing the three Cartesian components of the equation of motion (as pointed out by Beer (1974, pp. 51)). Explicitly, the x component of the Eulerian derivative of $\mathbf{u}$ is given

[2]This derivative goes by many names, for example convective, material, total, Stokes derivative, etc.

by:

$$\frac{du_x}{dt} = \frac{\partial u_x}{\partial t} + \left(u_x \frac{\partial u_x}{\partial x} + u_y \frac{\partial u_x}{\partial y} + u_z \frac{\partial u_x}{\partial z} \right) \tag{2.5.8}$$

The forces included in the summation in Equation 2.5.7 are:

$$\mathbf{a}_{\text{grav}} = \mathbf{g} \tag{2.5.9}$$

$$\mathbf{a}_{\text{pres}} = -\frac{1}{\rho}\nabla p = -\frac{R_s}{\rho}\nabla(\rho T) \tag{2.5.10}$$

$$\mathbf{a}_{\text{visc}} = \eta \nabla^2 \mathbf{u} \tag{2.5.11}$$

Here, terms (2.5.9) and (2.5.10) represent accelerations due to gravity and a pressure-gradient. Term (2.5.11) is an approximation to the acceleration due to viscous drag (Salby, 1995, pp. 408), where $\eta = \mu/\rho$ is the coefficient of kinematic viscosity (μ being the coefficient of dynamic viscosity). Another important momentum source at high latitudes comes from the Lorentz force acting on the ionized component of the atmosphere, which is coupled to the neutral component through collisions. This force is given by:

$$\mathbf{F} = \mathbf{j} \times \mathbf{B} \tag{2.5.12}$$

At high latitudes $\mathbf{B}$ is often assumed to be vertical, and is therefore given by $B_0\hat{\mathbf{z}}$. The current density $\mathbf{j}$ is given by $\sigma \mathbf{E}'_{\perp}$, where the electric field is assumed to have horizontal components only (Section 2.3.2), and σ is the conductivity tensor, namely:

$$\begin{bmatrix} \sigma_P & \sigma_H & 0 \\ -\sigma_H & \sigma_P & 0 \\ 0 & 0 & \sigma_0 \end{bmatrix} \tag{2.5.13}$$

Equation 2.5.12 then becomes (Larsen and Walterscheid, 1995):

$$\begin{aligned} \mathbf{F} &= \sigma \cdot (\mathbf{E}_{\perp} + \mathbf{u}_n \times B_0\hat{\mathbf{z}}) \times B_0\hat{\mathbf{z}} \\ &= \left[\sigma_H B_0^2 (v_y - u_y) + \sigma_P B_0^2 (v_x - u_x)\right] \hat{\mathbf{x}} \\ &\quad + \left[\sigma_P B_0^2 (v_y - u_y) - \sigma_H B_0^2 (v_x - u_x)\right] \hat{\mathbf{y}} \end{aligned} \tag{2.5.14}$$

Here u and v represent neutral and ion velocity components respectively, where the subscripts x and y represent the zonal and meridional directions. In the F-region only the Pedersen conductivity is retained, which gives an expression for the acceleration due to ion-drag:

$$\begin{aligned} \mathbf{a}_{\text{ion}} &= \frac{\sigma_P B_0^2}{\rho}(\mathbf{v} - \mathbf{u}) \\ &= \nu_{ni}(\mathbf{v} - \mathbf{u}) \end{aligned} \tag{2.5.15}$$

where ν_{ni} is in this case taken to be the mean collision frequency between ions and neutrals. Including acceleration terms (2.5.9 - 2.5.11) and (2.5.15), the momentum equation (2.5.7) becomes:

$$\frac{\partial \mathbf{u}}{\partial t} = \mathbf{g} - \frac{R_s}{\rho}\nabla(\rho T) + \eta\nabla^2\mathbf{u} + \nu_{ni}(\mathbf{v} - \mathbf{u}) - (\mathbf{u}\cdot\nabla)\mathbf{u} - (2\mathbf{\Omega}\times\mathbf{u}) + |\mathbf{\Omega}|^2\mathbf{R} \tag{2.5.16}$$

Some conclusions can be drawn regarding the behaviour of the thermospheric wind field from an analysis of the individual terms in the above equation. The centrifugal term is very small ($\lesssim 3\times10^{-2}$ m.s^{-2}), and if not neglected altogether is often incorporated into an effective gravity term, $\mathbf{g}^* = \mathbf{g} + |\mathbf{\Omega}|^2\mathbf{R}$. On a global scale, the dominant terms (at thermospheric altitudes) in Equation 2.5.16 are the gravitational, pressure gradient and ion-drag terms. In the steady-state, this reduces to:

$$\frac{R_s}{\rho}\nabla(\rho T) = \mathbf{g} + \nu_{ni}(\mathbf{v} - \mathbf{u}) \tag{2.5.17}$$

Hydrostatic equilibrium holds to high precision on a global scale, and thus the gravitational force is balanced by the vertical component of the pressure gradient. Neglecting the ion velocity (valid for low and middle latitudes), Equation 2.5.17 can be written in terms of the horizontal components only:

$$\begin{aligned} u_x\hat{\mathbf{x}} + u_y\hat{\mathbf{y}} &= -\frac{R_s}{\rho\nu_{in}}\left(\frac{\partial}{\partial x}\hat{\mathbf{x}} + \frac{\partial}{\partial y}\hat{\mathbf{y}}\right)(\rho T) \\ &\simeq -\frac{R_s}{\rho\nu_{in}}\left(\frac{\partial T}{\partial x}\hat{\mathbf{x}} + \frac{\partial T}{\partial y}\hat{\mathbf{y}}\right) \end{aligned} \tag{2.5.18}$$

where gradients in the mass density are assumed negligible. Equation 2.5.18 therefore reproduces the global-scale circulation outlined in Section 2.4.1, in which thermospheric wind flow is directed perpendicular to contours of constant temperature (or pressure), away from the high-temperature region on Earth's sunward side, directed toward the low-temperature nightside.

Energy Conservation: Energy conservation is expressed by the first law of thermodynamics:

$$de = \delta q - \delta w \tag{2.5.19}$$

Each of these quantities (e energy, q heat and w work) represent energy per unit mass. The symbol δ is used here to indicate an inexact differential (for which the integral is path dependent). The energy e is the sum of the internal, kinetic (due to bulk motion) and potential energy of the system (air parcel), thus:

$$e = e_i + \left(\frac{1}{2}\mathbf{u}\cdot\mathbf{u}\right) + |\mathbf{g}|z \tag{2.5.20}$$

In this equation e_i is used to represent the internal energy of the gas (per unit mass), while the second and third terms on the right are the usual kinetic and potential energy for a parcel with velocity $\mathbf{u}$ and altitude z relative to some reference altitude. The differential of Equation 2.5.20 is then:

$$\begin{aligned} de &= de_i + \frac{1}{2}\left(\mathbf{u}\cdot d\mathbf{u} + d\mathbf{u}\cdot\mathbf{u}\right) + |\mathbf{g}|dz \\ &= de_i + \left(\mathbf{u}\cdot d\mathbf{u}\right) + |\mathbf{g}|dz \end{aligned} \tag{2.5.21}$$

Taking the time derivative of Equation 2.5.19, and using the energy expression given by Equation 2.5.21 gives:

$$\frac{de_i}{dt} + \left(\mathbf{u}\cdot\frac{d\mathbf{u}}{dt}\right) + |\mathbf{g}|\frac{dz}{dt} = \frac{\delta q}{\delta t} - \frac{\delta w}{\delta t} \tag{2.5.22}$$

For an ideal gas, the internal energy is solely a function of temperature (Adkins, 1975, pp. 117-118), such that $de_i = c_v dT$, where c_v is the specific heat at constant volume. For a simple compressible substance the total work done by the system is equal to $p\ dV$, thus:

$$c_v\frac{dT}{dt} + \left(\mathbf{u}\cdot\frac{d\mathbf{u}}{dt}\right) - \left(\mathbf{g}\cdot\mathbf{u}\right) = \frac{\delta q}{\delta t} - \frac{1}{dm}\frac{d(p\ dV)}{dt} \tag{2.5.23}$$

where the sign of the potential energy term has changed to reflect the fact that an increase in z coordinate (positive u_z) will increase the gravitational potential energy (for a gravitational force $\mathbf{g} = -g\hat{\mathbf{z}}$), and dm is the mass of the air parcel element. The time rate-of-change of work is expressed in terms of the volume element dV, hence in Cartesian coordinates:

$$\begin{aligned} \frac{d}{dt}(p\ dV) &= p\frac{d}{dt}(dxdydz) \\ &= p\ dydz\frac{dx}{dt} + p\ dxdz\frac{dy}{dt} + p\ dxdy\frac{dz}{dt} \\ &= pu_x\ dydz + pu_y\ dxdz + pu_z\ dxdy \\ &= \frac{d(pu_x)}{dx}\ dxdydz + \frac{d(pu_y)}{dy}\ dydxdz + \frac{d(pu_z)}{dz}\ dzdxdy \\ &= dV\left(\nabla\cdot p\mathbf{u}\right) \end{aligned} \tag{2.5.24}$$

Since $dV = {}^{dm}/_{\rho}$, Equation 2.5.23 may be written:

$$c_v\frac{dT}{dt} = \frac{\delta q}{\delta t} - \frac{1}{\rho}\left(\nabla\cdot p\mathbf{u}\right) - \left(\mathbf{u}\cdot\frac{d\mathbf{u}}{dt}\right) + \left(\mathbf{g}\cdot\mathbf{u}\right) \tag{2.5.25}$$

Various processes contribute to $\frac{\delta q}{\delta t}$. Absorption of solar radiation was outlined in Section 2.3.1. This will be included as part of a radiative term $\dot{q}_r$, which will be used to describe the net rate of radiative heat transfer. Molecular heat conduction is given by Equation 2.3.7, collisional heating due to differential plasma-neutral velocities by Equation 2.3.2, and viscous heating (heating due to relative motion between gas parcels) will be represented by ϵ (see for example Rees (1989, pp. 125) for a description of the viscous heating term, or refer to Section 8.5). Incorporating these terms, Equation 2.5.25 becomes:

$$c_v \frac{dT}{dt} = \dot{q}_r + \frac{\kappa}{\rho}\nabla^2 T + \frac{1}{2}\nu_{ni}\left(\mathbf{u}_i - \mathbf{u}_n\right)^2 + \epsilon - \frac{1}{\rho}\left(\nabla \cdot p\mathbf{u}\right) - \left(\mathbf{u} \cdot \frac{d\mathbf{u}}{dt}\right) + \left(\mathbf{g} \cdot \mathbf{u}\right) \quad (2.5.26)$$

All that remains is to recast the above equation into an Eulerian reference frame by replacing the explicit time derivatives according to Equation 2.5.4. This results in the final energy equation:

$$\begin{aligned} c_v \frac{\partial T}{\partial t} = \dot{q}_r + \frac{\kappa}{\rho}\nabla^2 T + \frac{1}{2}\nu_{ni}\left(\mathbf{u}_i - \mathbf{u}_n\right)^2 + \epsilon - \frac{1}{\rho}\left(\nabla \cdot p\mathbf{u}\right) \\ - \mathbf{u} \cdot \left(\frac{\partial \mathbf{u}}{\partial t} + (\mathbf{u} \cdot \nabla)\mathbf{u} - \mathbf{g} + c_v \nabla T\right) \end{aligned} \quad (2.5.27)$$

which relates the time rate-of-change of temperature to the rate of heating/cooling, work done by expansion/contraction of the air parcel, energy due to bulk motion and gravitational potential, and advection.

2.6 Airglow and Aurora: Spectroscopic Emissions

The term "airglow" refers to the diffuse, unstructured, global emission of radiation by atomic and molecular species in the atmosphere. The terms "dayglow" and "nightglow" are often used to distinguish between airglow emissions observed during the day and night, where it is understood that different processes will assume different levels of importance at these times. Photodissociation, photoionisation, energetic electron impact and chemical processes can all leave atmospheric species in excited states. If this excitation energy is not lost through collisions with other species (a process known as 'quenching'), it can be lost via the emission of radiation. Airglow radiation occurs over a large range of wavelengths from the far infrared to the far ultra-violet. This section will focus only on the 630.0 nm ('red line') emission of atomic oxygen. In what follows, use is made of the *Rayleigh* (R), which is a measure of the height-integrated emission rate, where:

$$1 \text{ R} = 10^6 \text{ photon.cm}^{-2}.\text{s}^{-1} \quad (2.6.1)$$

Auroral emissions at 630.0 nm are distinguished from the airglow principally by intensity, spatial structure, and method of excitation. The airglow is in general unstructured, diffuse, and relatively weak, with a typical zenith brightness at 630.0 nm of ~60 R (Leinert et al., 1998), and the principal production mechanism (at night) is dissociative recombination. For the auroral emission, energetic electron impact is the primary production mechanism of excited oxygen atoms (see Table 2.1). As a consequence, the height of peak red line emission during aurora is quite variable, controlled as it is by the characteristic energy of the precipitating electrons, while the aurora themselves are often very highly structured. Spectroscopic emissions such as the oxygen red line provide an invaluable tool for studying the upper atmosphere. In this study, line profiles of the 630.0 nm line

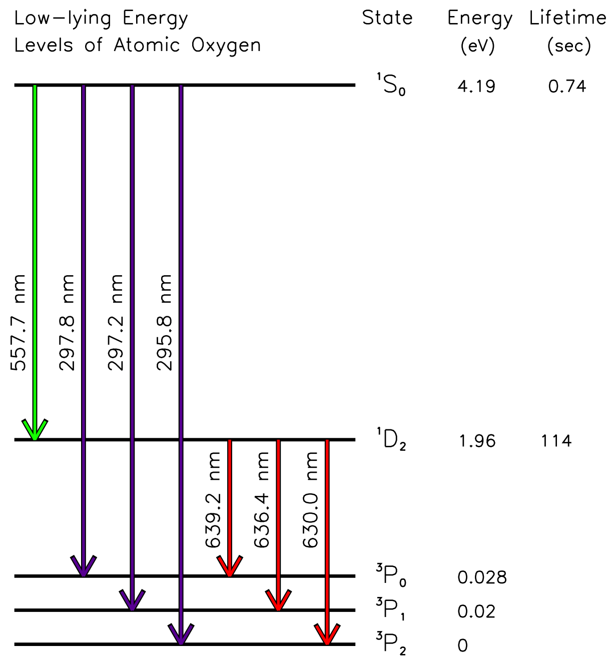


Figure 2.6.1: The low-lying energy levels of atomic oxygen, and the possible transitions between them. Adapted from Whitten and Poppoff (1971).

were used to infer neutral atmospheric temperatures (through Doppler-broadening), wind velocities (Doppler-shift), and emission intensities.

2.6.1 The Oxygen Red Line

The 630.0 nm emission line is the result of an intercombinational magnetic dipole transition between the excited ^1D state of atomic oxygen and its ground ^3P state. The transition can actually produce three lines, at 639.2, 636.4 and 630.0 nm, depending on the final energy state (see Figure 2.6.1), however the first of these is too weak to be observed (Whitten and Poppoff, 1971). Of the last two, the ratio of 630.0 to 636.4 nm is considered fixed at 3:1 (Rishbeth and Garriott, 1969). The magnetic dipole transition that produces the 630.0 nm line is a so-called "forbidden" transition in that it is not allowed by the electric dipole selection rules. There is however a finite (though low) probability of the transition occurring. This low transition probability - 0.0088 s^{-1} (Tachiev and Froese Fischer, 2002) - makes the ^1D state "metastable", since its lifetime (in the absence of collisions) is around 114 seconds.

A summary of the processes responsible for the production of $O(^1D)$ states is listed in Table 2.1. During the day, dissociation of molecular oxygen by solar radiation in the Schumann-Runge continuum (Reaction 1 in the table) is the major source of $O(^1D)$ atoms. At night, in the absence of aurora, the dissociative recombination reaction (2) is dominant. Both the ground (^3P) and excited (^1S, ^1D) states of atomic oxygen can be produced by this reaction, although the quantum yield (1.09 ± 0.15) is highest for the ^1D state (Kella et al., 1997). Sheehan and St.-Maurice (2004) give the temperature-dependent rate for this reaction as 1.95×10^{-7} $(300/T_e)^{0.7}$ $cm^3.s^{-1}$.

Type	Reaction
(1) Photodissociation	$O_2 + hv \rightarrow O(^1D) + O(^3P)$
(2) Dissociative Recombination	$O_2^+ + e \rightarrow O(^1D) + O$
(3) Electron Impact	$O(^3P) + e^* \ (E > 1.96\ eV) \rightarrow O(^1D) + e$
(4) Cascade	$O(^1S) \rightarrow O(^1D) + hv(\lambda = 557.7nm)$
(5) Ion-Atom Interchange	$N^+ + O_2 \rightarrow O(^1D) + NO^+$
(6) Quenching	$N^* + O \rightarrow O(^1D) + N(^4S)$
(7) Impact Dissociation	$O_2 + e^* \rightarrow O(^1D) + O + e$
(8) Atom interchange	$N(^2D) + O_2 \rightarrow O(^1D) + NO$

Table 2.1: Reactions leading to the production of the ^1D state of atomic oxygen.

Collisions between energetic electrons (with energy greater than the ^1D state threshold of 1.96 eV) and ground state oxygen atoms are another very important source of O(^1D). Under normal conditions, at an ionospheric electron temperature of 2000 K, only 0.01% of the (Maxwellian) electron population will posses sufficient energy to cause excitation (Mantas and Carlson, 1991). However, at times of increased magnetosphere-ionosphere coupling, when ionospheric electron temperatures can be greatly elevated, thermal electron impact is believed to be the primary source of stable auroral red (SAR) arcs (Cole, 1965; Kozyra et al., 1997). These arcs are steady, longitudinally extended, sub-visual enhancements of the 630.0 nm emission with intensities of up to at least 13 kR (Baumgardner et al., 2007) during intense magnetic storms, but more usually in the range of a few hundred to a few thousand Rayleigh (Sazykin et al., 2002). Of particular relevance to the present work is the excitation of atomic oxygen through collisions with energetic auroral electrons. These electrons can have energies in the range of hundreds of eV to hundreds of keV. As well as producing O(^1D) atoms and greatly enhancing the measured intensity of the 630.0 nm emission, incoming auroral electrons also enhance the local plasma density within auroral arcs, and can produce strong, localised heating of the neutral gas (Section 2.3.3).

Atomic oxygen in the metastable ^1S state can make the transition to a ^1D state by emitting a photon of wavelength 557.7 nm. At F-region heights, the main source of O(^1S) is Reaction 2, however the efficiency of this process is only around 5% (Kella et al., 1997). Thus the cascade reaction (4) constitutes a minor source of ^1D atoms, although during aurora it can be an important source at lower altitudes (Solomon et al., 1988). Other minor sources include ion-atom interchange (Reaction 5) between atomic nitrogen ions and molecular oxygen (Shematovich et al., 1999; Solomon et al., 1988), quenching (Reaction 6) of excited atomic nitrogen by atomic oxygen (also considered appreciable during aurora, see for example Solomon et al. (1988)), impact dissociation of O_2 (Reaction 7), and atom interchange between atomic nitrogen and O_2. This last reaction was proposed by Rusch et al. (1978), who, based on laboratory evidence, inferred a yield of O(^1D) from this

Reaction	Rate ($cm^3.s^{-1}$)	Reference
$O(^1D) + N_2 \rightarrow O + N_2$	$(2 \times 10^{-11}) \exp(107.8/T_n)$	Streit et al. (1976)
$O(^1D) + O_2 \rightarrow O + O_2$	$(2.9 \times 10^{-11}) \exp(67.5/T_n)$	Streit et al. (1976)
$O(^1D) + O \rightarrow O + O$	$(3.730 + 1.1965 \times 10^{-1} T_n^{0.5}$	
	$-6.5898 \times 10^{-4} T_n) \times 10^{-12}$	Sun and Dalgarno (1992)
$O(^1D) + e \rightarrow O + e$	6.6×10^{-10}	Sobral et al. (1992)

Table 2.2: Quenching reactions for the ^1D state.

reaction of 87%. However, Link (1983), based on decay rates of the 630.0 and 520.0 nm emissions from O(^1D) and N(^2D), inferred an upper limit for this yield of only 10%. Indeed, Link (1992) concluded that the evidence for this reaction as a source of O(^1D) was not compelling.

As mentioned previously, O(^1D) atoms may be lost either through collisional deactivation (quenching) with other thermospheric species (principally N_2, O and O_2), or via the emission of radiation (of which the majority is at 630.0 nm). The quenching reactions and their volume rate coefficients are listed in Table 2.2. While the rate of quenching due to O_2 is of the same order of magnitude as that due to N_2, the low O_2/N_2 ratio throughout the emission region ($\lesssim 0.1$) ensures that N_2 is the dominant quenching species (Hays et al., 1978). The last reaction in Table 2.2 represents quenching by thermal electrons.

2.6.2 Temporal and Spatial Variability

The 630.0 nm line is emitted from a broad layer between approximately 100 and 400 km altitude (see Figure 2.6.2), with an emission peak around 240 km (Bates, 1978; Witasse et al., 1999; de Meneses et al., 2008). Below about 200 km the emission rate decreases rapidly due to increasingly high quenching rates. Electron density plays an important role in nighttime 630.0 nm airglow emission due to its appearance in Reaction 2 (Table 2.1). The volume emission rate thus depends strongly on the behaviour of the bottomside F2 layer. Abreu et al. (1982) derived altitude profiles of 630.0 nm airglow emission from the Atmospheric Explorer E satellite (AE-E) at low latitudes, and found that regions of enhanced/depleted volume emission rates corresponded well with regions in which $\mathbf{E} \times \mathbf{B}$ drifts would drive the plasma to lower/higher altitudes. During the solstice, emission rates were greater overall in the northern (winter) hemisphere than in the southern (summer) hemisphere, while the presence of a pressure enhancement near midnight - associated with the midnight temperature maximum (Mayr et al., 1979) - produced correspondingly higher emission rates (Abreu et al., 1982).

Zhang and Shepherd (2004) studied around 130,000 O(^1D) dayglow emission height profiles from the WINDII instrument aboard the Upper Atmospheric Research Satellite (UARS). Using this data they were able to derive empirical relations between peak and height-integrated emission rates and the solar zenith angle (χ) and solar radio flux ($F_{10.7}$).

These relations were given by:

$$V_p = (a_0 \mathrm{F}_{10.7} + a_1) \cos^{\frac{1}{e}} \chi + a_2 \tag{2.6.2}$$

$$I = [(b_0 \mathrm{F}_{10.7}) + b_1] \cos^{\frac{1}{e}} \chi + b_2 \tag{2.6.3}$$

with:

$$a_0 = 1.791 \times 10^{26}\ \mathrm{Hz.J^{-1}.cm^{-1}}$$

$$a_1 = 76.4\ \mathrm{cm^{-3}.s^{-1}}$$

$$a_2 = 35\ \mathrm{cm^{-3}.s^{-1}}$$

$$b_0 = 24.80 \times 10^{26}\ \mathrm{R.Hz.cm^2.s.J^{-1}}$$

$$b_1 = 608\ \mathrm{R}$$

$$b_2 = 390\ \mathrm{R}$$

where V_p is the peak volume emission rate (in units of photon.cm^{-3}.s^{-1}) and I is the height-integrated emission rate in Rayleigh. Over the entire data set V_p was found to vary between 100-550 photon.cm^{-3}.s^{-1}, and I from 1000-8500 R. The height of peak emission ranged between 190-280 km.

Culot et al. (2005), using a one-dimensional fluid/kinetic model (TRANSCAR), found that varying geomagnetic activity had a negligible effect on red line emission intensity. These authors also noted that while the model results showed an increase in peak emission altitude (of <10%), no statistical variation of the peak emission altitude was found in an analysis of 4 years worth of data from the Wind Imaging Interferometer (WINDII) on board the UARS. Generally good agreement between modelled and observed 630.0 nm emission characteristics have been found by a number of researchers, for example

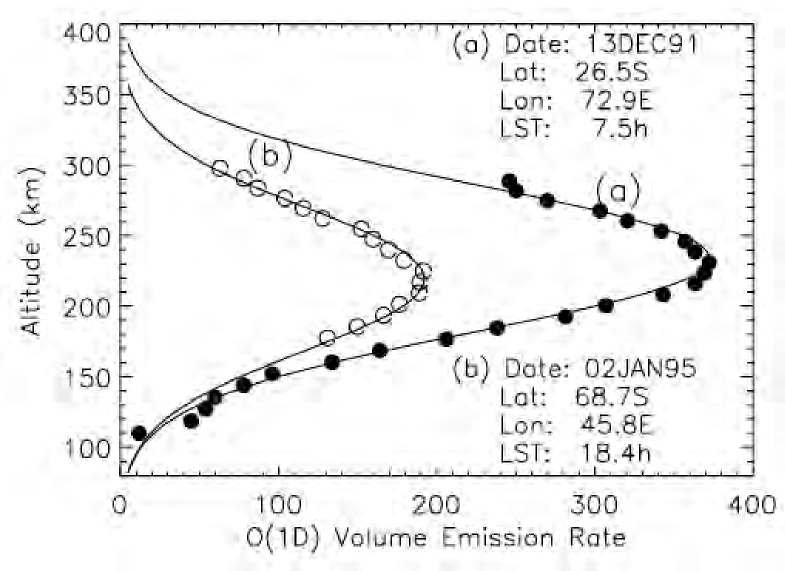


Figure 2.6.2: Samples of WINDII (Wind Imaging Interferometer) observations of the 630.0 nm dayglow from two different days. The volume emission rate is measured in units of photon.cm^{-3}.s^{-1}. Circles are the observed emission rates as measured by WINDII, solid lines are Gaussian curves fitted to the observed data. From Zhang and Shepherd (2004).

Hays et al. (1978), Witasse et al. (1999), Ogawa et al. (2002), Vlasov et al. (2005), and de Meneses et al. (2008), indicating that the dominant processes governing the emission are relatively well characterised.

2.6.3 Inferring Thermospheric Parameters

The 630.0 nm airglow line has been used extensively over many years to infer atmospheric parameters such as bulk velocity and temperature. A population of emitting species is assumed to have some random distribution of line-of-sight velocities both towards and away from the observer due to thermal motion. The range of Doppler-shifted wavelengths due to this distribution of velocities (which is assumed to be Maxwellian) will result in a broadening of the spectral line. This effect is known as Doppler-broadening, and has a Gaussian probability distribution given by:

$$P(\lambda)d\lambda = \frac{1}{f\pi^{\frac{1}{2}}}\exp\left(-\left[\frac{\lambda-\lambda_0}{f}\right]^2\right) \tag{2.6.4}$$

$$f = \sqrt{\frac{2kT\lambda_0^2}{Mc^2}} \tag{2.6.5}$$

where λ_0 is the central (emission) wavelength, T is the kinetic temperature of the emitting species, M is the atomic mass of the species, k is the Boltzmann constant and c is the speed of light. The full-width at half-maximum (FWHM) of this profile is given by:

$$\delta\lambda = \frac{2\sqrt{\ln(2)}}{f} = \frac{2\lambda_0}{c}\sqrt{\frac{2kT\ln(2)}{M}} \tag{2.6.6}$$

Equation 2.6.6 allows T to be estimated from recorded spectral line shapes of the emission if other sources of line broadening can be accounted for (and if the O(^1D) population is completely "thermalised", i.e. Maxwellian). Natural broadening, which arises due to the uncertainty in the energy associated with a transition between atomic energy levels, is negligible, owing to the (in atomic terms) very long lifetime of the metastable O(^1D) state ($\simeq$ 114 seconds). Pressure broadening, resulting primarily from collisions of the emitting species with atoms of the same kind ('self-broadening') or other atmospheric species ('foreign-gas broadening') is also negligible since, as the name implies, this process is proportional to the gas pressure, which in the upper atmosphere is very low. Broadening due to the finite spectral bandwidth of the instrument depends on the observing scheme adopted, and must be removed from the recorded line shape if Equation 2.6.6 is to be applied. This is usually achieved in practice by deconvolution from the recorded spectrum of an experimentally determined instrument function, which represents the instrument's response to a spectrally narrow source of radiation (such as a laser).

In the case that the emitting species is in thermal equilibrium with the ambient neutral gas, Equation 2.6.6 may be used to estimate the ambient neutral gas temperature. However, the newly created O(^1D) population will not in general meet this condition; for example, Reaction (2) in Table 2.1, which is the dominant O(^1D) production mechanism at night, is exothermic, the excess energy appearing as translational energy of the dissociation products (Yee, 1988). The degree of thermalisation depends upon the lifetime of the excited state, the collision frequency with ambient species (predominantly O and N_2 in the F-region), and the velocity distribution of the newly created $O(^1D)$ population (Yee, 1988). While collisions are required to thermalize the $O(^1D)$ population, collisional quenching will compete with radiative transition as a loss process (Witasse et al., 1999). Using a two-population model consisting of thermal and non-thermal $O(^1D)$ populations, Yee (1988) estimated the non-thermal fraction of the total $O(^1D)$ population. From these estimates he derived 630.0 nm airglow line profiles, and deduced that the intensity profiles which included the calculated non-thermal population differed by only a few percent from the profile for a completely thermalised population. The spectral line shape could thus be well approximated by a Gaussian function, and the signature of the non-thermal population was apparent only in the deduced temperatures, which could produce a systematic over-estimation of the ambient gas temperature of between approximately 50 and 100 K, depending on emission altitude (Yee, 1988). The two-population model neglected transport effects (such as diffusion) and did not attempt to model measurement noise introduced by the measurement apparatus.

2.7 Atmospheric Modelling

Many models have been developed of upper atmospheric composition, temperature and dynamics. Empirical models attempt to parameterize the dependence of various measured atmospheric parameters (horizontal wind components, species densities and temperature, for example) on factors such as geographical/geomagnetic location, local and universal time, geomagnetic and solar activity, solar wind and IMF conditions. This allows for a determination of the relative importance of each of these factors and their dependence on the geophysical parameters mentioned, and provides useful information against which direct measurements at a particular location can be compared. First-principles physics-based models, in contrast, attempt to calculate geophysical parameters from first principles, by solving the coupled equations of energy, momentum and mass conservation at discrete points in a spatial grid of 1, 2, or 3 dimensions. These "numerical" models can be used to test our understanding of the basic physical processes driving the atmosphere, through simulations of events for which observational data are available. A number of empirical and numerical models have been mentioned previously, and Table 2.3 provides a very brief summary of these and selected other models.

Table 2.3: Summary of some empirical and first-principles, physics-based "numerical" upper-atmospheric models.

Empirical Models	Description
NRLMSISE-00	The Naval Research Laboratory Mass Spectrometer and Incoherent Scatter Radar (Extended) Model, (Picone et al., 2002), based on the earlier MSISE90 model of (Hedin, 1991), which was itself an extension of the MSIS86 model (Hedin, 1987; Batten et al., 1987). It incorporates mass spectrometer, incoherent scatter radar and satellite drag data. Returns number densities of H, He, O, N, Ar, N_2, O_2 and Anomalous Oxygen, as well as total mass density and neutral temperature. Examples of model output are shown in Figures 2.1.1, 2.2.1, 2.2.3 and 2.4.1.
HWM93	The Horizontal Wind Model (Hedin et al., 1988, 1991). Returns zonal and meridional components of the neutral horizontal wind field for the given location, altitude, time, and solar/geomagnetic conditions. Examples of model output are shown in Figures 2.4.1 and 2.4.2.
IRI-2001	The International Reference Ionosphere (Bilitza, 2001). Returns the altitude-dependent electron density, electron temperature, ion temperature and composition (O^+, H^+, He^+, NO^+, O_2^+) and total electron content. Model output is shown in Figure 2.2.2.
Weimer 1996	Ionospheric electric potential model (Weimer, 1995, 1996). A more recent version "Weimer 2005" is currently available (Weimer, 2005). Returns the ionospheric electric potential at a given magnetic latitude/magnetic local time, for input IMF and solar wind parameters. Model output is shown in Figure 1.2.5.
First-principles, physics-based models	Description
NCAR TIEGCM	The National Center for Atmospheric Research's Thermosphere-Ionosphere-Electrodynamics General Circulation model (Dickinson et al., 1981; Roble et al., 1988; Richmond et al., 1992). Solves for the full 3D neutral wind vector, neutral, ion and electron temperature, and neutral and ion composition, at a horizontal resolution of 5° in both latitude and longitude.
TING	The Thermosphere-Ionosphere Nested Grid model (Wang et al., 1999). Extension of the NCAR TIGCM to incorporate selected regions of higher spatial resolution using nested grids.
CTIP	The Coupled Thermosphere/Ionosphere Plasmasphere (CTIP v1.0) model (Fuller-Rowell and Rees, 1980; Fuller-Rowell et al., 1996; Millward et al., 1996). Outputs the time-dependent temperature, density and velocity of the neutral atmosphere, as well as ion temperatures and densities over selected height ranges. Model output is shown in Figure 2.2.4.
GITM	The Global Ionosphere-Thermosphere Model (GITM) (Ridley et al., 2006). Solves for neutral, ion and electron velocity and temperature, and densities of 7 neutral and 7 ionic species, on a three-dimensional spherical grid. An altitude grid is used instead of a pressure grid, and hydrostatic equilibrium is not enforced.
TFM	The transfer function model (Mayr et al., 1984a,b, 1997), a linear, semi-analytic spectral model which describes perturbations in the thermospheric wind, temperature and composition with essentially unlimited temporal and horizontal resolution.

Chapter 3

The Fabry-Perot Spectrometer

3.1 General Principles

The design of the Fabry-Perot spectrometer (see Hernandez (1986) for a detailed treatment of Fabry-Perot spectrometer theory) is based around the 'etalon', a pair of flat, parallel, transparent plates separated by a spacer of thickness t and refractive index μ. The inner surface of each plate is coated with a highly reflective film, causing incident light rays to experience multiple reflections between the plates (see Figure 3.1.1). A fraction of the incident radiation is transmitted upon each reflection, so that a single incident ray is split into multiple emergent rays. Consider Figure 3.1.2, where two adjacent light rays are drawn, both originating on the same wavefront, with wavelength λ and incidence angle θ relative to the etalon plate normal. The refractive index is assumed to be the same both

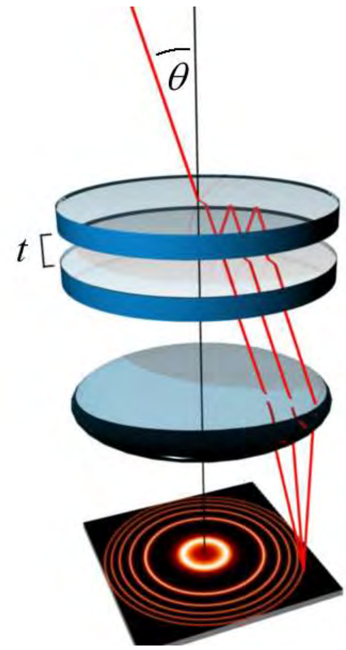


Figure 3.1.1: Fabry-Perot etalon schematic, showing the etalon plates separated by a gap t, and a positive power lens to focus the emergent rays onto a detector. The red line indicates a light path through the etalon.

above, below, and between the etalon plates.

After reflection from the lower, and then upper, plates, Ray 1 will be superposed on Ray 2 (they are shown separated in the diagram for clarity only). Any phase difference between the two rays will be due to the different paths travelled through the etalon, i.e. the segments of their respective paths which are not common to both rays (shown in red in Figure 3.1.2). The optical path difference (opd) between the reflected part of Ray 1 and Ray 2 is then equal to the product of the refractive index and the geometrical path difference, thus:

$$\begin{aligned}\Delta\text{opd} &= \mu\left(2t\tan(\theta)\sin(\theta) - \frac{2t}{\cos(\theta)}\right)\\ &= \frac{2\mu t}{\cos(\theta)}\left(\sin^2(\theta) - 1\right)\\ &= \frac{2\mu t\cos^2(\theta)}{\cos(\theta)}\\ &= 2\mu t\cos(\theta)\end{aligned} \tag{3.1.1}$$

where the negative sign has been ignored. The phase difference between these two rays is then:

$$\begin{aligned}\Delta\varphi &= \frac{2\pi}{\lambda}\Delta\text{opd} + 2\phi_r(\lambda)\\ &= \frac{4\pi\mu t\cos(\theta)}{\lambda} + 2\phi_r(\lambda)\end{aligned} \tag{3.1.2}$$

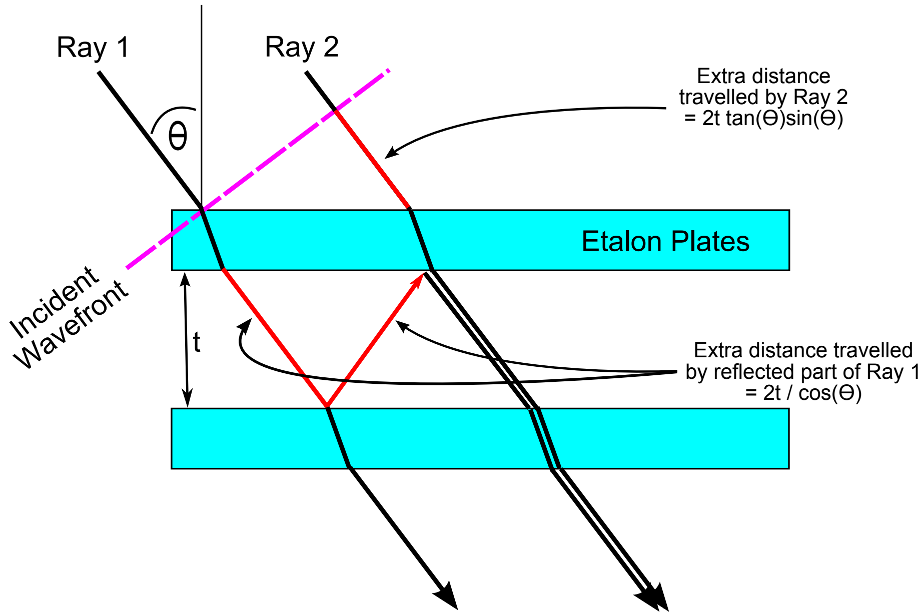


Figure 3.1.2: The paths of two light rays, originating on the same wavefront, with wavelength λ and incidence angle θ relative to the plate normal, are shown through the etalon. Ray 1 becomes superposed on Ray 2 after reflection at the lower, and then upper, etalon plate. The segments of each path which are not common to both rays are highlighted in red. The difference between these segments determines the relative phase difference between the emergent rays.

The $\phi_r(\lambda)$ term represents the phase-change upon reflection at the plate coating, which is usually small, and can be accounted for by a small adjustment to t. Setting $\xi = \cos(\theta)$, and accounting for the phase-change term as above, Equation 3.1.2 becomes:

$$\Delta\varphi = \frac{4\pi\mu t\xi}{\lambda} \tag{3.1.3}$$

Note that the form of Equation 3.1.3 appears non-intuitive in that the phase difference actually *decreases* with increasing incidence angle, whereas the geometry in Figure 3.1.2 might suggest the opposite. It is also important to note that interference occurs between rays from adjacent points on the wavefront, and therefore occurs everywhere these adjacent rays are superposed.

After passing through the etalon, the multiple, parallel emergent rays can be recombined by a focusing lens and imaged on to a detector located at the focal plane of the lens. The intensity recorded by the detector will depend upon the angular modulation introduced in the etalon, and will therefore maximise for:

$$\Delta\varphi = 2\pi m, \quad m = 0, 1, 2, ... \tag{3.1.4}$$

The integer m is termed the 'order' of interference, and will herein be treated as a continuous variable. Maxima in transmitted intensity therefore coincide with integral interference order. For a given λ, t and μ, the interference pattern recorded by the detector is dependent only on incidence angle θ. When a real-image of the incident light is produced at the detector, this incidence angle is mapped to radius on the detector. By axial symmetry (about the optical axis) contours of constant m map out points of constant radius at the detector, and therefore transmission maxima (or integral m) form a series of concentric, circular fringes. Two examples of laser interference fringes recorded by the Mawson SDI are shown in Figure 3.1.3. From Equations 3.1.3 and 3.1.4, the order of interference is given by:

$$m = \frac{2\mu t\xi}{\lambda} \tag{3.1.5}$$

The transmission of an ideal etalon may then be calculated by summing the amplitudes of each of the transmitted rays at a given order m. Assume that the wavefront is incident with amplitude A_0, and that the (complex) reflection and transmission coefficients of the inner etalon plate surfaces are r' and t' respectively. Relative to the incident ray, the transmitted rays will have amplitudes:

$$A_0 t'^2 e^{i\varphi},\ A_0 t'^2 r'^2 e^{2i\varphi},\ A_0 t'^2 r'^4 e^{3i\varphi}, ... \tag{3.1.6}$$

The sum of these amplitudes ($A_T(\varphi)$, the transmitted amplitude) forms a geometric series:

$$A_T(\varphi) = \sum_{k=0}^{N} A_0 t'^2 (r^2 e^{i\varphi})^k \tag{3.1.7}$$

At near-normal incidence, the number of reflections experienced by each ray is essentially infinite, and Equation 3.1.7 converges to:

$$A_T(\varphi) = \frac{A_0 t'^2}{1 - r'^2 e^{i\varphi}} \tag{3.1.8}$$

The transmittance is then given by the ratio of the transmitted to the incident intensity:

$$\begin{aligned} T_r(\varphi) &= \frac{A_T A_T^*}{A_0 A_0^*} \\ &= \frac{(t'^2)^2}{1 + (r'^2)^2 - 2r'^2 \cos(\varphi)} \end{aligned} \tag{3.1.9}$$

By assuming that no absorption of radiation occurs at the plate coatings, we have that $R = r'^2$ and $T = t'^2 = 1 - R$, which allows Equation 3.1.9 to be written as:

$$T_r(\varphi) = \frac{T^2}{1 + R^2 + 2R[1 - \cos(\varphi)]} \tag{3.1.10}$$

Which, after dividing through by $(1 - R)^2$, leads to the well-know Airy function (Hecht, 1974), which is expressed here in terms of the order of interference m:

$$A(m) = \frac{1}{1 + \frac{4R}{(1-R)^2}[\sin^2(m\pi)]} \tag{3.1.11}$$

The above equation reaches its maximum value at integral values of m. An important quantity relating to the interference fringe pattern given by Equation 3.1.11 is the *free spectral range*, Δ_λ. If the etalon is illuminated by two narrow-band sources, of wavelength λ_0 and $\lambda_1 = \lambda_0 + \Delta_\lambda$, then the (m-1)$^{\text{th}}$ order fringe of λ_1 will overlap the m$^{\text{th}}$ order

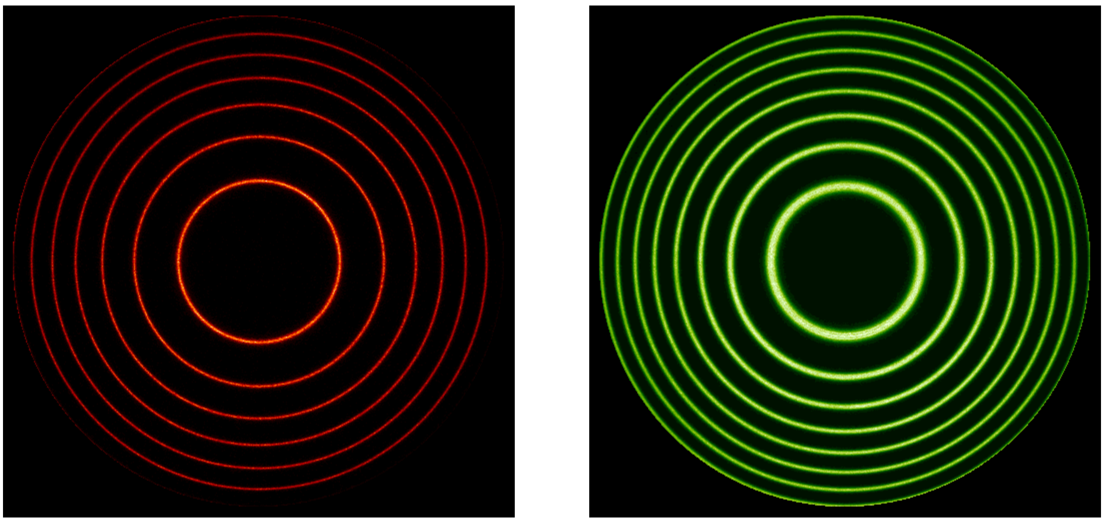


Figure 3.1.3: Examples of two sets of interference fringe images recorded by the Mawson SDI. Source illumination came from a laser in each case. The fringes on the left are due to a frequency stabilised laser emitting at 632.8 nm, whereas the fringes on the right are from a non-stabilised laser emitting at 543.5 nm. These are false colour images.

fringe of λ_0. The free spectral range thus corresponds to the wavelength interval between successive transmission peaks at constant μ, t and ξ. From Equation 3.1.5:

$$\begin{aligned} m_0\lambda_0 &= m_1\lambda_1 \\ \Rightarrow \quad m\lambda &= (m-1)(\lambda+\Delta_\lambda) \\ \Rightarrow \quad \Delta_\lambda &= \frac{\lambda}{m-1} \\ &\simeq \frac{\lambda^2}{2\mu t\xi} \end{aligned} \tag{3.1.12}$$

The full width at half maximum (FWHM) of a transmission peak, in 'units' of m, is given to a good approximation by (Wilksch, 1985):

$$\delta m_A = \frac{(1-R)}{\pi\sqrt{R}} \tag{3.1.13}$$

A measure of the performance of the etalon is given by the reflective finesse, N_R, defined as the ratio of the free spectral range to the width of a transmission peak, which is simply the reciprocal of δm_A:

$$\begin{aligned} N_R &= \frac{\Delta_\lambda}{\delta\lambda} \\ &= \frac{\pi\sqrt{R}}{(1-R)} \end{aligned} \tag{3.1.14}$$

The form of the Airy function is shown in Figure 3.1.4, along with graphical illustrations of finesse, peak width and free spectral range.

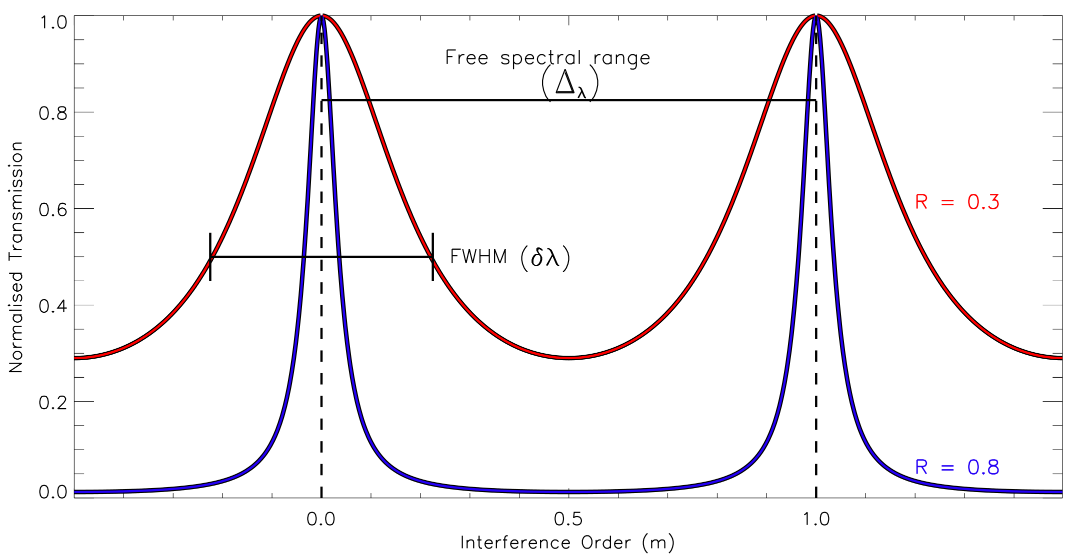


Figure 3.1.4: Normalised Airy function dependence on interference order m, for reflectances $R = 0.3$ (red curve) and $R = 0.8$ (blue curve). The parameters Δ_λ and $\delta\lambda$ are shown for $R = 0.3$.

3.2 The Real Etalon

Implicit in the above discussion is the notion of an 'ideal' etalon, which transmits normally incident monochromatic radiation at a constant order everywhere across the etalon plate aperture. In practice, any real etalon will incorporate many small imperfections whose cumulative effect is to broaden the resulting fringe pattern. Imperfections in plate polishing, coating and alignment all have the effect of varying the effective plate separation across the etalon, which, in addition to the necessarily finite solid angles subtended by each pixel on the detector, will result in a broadening of the recorded spectrum relative to the source spectrum. Additional effects due to imperfect uniformity of the plate coatings, thermal and mechanical stress on the etalon plates, and vibrations, may also contribute.

Even an ideal etalon will result in some broadening. In Section 3.1 the transmission of an ideal etalon was described by the Airy function (Equation 3.1.11), and the width of the transmission peaks were seen to depend on the reflectance of the plate-coatings (Figure 3.1.4). The transmission peaks of an ideal etalon viewing a monochromatic source of radiation will thus have a finite width, given by Equation 3.1.13. This effect is called reflectance broadening.

The extension to a non-ideal etalon follows the formulation given by Wilksch (1985). Looking first at the effect of plate defects, it is reasonable to assume, for a spectrometer operating at near-normal incidence (as is the case with the Mawson SDI), that the imperfect etalon can be treated as a juxtaposition of a large number of elementary etalons, whose plate separation is given by t_d relative to the mean plate separation t_0. The contribution ds to the total etalon area S by elementary etalons having a plate separation between $t_0 + t_d$ and $t_0 + t_d + dt_d$ is then given by:

$$ds = D(t_d)dt_d \tag{3.2.1}$$

where $D(t_d)$ is referred to as the defect function. It follows that the integral of $D(t_d)$ over all separations t is the total area S. For light of wavelength λ and incidence angle θ normal to an elementary etalon whose separation is exactly t_0 ($t_d = 0$), the transmission is given by the Airy function at order m_0, $A(m_0)$. For $t_d \neq 0$, the order of interference will be shifted to $m - m_d$, where m_d is given by:

$$m_d = -\frac{t_d}{t_\Delta} \tag{3.2.2}$$

where t_Δ is the change in t corresponding to an integral change in m, and the negative sign is introduced so that the final expression will have the form of a convolution product (Equation 3.2.6). The fractional contribution to the total transmission E by the elementary area ds can therefore be written as:

$$dE = A(m - m_d)\frac{D(t_d)}{S}dt_d \tag{3.2.3}$$

Introducing the unit defect function defined by:

$$\hat{D}(m_d)dm_d = \frac{ds}{S} = \frac{D(t_d)}{S}dt_d \tag{3.2.4}$$

allows Equation 3.2.3 to be written:

$$dE = A(m - m_d)\hat{D}(m_d)dm_d \tag{3.2.5}$$

The total transmission from all elementary etalons is then the integral over all defect-induced variations in m, and is called the etalon transmission:

$$E(m) = \int_{-\infty}^{\infty} A(m - m_d)\hat{D}(m_d)dm_d \tag{3.2.6}$$

Note that the integration limits of $(-\infty, \infty)$ in this equation are a mathematical convenience only; in reality the defect function will go to zero very quickly either side of $m_d = 0$. Equation 3.2.6 defines a convolution product:

$$E(m) = \hat{D}(m) * A(m) \tag{3.2.7}$$

where the $*$ symbol defines the convolution operation. It is convenient to refer to $E(m)$ as the as the etalon function. It gives the transmission of the etalon as a function of the interference order m taking into account plate defects. If the FWHM of $D(t_d)$ is given by δt_d, then the FWHM of $\hat{D}(m_d)$ is given by:

$$\delta m_d = \frac{\delta t_d}{t_\Delta} \tag{3.2.8}$$

A detector placed in the focal plane of the lens must necessarily subtend a finite range of angles $d\xi$ ($\xi = \cos(\theta)$) if any signal is to be recorded. From Equation 3.1.5, any range of angles $d\xi$ will result in radiation transmitted at a range of interference orders dm, which will broaden the recorded spectrum. In the case of the Mawson SDI, each pixel on the camera's CCD detector can be treated as an individual detector which records its own spectrum. Let a pixel-dependent parameter be denoted with the subscript p, where it is understood that p also corresponds to a particular (x, y) location on the detector. Consider a pixel which subtends a small range of incidence angles relative to the plate normal direction, denoted by $\xi_0 = 1$. If ξ_f denotes an incidence angle (relative to ξ_0) subtended by the pixel, then rays having the cosines of their angles between ξ_f and $\xi_f + d\xi_f$ will constitute a fraction ($d\omega_p$) of the total solid angle subtended by the pixel at the etalon (Ω_p) and will be a function of ξ_f:

$$d\omega_p = F_p(\xi_f)d\xi_f \tag{3.2.9}$$

The function $F_p(\xi_f)$ is commonly called the aperture function, whose integral over all incidence angles ξ_f is Ω_p. If the rays at normal incidence ($\xi = 1$, for which $\xi_f = 0$)

correspond to transmission at order m_0, then the change in m due to viewing an incidence angle ξ_f will be given by:

$$m_f = -\frac{\xi_f}{\xi_\Delta} \tag{3.2.10}$$

where, analogous to the discussion of plate defects, ξ_Δ is the change in incidence angle which will result in an integral change in m. Introducing the unit aperture function:

$$\hat{F}_p(m_f)dm_f = \frac{d\omega_p}{\Omega_p} = \frac{F_p(\xi_f)}{\Omega_p}d\xi_f \tag{3.2.11}$$

the contribution to the total transmission for the pixel, at wavelength λ, from the rays having the cosines of their angles within the small range between ξ_f and $\xi_f + d\xi_f$ is given by:

$$dI_p = E(m - m_f)\hat{F}_p(m_f)dm_f \tag{3.2.12}$$

where $E(m)$ is the etalon function introduced previously (Equation 3.2.7). The integral over all m_f gives the total transmission for the pixel:

$$I_p(m) = \int_{-\infty}^{\infty} E(m - m_f)\hat{F}_p(m_f)dm_f \tag{3.2.13}$$

which may be written as the convolution product:

$$I_p(m) = \hat{F}_p(m) * E(m) \tag{3.2.14}$$

The function $I_p(m)$ is called the instrument function, and describes the broadening effect of etalon plate defects and the finite collecting area of each pixel. In this case, the derived instrument function applies to a single pixel on the detector, each of which records an individual intensity variation during a scan. Each pixel will have a unique aperture function, although due to symmetry these will be similar for pixels at similar distances from the optical axis, and since the etalon used in this work operates at near-normal incidence, Ω_p will vary by only a small amount across the detector. In addition, the pixels themselves are typically much smaller than the width of the 630.0 nm airglow fringe, and therefore aperture broadening is very small (in contrast to older-style "pinhole scanners", where aperture broadening is comparable to the width of the etalon function).

In the case of the Mawson SDI, the spectra from a group of pixels are co-added during a scan (details of the Mawson SDI are given in Chapter 4, where these groups are referred to as "zones"). This group of pixels, or zone, which is defined in software, forms an effective collecting area (or an effective detector) larger than an individual pixel, increasing the amount of radiation collected per unit time. However care must be taken when co-adding the intensity variations from different pixels to account for the fact that the Airy function will in general maximise at a different value of the scanning parameter (see Section 3.3) for different pixels, due to different radial distances from the optical axis. A scheme for carrying out this co-addition was given by Conde and Smith (1997), for which the details are provided in Section 4.5. This scheme effectively removes the aperture broadening that would otherwise apply to a single detector subtending the same range of incidence angles as the overall zone does.

3.3 Spectral Scanning

From Figure 3.1.4 it can be seen that varying the interference order m over some range of values Δm will result in a corresponding intensity variation (described by $A(m)$) for that range. Equation 3.1.5 shows that, for incident radiation of a fixed wavelength λ, this scanning can be achieved by variation of any of the parameters t (separation scanning), μ (refractive index scanning) or ξ (spatial scanning). As the particular parameter is varied, the spectrometer is progressively 'tuned' to maximise the transmission of radiation at progressively different wavelengths. Each discrete step in a scan is interpreted in this way; that is, we ask, at each discrete step, what wavelength would be required to produce a maximum in the transmission intensity for that particular value of the scanning parameter.

Alternatively, m can be thought of as the order to which a fixed wavelength λ would correspond as t, μ or ξ are varied (Wilksch, 1985). Thus, the interval Δm can always be interpreted as a wavelength interval $\Delta\lambda$, and one which, for a linear variation in the scanning parameter, is precisely linear in λ (Wilksch, 1985). In order that the recorded spectrum contain only one intensity maximum (or wavelength 'peak'), the range over which m is scanned can be intentionally limited to one order of interference, thus $\Delta m = 1$ and the wavelength interval is equal to one free spectral range ($\Delta\lambda = \Delta_\lambda$).

Consider a spectrometer viewing an extended source of radiation whose spectral radiance is given by $B(\lambda)$ W.sr^{-1}.m^{-2}.nm^{-1}, about a peak wavelength λ_0. The radiance[1] dl in the wavelength interval between $\lambda_0 + \lambda_b$ and $\lambda_0 + \lambda_b + d\lambda_b$ is given by:

$$dl = B(\lambda_b)d\lambda_b \tag{3.3.1}$$

The integral over all λ must therefore equal the total radiance of the source, denoted by L. If λ_0 is transmitted at order m_0, then $\lambda_0 + \lambda_b$ will be transmitted at order $m_0 - m_\lambda$, where:

$$m_\lambda = \frac{\lambda_b}{\Delta_\lambda} \tag{3.3.2}$$

where Δ_λ is the free spectral range. The unit source function, defined by:

$$\hat{B}(m_\lambda)dm_\lambda = \frac{dl}{L} = \frac{B(\lambda_b)}{L}d\lambda_b \tag{3.3.3}$$

allows the fraction of the total radiance transmitted in the range $d\lambda_b$ about $\lambda_0 + \lambda_b$ to be

[1]Here the term radiance refers to the energy per unit time emitted from a surface of unit area into a given solid angle, and therefore has units of W.sr^{-1}.m^{-2}. The *spectral* radiance is then the radiance per unit wavelength, with units W.sr^{-1}.m^{-2}.nm^{-1}. Radiance is equal to the integral of the spectral radiance over a given wavelength interval.

written:

$$dY_p = I_p(m - m_\lambda)\hat{B}(m_\lambda)dm_\lambda \tag{3.3.4}$$

and thus:

$$\begin{aligned} Y_p(m) &= \int_{-\infty}^{\infty} I_p(m - m_\lambda)\hat{B}(m_\lambda)dm_\lambda \\ &= \hat{B}(m) * I_p(m) \end{aligned} \tag{3.3.5}$$

Again, the integration limits in Equation 3.3.5 are a mathematical convenience. In practice the integral spans (and is weighted by) the spectral response of the detector. The function $Y_p(m)$ is the recorded intensity variation, and represents the fraction of the total source radiance which is transmitted to the detector pixel. It is a function of m, the order of interference, which here represents a 'tuning parameter' which is used to vary the wavelength at which the transmission of radiation maximises. From the discussion at the beginning of this section, m represents the order to which a fixed wavelength corresponds as the spectrometer is tuned by varying one of the scanning parameters t, μ or ξ, and from Equations 3.3.2 and 3.1.12 is related to λ by:

$$\frac{m}{m_0} = \frac{\lambda}{\lambda_0} \tag{3.3.6}$$

Since the Airy function is periodic in m, with period unity, the scanning parameter is typically varied over a range equivalent to $\Delta m = 1$. We therefore have that:

$$m = \frac{m_0}{\lambda_0}\lambda = \frac{\lambda}{\Delta_\lambda} \tag{3.3.7}$$

And thus the wavelength interval corresponding to a scan over $\Delta m = 1$ is:

$$\begin{aligned} \lambda - \lambda_0 &= (m - m_0)\Delta_\lambda \\ &= \Delta_\lambda \end{aligned} \tag{3.3.8}$$

If the scanning parameter is varied in discrete steps of constant size (as is the case with the Mawson SDI) then the change in wavelength corresponding to maximum transmission is also constant at each step, and the resulting intensity variation $Y_p(\lambda - \lambda_0) = Y_p(\Delta_\lambda[m - m_0])$ is interpreted as a set of samples of the source wavelength spectrum convolved with the instrument transmission function. The total energy per unit time transmitted to the pixel will be:

$$\Phi_p(\lambda) = LS\Omega_p\tau Y_p(\lambda) \tag{3.3.9}$$

where τ is a coefficient which takes into account the net effect of absorption and scattering in the instrument. The actual 'data number' recorded output by the detector at the given pixel will be modified by the quantum efficiency of the detector, ϵ, and may be written as:

$$\Theta_p(\lambda) = k\epsilon\Phi_p(\lambda) \tag{3.3.10}$$

where the constant k takes care of the conversion of units between incident power (J.s^{-1}) and the data number which is physically recorded by the detector.

Chapter 4

The Mawson Scanning Doppler Imager

4.1 Introduction

After initial testing at La Trobe University, the Mawson Scanning Doppler Imager (SDI) was disassembled and shipped to Davis station, Antarctica, during the 2006/2007 summer season. From Davis the crates containing the instrument parts were flown to a site on the Antarctic plateau approximately 5 kilometers south of Mawson, before being transported to Mawson station by land (Figure 4.1.1). The crates were unloaded in the Mawson Aeronomy building, and installation began once the frame of a previous Fabry-Perot spectrometer had been dismantled. The design of the instrument was such that it could be easily assembled by myself and Mr. Theo' Davies. The frame consisted of lower and upper sections. The upper section was assembled first, then lifted by a winch (lower left image of Figure 4.1.1) and held suspended while the lower section was assembled beneath it. Once the instrument frame was in place, a smaller winch attached to the upper section of the frame was then used to raise the optical components which were suspended in the centre of the frame. The optical components were assembled in a similar manner to the frame: components higher up in the optical assembly (a schematic of which is shown in Figure 4.2.1) were installed first, then hoisted up using the winch, and components lower down were added on sequentially.

Once the optical components were assembled, the winch was used to raise the instrument into its operating position, in which the optical components were held in place by a collar which was supported at three radially-separated points by 'legs' attached to the upper section of the frame. In this position the all-sky lens at the top of the optical assembly was raised above the level of the building's roof, and protected by a glass dome. The image at the bottom right of Figure 4.1.1 shows the fully assembled instrument with the optical components in the lowered position, supported by the winch. Directly behind the instrument in this image is an electronics rack which holds the etalon control unit, filter wheel control unit, servo-motor power supply and calibration lasers (which couple to the instrument via optical fibre). To the right of the instrument is the computer which runs the instrument control software (described in Section 4.11). A water reservoir and circulating pump (not visible in this image), used to circulate water through the EMCCD camera (just visible at the bottom of the optical assembly), is located on the wooden platform to left of the instrument. The instrument frame itself rests on a concrete plinth which is isolated from the main building, an important consideration given the consistently strong winds experienced at Mawson, and the consequent building vibration.

4.2 Instrument Design

The Mawson SDI is a separation-scanned, all-sky (wide-field) imaging spectrometer, of the type described by Conde and Smith (1997). A schematic is shown in Figure 4.2.1. An all-sky lens collects light from a large (approximately 144° full-angle) field-of-view of the sky. The airglow wavelength of interest is selected by a six-position filter wheel located beneath the all-sky lens. This filter wheel is controlled in software, and currently carries three interference filters, two at wavelengths of 630.0 nm and one at 557.7 nm.

Calibration light from up to four different sources can be introduced directly beneath the filter wheel via an integrating sphere. Selection of the calibration source is made possible by a servo-driven shutter mechanism, controlled in software. A second software-controlled servo drives a small mirror used to switch the viewing direction between the sky and the integrating sphere. Two calibration light sources were available during 2007, both Helium-Neon (HeNe) lasers, one at a wavelength of 543.5 nm and the other (frequency-stabilised) at 632.8 nm.

Light from the desired source (sky or calibration laser) passes through a field-lens and a Taylor triplet collimator before entering a 150 mm aperture, capacitance-stabilised Fabry-Perot etalon. The gap between the etalon plates is piezoelectrically scannable over approximately 2.5 orders of interference at λ630.0 nm, about a nominal gap of 25 mm (key etalon properties are listed in Table 4.1, along with the focal lengths of the lenses comprising the fore-optics). In its usual configuration, a complete scan over one order of interference in 128 discrete steps requires approximately 40 seconds.

Light emerging from the etalon is angularly modulated by approximately 8 orders of interference fringes (for example fringes see Figure 3.1.3). A very fast, 300 mm f/2 Nikon telephoto lens images these fringes onto an Andor "Ixon" electron-multiplying CCD camera (512 x 512 pixels). The camera is cooled by a Peltier thermoelectric cooler, and heat is removed from the hot side of the Peltier junction by circulating liquid water, allowing the camera to be cooled to around -80°C. The quantum efficiency of the camera is greater than 90% throughout most of the visible spectrum, with essentially zero read noise due to the deep cooling, while electron-multiplication allows for readout at high frame rates without any significant penalty due to read noise.

Figure 4.1.1: Sequence of images (ordered from top left to bottom right) showing the stages of installation of the Mawson SDI. They show, in order: boxes containing instrument parts unloaded from aircraft onto the Antarctic plateau; transfer of boxes onto a truck for transport to Mawson station; arrival of boxes at Mawson station and transfer onto a utility vehicle for transport to the Aeronomy building; unloading of boxes into the Aeronomy building; raising of the top-level frame to enable placement of lower-level frame beneath; the fully assembled instrument with optical components (suspended in the centre of the light-blue frame) shown in the lowered position. Images courtesy of Mr. Theo' Davies.

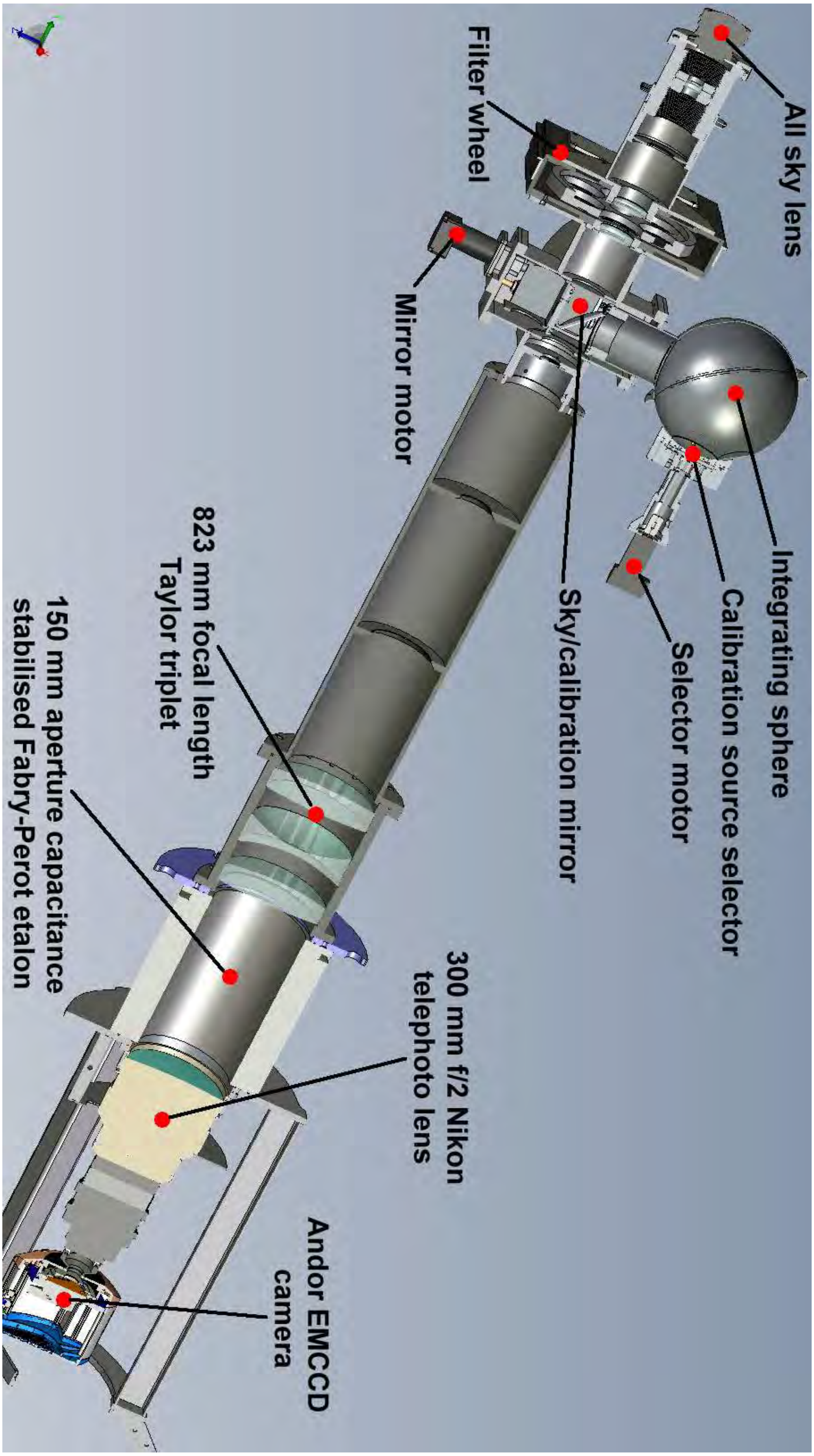


Figure 4.2.1: Schematic of the Mawson SDI.

Table 4.1: Etalon and Lens Specifications

Etalon	
Nominal gap	25mm
Aperture	150mm
Max light incidence angle	0.93°
Reflective finesse	13.5
Lens	Effective Focal Length (mm)
All-sky lens	8
Upper field lens	50
Filter upper collimator	250
Filter lower collimator	300
Lower field lens	250
Taylor triplet	823
Nikon f/2 telephoto	300

4.3 Etalon Parallelism

The separation between the etalon plates is controlled by three capacitive sensor/piezoelectric transducer pairs (each pair will henceforth be referred to as a 'leg', and numbered 1, 2, and 3) placed between the etalon plates and at equal angles around the circumference of the plates (see panel A on the right of Figure 4.3.1). The gap is physically adjusted by applying a voltage across the piezoelectric transducers. The capacitive sensors consist of a reference capacitor attached to the lower plate, and a 'driven' capacitor, of which half is attached to the top of the reference capacitor (and separated from it by a spacer) and the other half is attached to the upper plate. Any displacement of the etalon plates therefore varies the capacitance ratio of the (nominally equal) driven and reference capacitors (Hicks et al., 1974) which form part of an AC excited capacitance bridge. The output of the capacitive sensors is a voltage which is proportional to the difference between the reference and driven capacitors (and hence the etalon plate separation), and an error-integrator is used to compare this output voltage with a control voltage, and drive the piezoelectric transducers until the difference between the capacitive sensor and control voltages is minimal (Rees et al., 1981). The control voltage thus corresponds to a demanded etalon plate separation.

A control unit, interfaced to the control computer via a serial link, interprets commands sent to it by the computer and drives the individual leg voltages (and hence leg heights) until the sensed leg spacings match those commanded by the computer. It is possible to assign to each leg a 12-bit number (0 - 4095) representing a demanded spacing at the leg. This 12-bit number will henceforth be referred to as a 'digital' voltage, to differentiate it from the resulting physical voltage to which the piezoelectric transducers are driven by the servo-loop described above. A digital voltage corresponds to the control voltage mentioned above, and should thus be interpreted as a demanded leg spacing (or etalon plate separation at that leg).

For optimum performance the etalon plates must be aligned as close as possible to parallel. Non-parallel plates will result in broadened and distorted interference fringes.

The width of such degraded fringes varies with azimuthal angle around the fringe. The etalon of the Mawson SDI is scanned by varying the plate separation, and the etalon plates must be kept parallel throughout the scan. It is thus necessary to ensure that the starting value of each digital leg voltage is such that the resulting fringes are free from this distortion. In practice this is achieved by individually varying the digital voltages applied to each leg and identifying by eye the sharpest fringe image (of a laser calibration source) which results. The individual (digital) leg voltages which produce this 'best' fringe image are then taken as the starting values (or offsets) for each leg in any subsequent etalon scan, which for leg number n may be written as $V_n(0)$.

Since the individual capacitive sensors will not respond identically to changes in plate separation, it is also necessary to measure any leg-to-leg variation and account for it during any scan of the etalon, to ensure that the plates remain as parallel as possible over the entire scan range. As mentioned previously, the servos can be commanded with a range of digital values from 0 through 4095. Each capacitive displacement sensor is expected to respond most linearly throughout the middle of this range. The first parallelism step therefore, outlined above, is performed about a nominal digital voltage of around 1000, or around 25% of the possible range. This parallelism step is then repeated about a nominal digital voltage of around 3000, or approximately 75% of the possible range. If the difference in voltage at each fraction of the scan range (or nominal voltage) for leg number 1 (for example) is labelled ΔV_1, then the response of leg number n to a unit increment in digital voltage is calculated relative to the response of leg number 1, by defining the relative gain R_n:

$$R_n = \frac{\Delta V_n}{\Delta V_1} \tag{4.3.1}$$

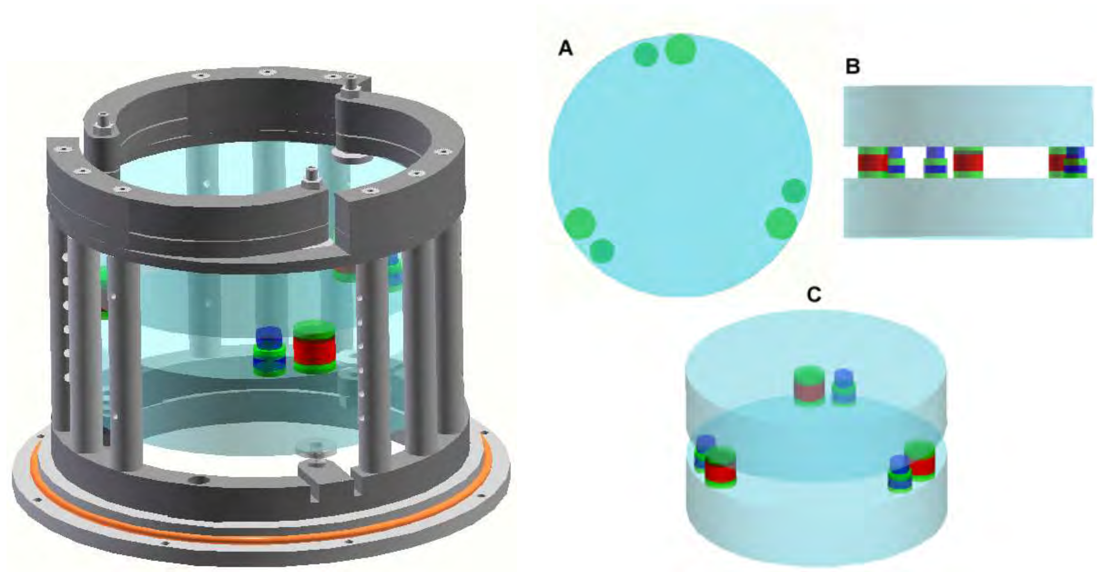


Figure 4.3.1: Left: 3D model of the etalon and supporting cage. Right: schematic views of the etalon plates, piezoelectric transducers (green- and red-shaded cylinders) and capacitive sensors (green- and blue-shaded cylinders). Images courtesy of Dr. David Rees, Hovemere Ltd. (private communication (2009)).

By this definition, R_1 is equal to unity, and, for a given digital voltage increment ΔV_1 applied to leg number 1, $R_n \times \Delta V_1$ gives the digital voltage increment which should be applied to leg number n in order that the etalon plates remain parallel after the increment, accounting for the slightly different responses of the capacitive displacement sensors. For the etalon used in this work, values for R_n of around 1.05 were typical for legs 2 and 3.

4.4 Scan Range

Before any separation scanning can take place, it is necessary to determine how much digital voltage must be applied to each of the legs to change the etalon plate separation by one order of interference. Assuming that the SDI is observing a narrow-band monochromatic light source, increasing the plate separation results in interference fringes of increased radius. If an image is recorded of the fringes at some initial plate separation t_0, and the plate spacing gradually increased, the fringe radii will increase accordingly, until a new plate separation t is reached at which an (m+1)$^{\text{th}}$ order fringe is now overlapping an m$^{\text{th}}$ order fringe of the reference image. This change in etalon plate spacing Δt (with wavelength held fixed) is equivalent to varying the incident wavelength (with the etalon gap held fixed) by one free spectral range (Δ_λ) of the etalon (Section 3.1).

The aim is thus to determine the digital voltage which must be applied to each leg such that $\Delta m = 1$ (refer to Figure 4.4.1). First, a laser calibration source is selected, and the etalon leg voltages set to their initial offsets $V_n(0)$. A 512x512 pixel reference image I_0 is recorded of the interference fringes at this initial plate separation. A series of images I_p are then recorded, each at a different set of leg voltages $V_n(p)$ defined by:

$$V_n(p) = V_n(0) + pR_n\delta V, \qquad p = 1, 2, 3...P \tag{4.4.1}$$

The number (P) and size (δV) of the voltage increments are chosen such that the final change in plate separation is slightly more than one free spectral range. Each image I_p is cross-correlated with the reference image I_0 so that:

$$X(p) = \sum_{x=0}^{N_x-1} \sum_{y=0}^{N_y-1} [I_0(x, y) \times I_p(x, y)] \tag{4.4.2}$$

gives the magnitude of this correlation, where $N_x = N_y = 512$, and (x, y) denotes a pixel location in the image. Since fringes recorded at plate spacings separated by one order of interference should appear (practically) identical, the correlation product $X(p)$ will maximise for p corresponding to any integral multiple of one free spectral range. A suitable polynomial function fitted to $X(p)$ is used to solve for p_{max}, and the total change in applied digital voltage found from Equation 4.4.1:

$$\Delta V_n^{fsr} = p_{max}R_n\delta V \tag{4.4.3}$$

ΔV_n^{fsr} then represents the leg-dependent digital voltage which must be applied to vary

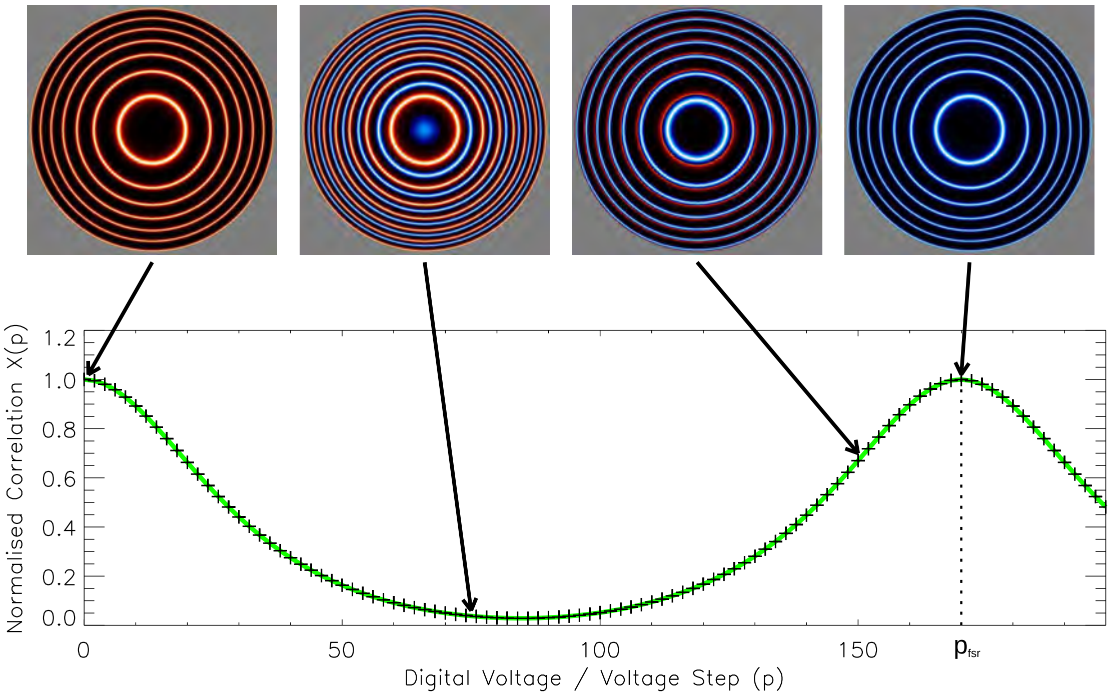


Figure 4.4.1: Illustration showing how the $X(p)$ series is generated from the cross-correlation of a reference fringe image (shown in red) and a scanning fringe (blue), and used to find the digital voltage needed to vary the plate separation by one free spectral range. Images of four steps in the scan, and their corresponding values of $X(p)$ (normalised) are shown for this simulated case.

the etalon plate separation by one order of interference, at wavelength λ_0. For this number to be useful, it must be applicable to wavelengths other than the calibration wavelength used to calculate it. From Equation 3.1.5, the change in t corresponding to one order of interference ($\Delta m = 1$) is given at normal incidence by:

$$\begin{aligned} \Delta t_{\lambda_0} &= \frac{\lambda_0}{2\mu} \\ \Rightarrow \quad \Delta t_{\lambda_1} &= \Delta t_{\lambda_0} \frac{\lambda_1}{\lambda_0} \end{aligned} \tag{4.4.4}$$

Since ΔV_n^{fsr} is directly proportional to t, the digital voltage which must be applied to each leg at some new wavelength λ_1 will be 'scaled' by the factor $\frac{\lambda_1}{\lambda_0}$.

When acquiring spectra, a single scan of the etalon over one order of interference is performed in a series of Q steps, called "channels". The number of channels in a scan is configurable, but is commonly chosen to be 128. Thus one complete scan at wavelength λ_1 will require Q voltage increments of size:

$$S_n = \frac{\Delta V_n^{fsr}}{Q} \frac{\lambda_1}{\lambda_0} \tag{4.4.5}$$

In practice it is not necessary to cover the whole range of p values from 1 through P when searching for p_{max}. All that is required is that the range of p values chosen records the

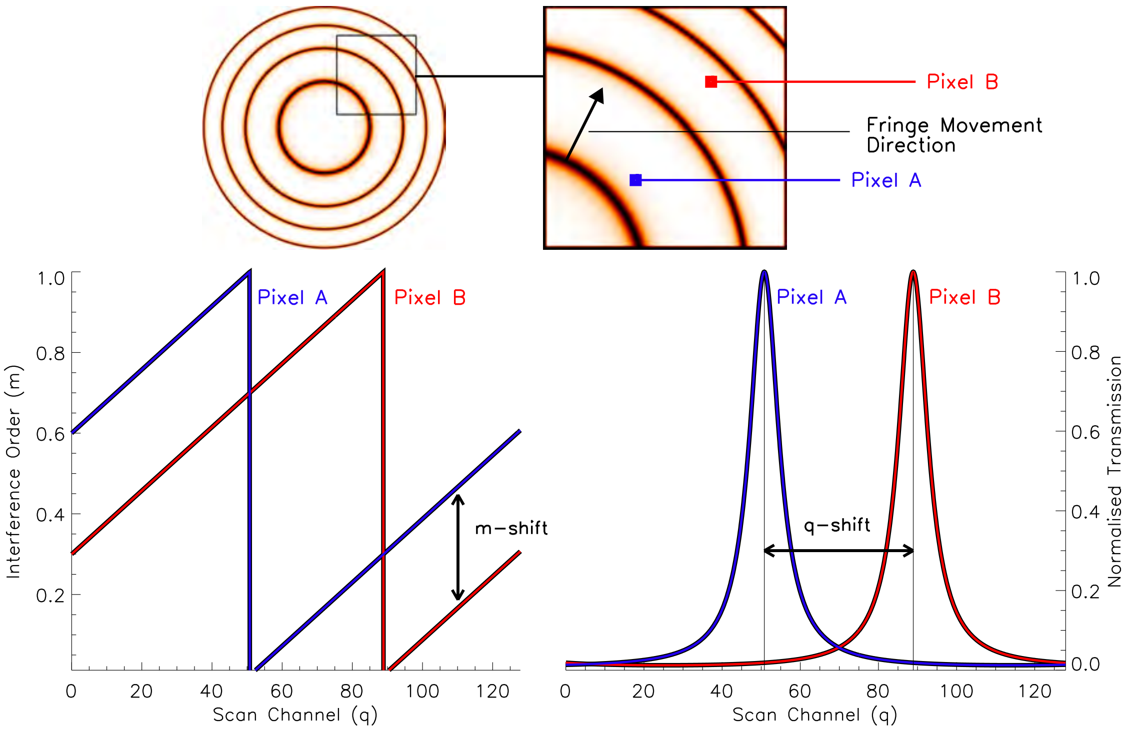


Figure 4.5.1: The variation of recorded spectra between two pixel locations (x, y) on the detector. Pixel locations relative to interference fringes are shown in the top panels. The Airy function maximises for integral values of m. Pixel A, since it starts at a higher fractional interference order than Pixel B (left panel) , will reach record the fringe peak at a lower channel number than Pixel B (right panel), resulting in a phase-shift between the spectra recorded by the two pixels.

maximum in $X(p)$, so that p_{max} may be solved for. Also, it often useful to co-add the $X(p)$ series from multiple scans, in order to obtain a sufficiently high signal-noise ratio such that the uncertainty in the fitted polynomial function is minimal, and to suppress the effects of one-off events like cosmic ray strikes on the detector.

4.5 Phase Mapping

A positive power lens placed after the etalon transforms the angular distribution of transmitted intensity to a spatial distribution at the focal plane of the lens. Since m also varies with θ (and hence with (x, y) position on the detector), pixels at different locations will not necessarily share the same value of m at a particular value of the scanning parameter (in this case t). For an etalon operating with a very narrow range of incidence angles (as is the case with the Mawson spectrometer), the variation of m with plate separation will be, to a very good approximation, equal for pixels mapping to different values of θ. Thus any difference in the value of m between pixels with different values of θ (which effectively means different radii) at the start of a scan will be preserved throughout the entire scan range.

To see the effect that a difference in interference order will have on the derived spectra, consider Figure 4.5.1. Since the Airy function is periodic in m with a period of unity, the important quantity here is not the absolute interference order m, but rather the fractional part of m, given by m mod 1. At the beginning of a scan (channel number 0, corresponding to some plate separation t_0), Pixel A is at a higher fractional order of interference than Pixel B. This means in practice that as the etalon plate separation is increased, the interference fringes (whose radii increase with increasing plate separation) will cross the radius of Pixel A before they cross that of Pixel B. Pixel A is in a sense closer to the fringes (as they appeared at the start of the scan) considering the direction they will move during the scan.

As the Airy function peaks for integral values of m, and Pixel A reaches this value at a lower channel number than Pixel B (left panel in Figure 4.5.1), the spectra recorded by Pixel A will peak at a lower channel number than Pixel B (right panel in Figure 4.5.1). Since a recorded intensity variation is interpreted as a function of wavelength regardless of the scanning parameter (Section 3.3), the phase-shift introduced by this variation in pixel location across the detector must be accounted for lest it be mistaken for a shift in

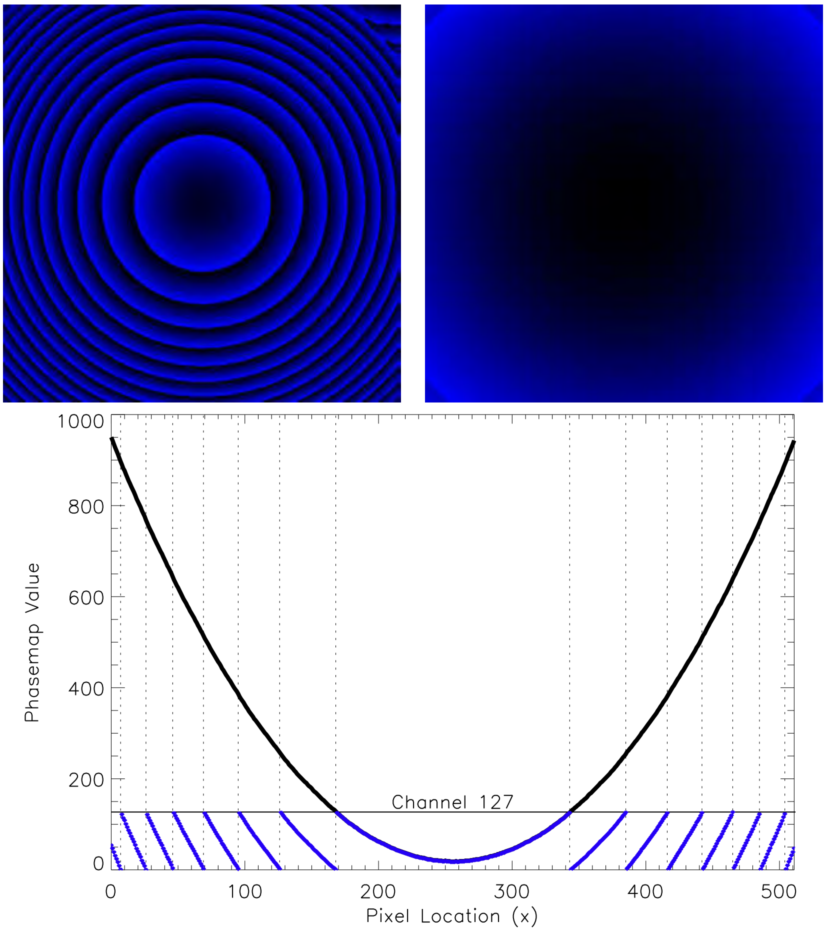


Figure 4.5.2: Example of a phase map used by the Mawson SDI to acquire spectra. Top left: the 'wrapped' version of the phase map as recorded directly by the instrument, darker shades corresponding to lower values. Top right: the 'unwrapped' version of the phase map, in which the discontinuities of the wrapped version have been removed, to allow for scaling in wavelength. Bottom: cross-section through the centre of both the wrapped (blue curve, discontinuous) and unwrapped (black curve) phase maps, in units of channels, for the case $Q = 128$.

incident wavelength. This can be done by determining, for each pixel on the detector, the value of the scanning parameter t (or channel number q which represents it) for which the spectrum recorded by the pixel maximises. Spectra can then be effectively 'shifted' by this amount along the channel/wavelength axis, ensuring that spectra recorded by different pixels, with the same incident wavelength, will produce spectra which peak at the same channel. The process of encoding the variation of spectral peak position with (x, y) position on the detector is herein referred to as "phase mapping".

Having determined the values of V_n^{fsr}, the phase map is recorded by viewing the calibration laser. The etalon plate separation is then varied over Q channels, where the separation at each channel q is given by:

$$V_n(q) = V_n(0) + qS_n \quad q = 0, 1, 2...Q - 1 \tag{4.5.1}$$

and S_n is defined by Equation 4.4.5. At each channel, a two-dimensional interference fringe image is recorded. Let this image be denoted by $B(x, y, q)$, where $0 \le x, y \le 511$. For each pixel in $B(x, y, q)$, the following summations are performed:

$$a(x, y) = \sum_{q=0}^{Q-1} B(x, y, q) \cos\left(\frac{2\pi q}{Q-1}\right) \tag{4.5.2}$$

$$b(x, y) = \sum_{q=0}^{Q-1} B(x, y, q) \sin\left(\frac{2\pi q}{Q-1}\right) \tag{4.5.3}$$

Since Equation 4.5.1 ensures a scan range of one order of interference, the intensity recorded by $B(x, y, q)$ will span at most one peak of Figure 3.1.4. The arrays $a(x, y)$ and $b(x, y)$ then contain the first two Fourier coefficients ('scaled' by a factor of $Q/2$) of the spectra recorded by each pixel. What is required for the phase map is the position of the peak within the scan range. From the Fourier coefficients calculated in Equations 4.5.2 and 4.5.3, the intensity variation at pixel (x_i, y_i) may be written as a function of the phase-angle ψ_q:

$$\psi_q = \frac{2\pi q}{Q-1} \tag{4.5.4}$$

which gives:

$$Y(x_i, y_i, \psi_q) \simeq a(x_i, y_i) \cos(\psi_q) + b(x_i, y_i) \sin(\psi_q) \tag{4.5.5}$$

The peak of the spectrum is obtained by solving:

$$0 = \frac{\partial}{\partial \psi_q} Y(x_i, y_i, \psi_q) \tag{4.5.6}$$

$$= a(x_i, y_i)\sin(\psi_q) - b(x_i, y_i)\cos(\psi_q) \tag{4.5.7}$$

$$\Rightarrow \quad \tan(\psi_q) = \frac{b(x_i, y_i)}{a(x_i, y_i)} \tag{4.5.8}$$

$$\Rightarrow \quad \psi_q = \psi(x_i, y_i) = \arctan\left(\frac{b(x_i, y_i)}{a(x_i, y_i)}\right) \tag{4.5.9}$$

The two-dimensional array $\psi(x, y)$ is referred to as a phase map. It encodes the variation across the detector of the phase-angle at which a spectrum peaks. The channel number q corresponding to phase-angle ψ_q is given simply by rearranging Equation 4.5.4:

$$q = \frac{\psi_q}{2\pi}(Q - 1) \tag{4.5.10}$$

which allows the $\psi(x, y)$ to be stored in the more convenient units of channel number. An example of a recorded phase map is shown in Figure 4.5.2. The figure in the top left corner of this group of images shows the variation of ψ across the detector as a variation in color, where darker shades correspond to lower values. Note that Equation 4.5.9 can only return a value for $\psi(x_i, y_i)$ between $[0, Q - 1]$, regardless of the order of interference, which explains the sharp discontinuities in the image. This is an important point, because in order to use the phase map at a wavelength other than that at which it was recorded (λ_0, the calibration laser wavelength), it needs to be scaled by an amount equal to:

$$\frac{\psi}{\psi_0} = \frac{\lambda_0}{\lambda} \tag{4.5.11}$$

which follows from Equations 3.1.4 and 3.1.5. The discontinuous form is the one appropriate for acquiring spectra, however no scaling can take place while these discontinuities exist. To understand why this is so, consider the phase map shown in the top left of Figure 4.5.2. Any pixel shaded in black corresponds to a spectrum which peaks at $q = 0$, ie. a pixel which samples a fringe peak at the initial plate separation, for which $\psi(x, y) = 0$. If a new light source is selected, such that the new wavelength is higher than λ_0 by some fraction of a free spectral range (for example), the new fringe peaks will all be shifted toward smaller radii. Thus a pixel which sampled a fringe peak at λ_0 will no longer do so at the new wavelength, which implies a corresponding shift to $\psi(x, y) \neq 0$ for that pixel. Clearly, simply scaling the discontinuous $\psi(x, y)$ by Equation 4.5.11 will produce no such change for this pixel, and so the discontinuities must first be removed from the recorded phase map before it can be applied (scaled) to light of a different wavelength. This process of removing phase map discontinuities is referred to as 'unwrapping'.

The unwrapped version of the phase map, $\tilde{\psi}(x, y)$, is shown in the top right corner of Figure 4.5.2, and a cross-section through both maps plotted below, for the case of $Q = 128$ channels. The unwrapped phase map can be freely scaled to a new wavelength using Equation 4.5.11. In order to use this scaled phase map to acquire spectra, it must be 'wrapped' back to the discontinuous form, a process which is computationally very simple compared to unwrapping, and is given simply by $\psi(x, y) = \tilde{\psi}(x, y) \bmod Q$. In practice only the integer part of $\psi(x, y)$ is used, since spectra are acquired at discrete values of the scanning parameter.

For a given pixel (x_i, y_i), the quantity $\psi(x_i, y_i)$ can be used to define the 'spectral channel' (q_{spec}) to which the signal recorded at $B(x_i, y_i, q)$ will be added. For the spectrum $Y(x_i, y_i, q)$ recorded by this pixel:

$$Y(x_i, y_i, q_{\text{spec}}) = Y(x_i, y_i, q_{\text{spec}}) + B(x_i, y_i, q) \tag{4.5.12}$$

where:

$$q_{\text{spec}} = [q - \psi(x_i, y_i)] + \frac{Q}{2} \tag{4.5.13}$$

where it is understood that values of q_{spec} which lie outside the range $[0, Q-1]$ are 'wrapped' into it by adding or subtracting integer multiples of Q, and that $\psi(x_i, y_i)$ is given in units of channel number. Defined in this way, q_{spec} accounts for pixels at different orders of interference by shifting their recorded spectra along the q-axis, so that all pixels viewing the same wavelength will peak at the same value of q_{spec} (and in the case that this wavelength is equal to that at which the phase map was recorded, or scaled to, the peak will occur at $\frac{Q}{2}$, in the middle of the spectrum).

4.6 Zone Mapping

In general a single pixel will not gather enough photons from the airglow source to achieve an adequate signal-to-noise ratio in a reasonable amount of time to record a useful spectrum from every pixel individually. In order to achieve an adequate signal-to-noise ratio within a reasonable integration time it is necessary to add-up the signal from a group of many pixels (herein referred to as a "zone"). The "zone map" is used to describe how the field-of-view should be divided up into these zones. It is a two-dimensional look-up table, defined entirely in software, which contains, for each pixel on the detector, the index of the zone to which that pixel belongs. Because all phase-offsets between spectra from different pixels have already been accounted for by the phase map, in principle any set of pixels may be used to form a zone, and there is no limitation (other than the number and finite size of the available pixels) on the size or shape of the resulting zones. Nor is it even necessary that the zone map remain constant between exposures, since it could be easily reconfigured in software on-the-fly, so that a given zone could, for example, be made to change position in order to track a moving feature in the sky. In order to preserve angular (directional) information, however, it is necessary that the pixels forming each zone be adjacent to one-another. Also, since smaller zones will be exposed to less total radiation for a given exposure time, the zone map represents a compromise between spatial and temporal resolution.

Two zone map configurations were used during 2007, both of which are shown in Figure 4.6.1. For the majority of 2007, the zone map in the top left panel of Figure 4.6.1 was

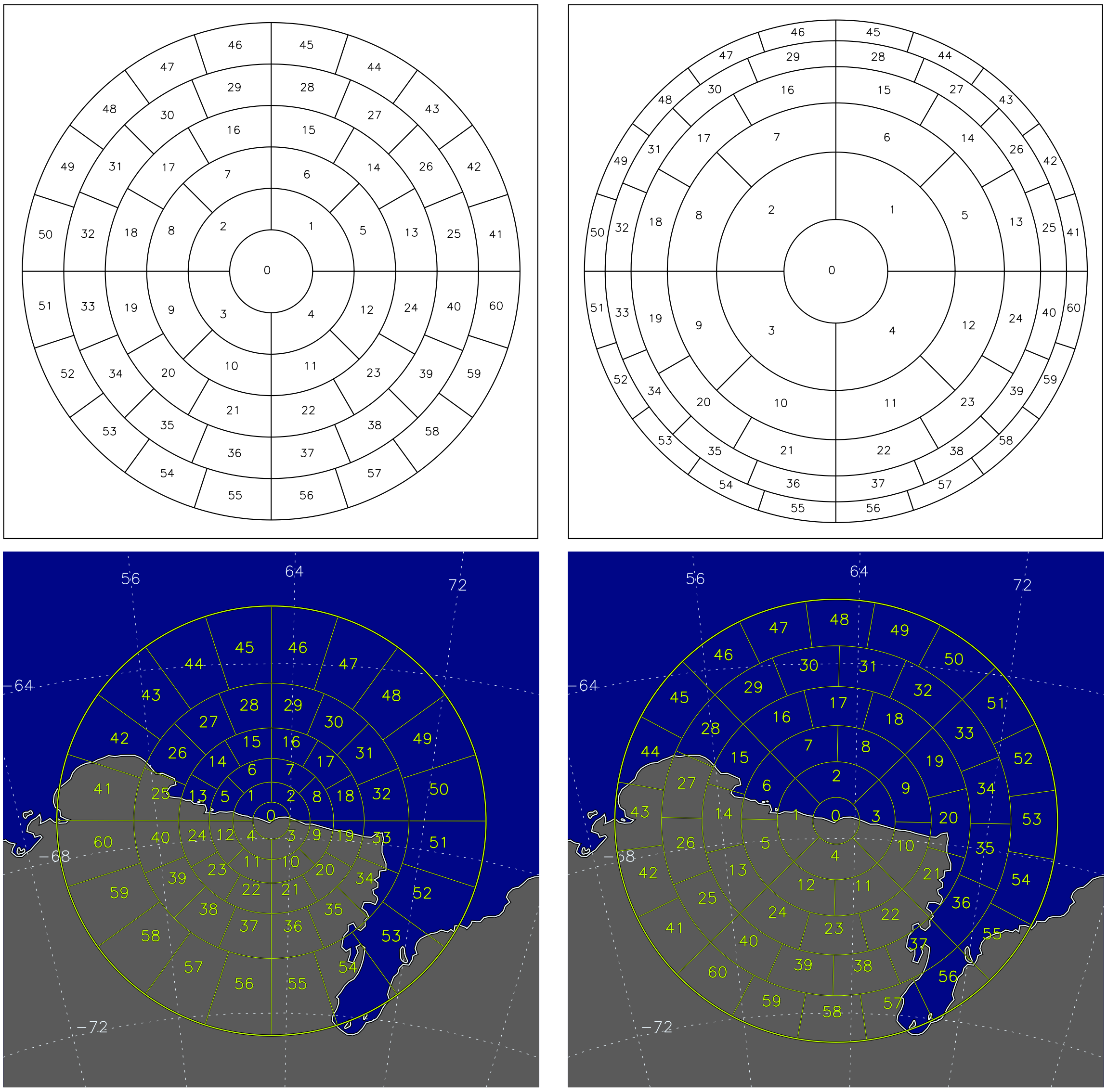


Figure 4.6.1: Example zone maps. Top row shows how the zone map groups sets of adjacent pixels to divide the detector into zones. The bottom row shows these same zone maps as seen by an observer in space looking down in altitude, where the zones have been projected onto the sky at a height of 240 km. The grey area here is the Antarctic continent. The left-hand maps were used for the majority of 2007, while the right-hand maps were used throughout 2008.

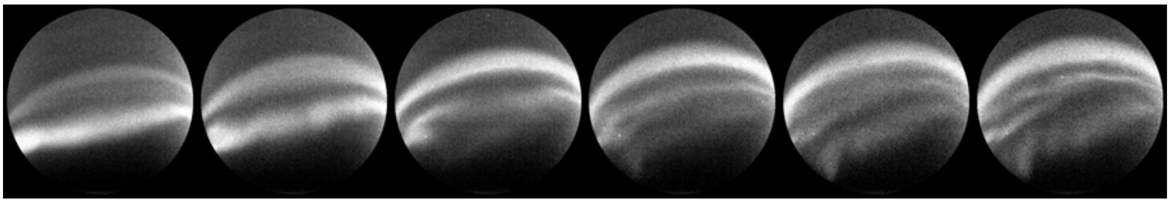


Figure 4.7.1: Sequence of all-sky images $A(x, y)$, built up from superpositions of interference fringe images recorded during five exposures. This sequence was taken from the night of April 21, 2007, and spans approximately 22 minutes. Discrete auroral forms are clearly visible. These images have been scaled and smoothed for clarity. Note that in 2007 the detector was aligned with the top directed toward compass north, which is why the auroral arcs are oriented slightly off-horizontal.

used. A total of 61 zones were defined, which divided the detector into five annuli of equal radial width, along with a circular central zone. Each annulus was divided into a number of zones of equal angular width. Outward from the innermost ring, annuli were divided into four, eight, twelve, sixteen and twenty zones respectively. The projection of this zone map onto the sky, at the assumed altitude of the airglow emission (240 km), is shown in the bottom left panel of Figure 4.6.1. Since a given radial distance on the sky is proportional to the tangent of the zenith angle, the projection of this zone map onto the sky results in zones which map to increasingly larger radial widths toward the edge of the field-of-view. The azimuthal width of the zones of course remains unchanged.

A different zone map configuration was introduced after August 1, 2007, which is shown in the top right panel of Figure 4.6.1. The number of zones within each annulus remained the same, however instead of using annuli of fixed radial width, the width was arranged to decrease roughly in proportion to the tangent of the zenith angle. This ensured that the projection of this zone map onto the sky (bottom right panel of Figure 4.6.1) resulted in zones of nearly constant radial width across the entire field-of-view.

As well as changes to the zone map, the detector orientation was also changed during 2007. For the majority of the year, the top of the detector was aligned along magnetic (compass) north, which is approximately 66° west of geographic north at Mawson. A small correction was made to this alignment on October 8, 2007, in order to align the detector perpendicular to the auroral oval (in the CGM or corrected geomagnetic north direction), which is often a more useful orientation in the auroral zone. The top of the camera was therefore aligned along a bearing of 44° west of geographic north. The map projections shown in Figure 4.6.1 correspond to zone maps aligned along the original (compass north) direction (left-hand column) and the current (CGM north) direction (right-hand column). These zone maps are centred on Mawson station, with the Antarctic continent shown in grey.

4.7 Spectral Accumulation

Once accurate values for $V_n(0)$ and S_n have been obtained, and a phase map and zone map defined, spectral accumulation can proceed. For the majority of the observing time the spectrometer is acquiring spectra, and the typical sequence of events is outlined in Figure 4.7.2.

In Step 9, the 512x512 pixel fringe image recorded at a particular channel in the scan is added to an array ($A(x, y)$) which contains a 'running-total' of all fringe images within the scan. The interference pattern imaged by the detector at each channel in the scan is a normal (monochromatic) image of the sky at the wavelength of observation, but one which is modulated by the Fabry-Perot interference fringes. Since a scan is performed over one order of interference, the interference fringes will sweep across every pixel in the field-of-view over the course of a scan. Thus, the superposition of all such images within a scan will show a complete (unmodulated) image of the sky (assuming the scene brightness did not vary significantly during the scan). Thus, at the end of an exposure, $A(x, y)$ contains a monochromatic image of the sky at the wavelength being observed, over an identical field-of-view and exposure time. This is extremely useful for comparing wind and temperature maps with auroral forms in the sky. Figure 4.7.1 shows a sequence of such all-sky images, from the night of April 21, 2007, recorded over approximately 22 minutes. The discrete auroral bands are clearly visible.

In Figure 4.8.1 an example of spectra resulting from a single exposure of the 630.0 nm airglow is shown. Laser calibration spectra (shown in blue), recorded prior to this exposure, are plotted along with the recorded sky spectra (shown in red) for each zone. The variation of recorded spectra with zone number is illustrated by comparing spectra recorded in the central zone to those recorded in one of the outer zones, in this case zone number 40. A bright auroral arc (not shown) was present in the sky, and mapped to zones in the top half of the zone map of Figure 4.8.1, which resulted in the spectra from these zones showing a higher signal-to-noise ratio than those spectra recorded in zones in the lower half of the zone map.

4.8 Spectral Fitting

Recorded spectra are fitted with an analytic model spectrum to yield estimates of Doppler-shift, Doppler-width, emission intensity and continuum background intensity. The finite spectral bandwidth of the instrument will result in recorded spectra which are broadened relative to their emission profile. This instrumental broadening is seen physically in the widths of spectra recorded when viewing a spectrally narrow source (such as a calibration laser). The function which characterises this broadening is called the instrument function (I), a discrete representation of which is obtained by recording a laser spectrum. The spectral fitting algorithm employed here is described by Conde (2001). This algorithm was originally designed for Lidar backscatter spectra. Lidar is similar to radar, except light at visible wavelengths is used rather than radio wavelengths. The Lidar spectra may contain contributions from light backscattered by both aerosols and molecules. The algorithm can however be run in a simplified mode which is optimum for fitting passive

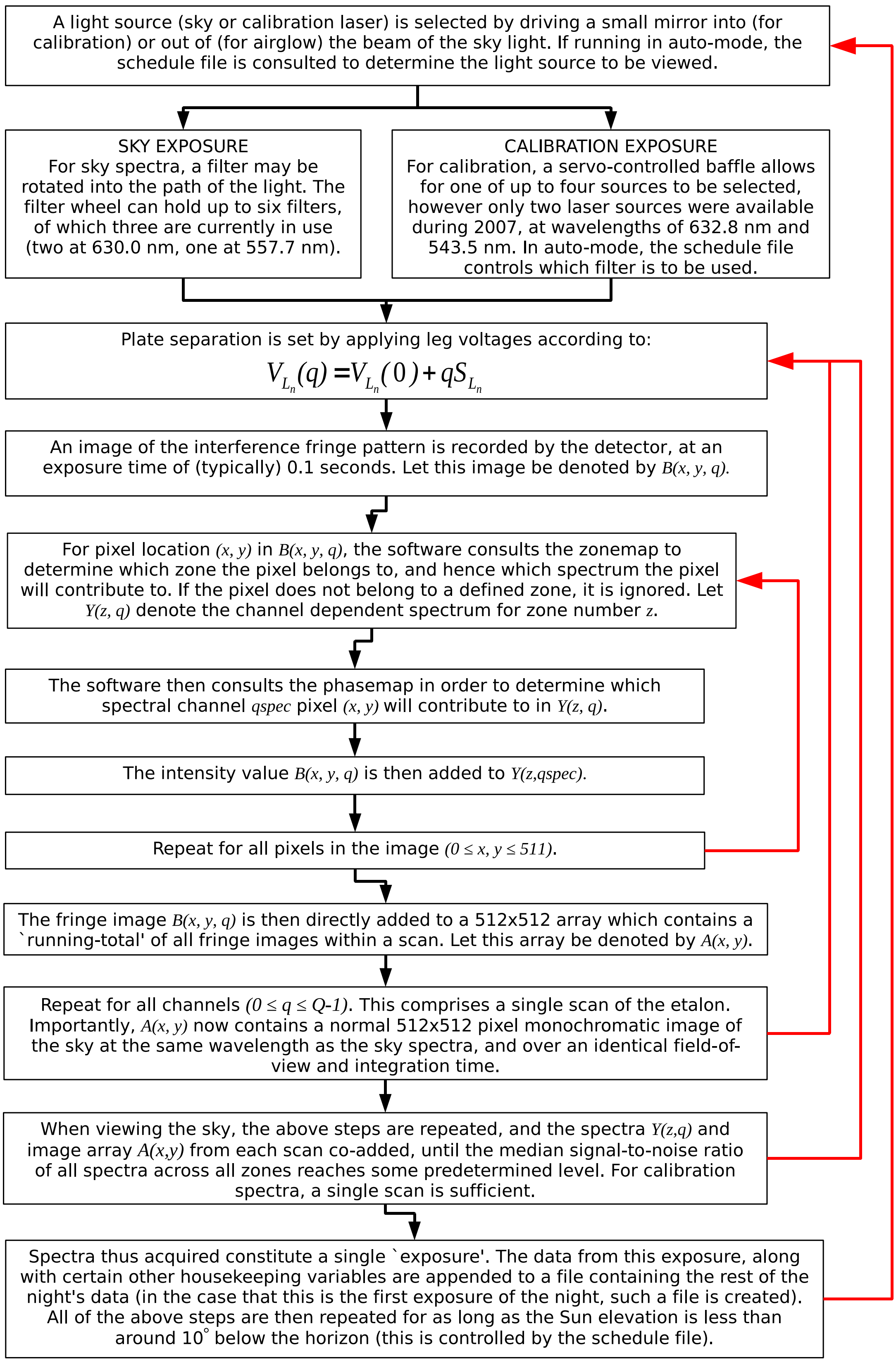


Figure 4.7.2: Flowchart of the spectral acquisition process.

airglow emission spectra, which are not affected by the presence of aerosols[1]. A very brief description of the algorithm is given here, highlighting the important steps.

The fitting algorithm models the recorded spectrum $\{y_q\}$ as a discrete set of Q samples $\{s_q\}$ of the function:

$$s(\lambda) = \sum_{j=0}^{2} a_j S_j(\lambda) \tag{4.8.1}$$

where the coefficients a_j are the intensities of each of the source terms $S_j(\lambda)$, given by:

$$S_0(\lambda) = 1 \qquad \text{Continuum background} \tag{4.8.2}$$

$$S_1(\lambda) = I(\lambda) * \delta(\lambda - \lambda_0) \qquad \text{Aerosol} \tag{4.8.3}$$

$$S_2(\lambda) = I(\lambda) * \sum_{n=0}^{N-1} q_n \exp\left(-\left(\frac{\lambda - \lambda_0}{w_n(T)}\right)^2\right) \qquad \text{Molecular} \tag{4.8.4}$$

where $*$ denotes the convolution operation. The molecular term, Equation 4.8.4, allows for backscatter from a multi-component atmosphere (with up to N species), where the q_n are the (pre-determined) *relative* backscatter intensities for each species, such that:

$$\sum_{n=0}^{N-1} q_n = 1 \tag{4.8.5}$$

and the $w_n(T)$ term describes the width of the emission/backscatter profile for the n^{th} species, which is a function of the kinetic temperature T. The λ_0 term is the Doppler-shifted wavelength of the emitting/scattering species moving at a bulk velocity v.

In general, the algorithm solves for the intensity coefficients a_j analytically, while numerically searching for the best combination of λ_0 and T. However it is versatile enough to allow for any of these parameters to be held constant at a user-supplied value. For example, the aerosol term (Equation 4.8.3), which describes the backscatter of lidar radiation by aerosols (in terms of the instrument function convolved with the Dirac delta function) is not required for fitting spectra acquired by a passive FPS, and thus the a_1 term is held constant at 0. In a similar manner, for airglow emission from a single atomic species (in this case oxygen), we have $N = 1$ and $q_0 = 1$. From Section 2.6.3 (Equation 2.6.5), the width function $w_0(T)$ for a Doppler-broadened emission line is given by:

$$w_0(T) = f = \lambda_0 \sqrt{\frac{2kT}{Mc^2}} \tag{4.8.6}$$

where M is the atomic mass of oxygen, k is the Boltzmann constant and c is the speed of

[1]Aerosol scattering does affect the angular distribution of the signal. But it does not (directly) affect the spectra.

light. Equation 4.8.1 then becomes:

$$s(\lambda) = a_0 + I(\lambda) * a_2 \exp\left(-\frac{(\lambda - \lambda_0)^2 Mc^2}{2kT\lambda_0^2}\right) \tag{4.8.7}$$

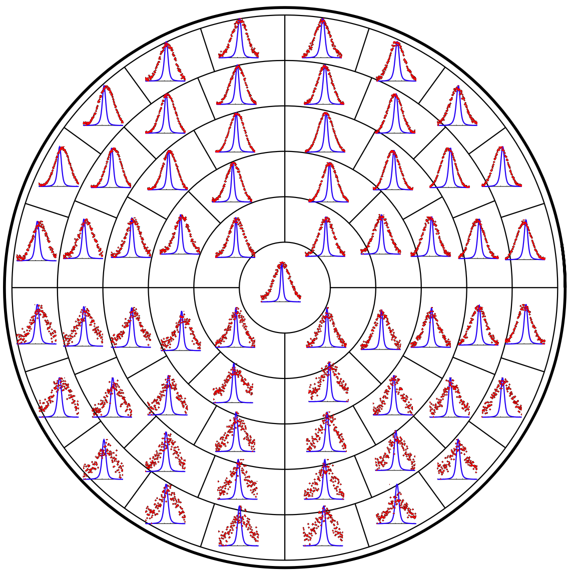
The 'best-fit' parameters (p={a_0, a_2, λ_0, T}) are taken to be those which minimise:

$$\chi^2 = \sum_{q=0}^{Q-1} \frac{(y_q - s_q)^2}{\sigma_q^2} \tag{4.8.8}$$

where σ_q^2 is the variance of the qth element of the observed spectrum (y_q). That is, the optimum choice of parameters $\{p_j\}$ should satisfy

$$\frac{\partial \chi^2}{\partial p_j} = 0 \tag{4.8.9}$$

As mentioned above, the algorithm uses a combined approach to find the best choice of parameters $\{p_j\}$ appearing in Equation 4.8.7. At each iteration, the algorithm uses the current 'best-guess' values of λ_0 and T (or initial estimates of these parameters in the first iteration) to solve for a_0 and a_2 analytically. New estimates of λ_0 and T are then generated using the Levenberg-Marquardt approach (Press et al., 1986), and the goodness-of-fit (Equation 4.8.8) evaluated. These steps are repeated until either successive values of χ^2 differ by a very small (configurable) amount, or a predefined maximum number of iterations is reached, whichever condition is met first.



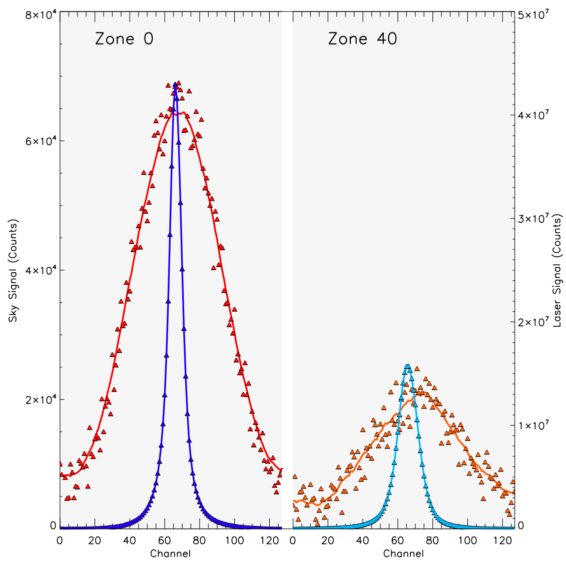


Figure 4.8.1: Left: Normalised laser (blue) and sky (red) spectra recorded on 23rd May 2007. These spectra were summed from 3 scans of the etalon, with a total exposure time of 2.5 minutes. Right: Close-up of spectra from zone 0 and zone 40, illustrating the variation of recorded spectra with zone number. Data points are the observed spectra, solid curves are smoothed versions of the observed spectra. These spectra have not been normalised.

4.9 Deriving Line-of-Sight Wind Speed

The spectral fitting algorithm returns a time-series of estimates of peak wavelength ('peak position'), temperature, emission and continuum background intensity, along with the associated 1σ statistical uncertainties, for each spectrum. Peak positions are calculated in units of channels relative to the first recorded calibration spectrum of the night for the corresponding zone. Real etalons are not perfectly stable (at the required level of a few nanometers), so to infer wind speeds to m.s^{-1} accuracy requires the removal of 'drift' of the etalon plates, a term which refers to the gradual, systematic variation in the *inferred* laser wavelength (since the actual laser wavelength is assumed constant), due for example to changes in the mean etalon plate separation throughout the night.

Since the phase map (which encodes the channel number corresponding to the laser peak position for every pixel) is only recorded at the beginning of the night's observations, any variation in mean plate separation l will shift the inferred wavelength peaks toward higher (decreased l) or lower (increased l) channel number. This drift can be tracked by observing a calibration laser source which is assumed to remain stable (fixed in wavelength) throughout the night. At Mawson, a frequency-stabilised Helium-Neon laser is used as the calibration source for drift correction. A plot of the variation in recorded laser peak position throughout the night (relative to the first recorded peak position of the night) is shown in Figure 4.9.1. In this figure, the drift in the laser peak position for zone z is given by:

$$d(z,t) = p_{las}(z,t) - p_{las}(z,t_0) \tag{4.9.1}$$

where $p_{las}(z,t)$ is the laser peak position in zone z at time t (in this case $0 \leq z \leq 60$),

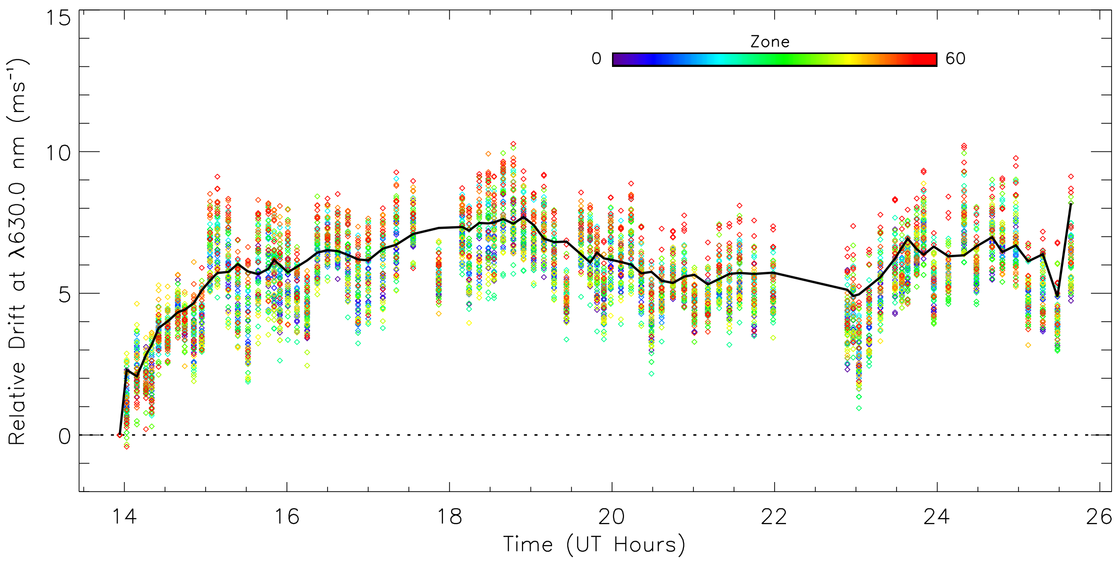


Figure 4.9.1: Example laser drift for the night of the 17$^{\text{th}}$ April 2007. Peak wavelength variation throughout the night (relative to the first peak wavelength) is shown for each zone, expressed in terms of the change in line-of-sight velocity the drift would produce at an observing wavelength of λ630.0 nm. The black curve is the smoothed laser drift recorded in zone 0.

and $p_{las}(z, t_0)$ corresponds to the first peak position of the night. In order to correct the sky peak positions for laser drift, $d(z, t)$ must be subtracted from the sky peak positions. Since in practice calibration spectra are recorded less frequently than sky spectra (one calibration to five sky exposures is routine), it is necessary to interpolate $d(z, t)$ in time to match the times at which the sky spectra were recorded. A small amount of smoothing (in time) is also applied to $d(z, t)$ at this stage to suppress measurement noise in the drift calibration. If $\widetilde{d}(z, t)$ is used to denote the time-smoothed, interpolated drift, then the drift-corrected sky peak positions are given by:

$$p_{sky}^{dc}(z,t) = p_{sky}(z,t) - \left[\widetilde{d}(z,t) \times \left(\frac{\lambda_0}{\lambda}\right)\right] \tag{4.9.2}$$

where λ_0 and λ are the laser and sky wavelengths respectively, which are used to scale the laser drift to the sky wavelength, as was done with the phase map (Equation 4.5.11).

In order to convert these drift-corrected line-of-sight peak positions into line-of-sight speeds, it is necessary to determine the peak position corresponding to a zero Doppler-shift, in other words a zero-velocity reference. In the absence of a convenient laboratory source of the 630.0 nm spectral line, the usual method is to assume that the line-of-sight wind speed in the zenith (the vertical wind, measured in zone 0) will be close to zero for data averaged over many hours. In general this is a reasonable assumption to make, since hydrostatic equilibrium should oppose any sustained vertical motion of the atmosphere beyond that due to a changing vertical pressure gradient (from solar heating for example). Strictly, for this assumption to be valid, a full 24-hour period of observation of the zenith wind would be required in the absence of strong geomagnetic heating, however (Section 2.4.2) even if these conditions are not met, the systematic error introduced through the assumption of zero mean zenith wind speed is at most 10-20 ms^{-1}, and typically is likely to be very much less (Aruliah and Rees, 1995).

The zero Doppler-shift baseline is thus computed by finding the median of the drift-corrected zenith peak position time-series. This value is then subtracted from the drift-corrected line-of-sight peak positions from all zones. Peak positions from the zenith-looking zone (zone 0) will therefore have a median of zero, while those from all off-zenith zones will now represent the line-of-sight Doppler-shifts (in units of channels) relative to the assumed Doppler baseline. The Doppler formula relates a change in observed wavelength to a change in the (line-of-sight, non-relativistic) speed v of the emitting source by:

$$\Delta v = \frac{c\Delta\lambda}{\lambda_0} \tag{4.9.3}$$

A spectrum spans one free spectral range, which from Section 3.1 is given, at normal incidence, by:

$$\begin{aligned}\Delta_\lambda &= \frac{\lambda_0^2}{2\mu l}\\ &= \frac{(630.0\times 10^{-9}\ \mathrm{m})^2}{2\times 25\times 10^{-3}\ \mathrm{m}}\\ &= 7.94\times 10^{-12}\ \mathrm{m}\end{aligned} \tag{4.9.4}$$

where l is used now to represent the mean etalon spacing, and $\mu = 1$. Thus, the line-of-sight speed change that would shift the observed emission wavelength by one free spectral range is equal to:

$$\begin{aligned}\Delta v &= \frac{(3\times 10^{8}\ \mathrm{m.s^{-1}})(7.94\times 10^{-12}\ \mathrm{m})}{630.0\times 10^{-9}\ \mathrm{m}}\\ &= 3781\ \mathrm{m.s^{-1}}\end{aligned} \tag{4.9.5}$$

The Mawson instrument is configured to scan over $Q = 128$ channels, and so one channel is equivalent to a line-of-sight speed change of $\Delta v/128 = 29.54$ ms^{-1}. The final line-of-sight wind speeds are therefore calculated using:

$$v_{sky}(z,t) = \left(\frac{\Delta v}{128}\right) p_{sky}^{dc}(z,t) \tag{4.9.6}$$

4.9.1 Offset Correction

Cloud appearing in the instrument's field-of-view can affect line-of-sight velocity estimates, by scattering light into and out of the line-of-sight direction. The signal seen when viewing the cloud will therefore contain contributions from both the unscattered line-of-sight brightness and the brightness scattered into the line-of-sight from other regions of the sky. If the scattered component is much brighter than the unscattered component, as could happen when viewing a sky region adjacent to a bright auroral arc for example, then the observed signal may contain a significant or even dominant contribution from directions other than the original line-of-sight.

Thus any zone viewing cloud essentially records an intensity-weighted spatial average of the spectrum over a region of the sky that extends beyond the original viewing zone. Such scattering can shift the apparent wavelength seen when viewing cloud relative that which would be seen when viewing the sky directly, dependent on the angular distribution of emission brightness and Doppler-shift. If the sky is very overcast, such that all zones are viewing heavy cloud, the averaging effect of the cloud is so complete that the line-of-sight velocity measured by *all* zones will approach zero.

However, examination of exposures recorded by the Mawson instrument during such periods of very heavy cloud (covering the entire field-of-view) showed non-zero line-of-sight velocities whose magnitude varied with azimuth around the detector, and were consistent with a constant, non-zero offset aligned along the detector's top-bottom axis. When this axis was aligned along the magnetic meridional direction (with magnetic north directed from the bottom to the top of the detector) the effect of this offset was to add a magnetic meridional component of approximately -45 ms^{-1} (i.e. magnetically southward) to the fitted horizontal wind field (see Section 4.10 below).

The exact mechanism responsible for this offset is currently unclear, however there is no doubt that this effect is an instrumental artifact and there are strong indications from both the Mawson instrument and from a similar one operating in Alaska that it is related to the way charge was read out from the camera's CCD chip (as both instruments used almost identical cameras). However, correcting for the offset was straightforward. For each year (2007 and 2008) a minimum of 12 days which had very heavy cloud cover for the entire night were identified. These days were easily identified both by cloud observations from the Australian Bureau of Meteorology and by the very flat (but non-zero) all-sky average wind time-series. The average line-of-sight velocity in each zone was then calculated from these heavily overcast days. The result was an average 'cloudy day' line-of-sight wind map which characterised the offset (the actual offset maps for 2007 and 2008 are reproduced in Figure 4.9.2). Averaging over (at least) 12 days ensured that any residual angular variation of Doppler-shift that may have "leaked" past the actual scattering of the clouds was suppressed. This map was then subtracted from the sky maps of Doppler-shift for each exposure before fitting the vector wind field (as described below). The results presented in Chapter 7 confirmed unambiguously the need for this offset correction.

4.10 Vector Wind Fitting

Having reduced the recorded spectra to drift-corrected line-of-sight velocities, it is then possible to 'fit' a two-dimensional (horizontal) vector wind field to every sky exposure.

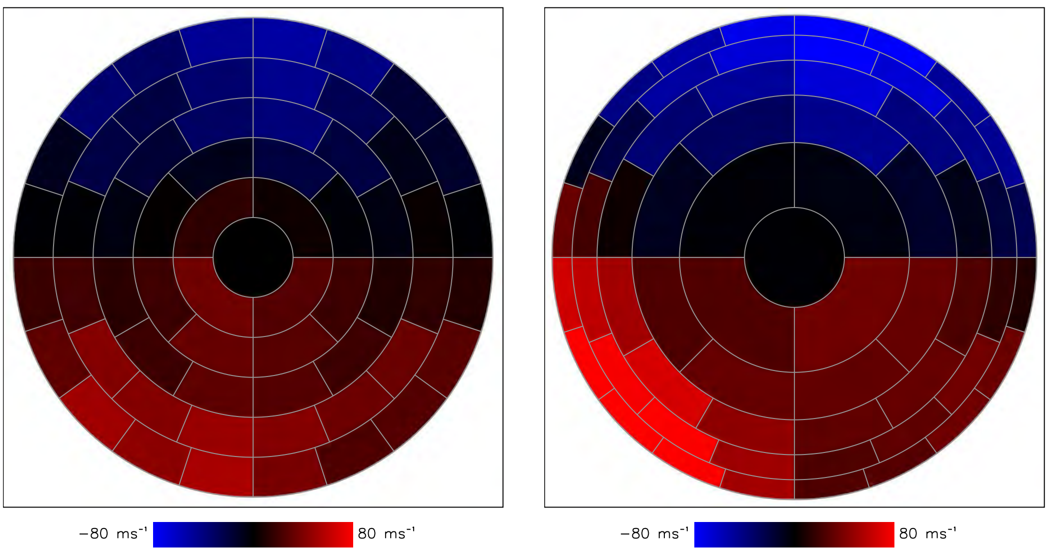


Figure 4.9.2: Average 'cloudy day' line-of-sight wind maps, used for offset correction. The map calculated for 2007 is on the left, that for 2008 on the right. Red hues indicate line-of-sight winds moving away from the observatory ('red-shift'), blue lines indicate motion toward the observatory ('blue-shift'). The velocity scale is indicated at the bottom of each map.

While it is possible to analyse the observed wind-field solely in terms of the spatial distribution of Doppler-shifts calculated as above, the vector-field representation is far more amenable to physical interpretation, and can reveal details in the winds which would be missed in the Doppler approach. The vector fitting method derived here follows that of Conde and Smith (1998), which is a slightly modified version of that by Browning and Wexler (1968), who applied the technique to azimuth-scanned Doppler radar measurements, and which was extended to the analysis of thermospheric airglow observations by Burnside et al. (1981).

The example zone maps shown in Figure 4.6.1 contain zones which are defined by the radius (outward from the central zone) and azimuthal angle θ_k (relative to some reference direction) of the centre of the zone. A zone's radius corresponds to the spectrometer viewing some zenith angle ϕ, and so the set of line-of-sight wind velocities for a single exposure may be written as $\{V_{\|}(\theta_k, \phi)\}$. The basic 'unit' of data used in this analysis is the line-of-sight wind measured in one ring of zones, and therefore it is only necessary to subscript the azimuthal angle θ, and not the zenith angle ϕ, since a ring of zones will have a common ϕ. Also note that the usage of θ and ϕ here does not follow the standard usage in physics (where θ would usually be associated with the zenith angle, and ϕ with the azimuthal angle), however it is commonly adopted by mathematicians.

The line-of-sight wind speed will be made up of a vertical and a horizontal component, which will be labelled by $\{Z_{\|}(\theta_k, \phi)\}$ and $\{H_{\|}(\theta_k, \phi)\}$ respectively (these quantities are therefore the vertical and horizontal *components* of the wind speed along the line-of-sight). Thus:

$$V_{\|}(\theta_k, \phi) = Z_{\|}(\theta_k, \phi)\cos(\phi) + H_{\|}(\theta_k, \phi)\sin(\phi) \tag{4.10.1}$$

A diagram of the relevant angles and vector components is shown in Figure 4.10.1. If the vertical wind is assumed to be uniform throughout the field-of-view[2], then:

$$Z_{\|}(\theta_k, \phi) = V_z = V_{\|}(\theta_0, 0) \tag{4.10.2}$$

and the contribution to the off-zenith $V_{\|}$ from the vertical wind can be removed to yield the horizontal component of the line-of-sight velocity:

$$H_{\|}(\theta_k, \phi) = \frac{V_{\|}(\theta_k, \phi) - V_z\cos(\phi)}{\sin(\phi)} \tag{4.10.3}$$

In the (horizontal) xy plane, let θ_k measure the clockwise angle between the positive y-axis and the unit vector from the origin (representing the location of the observatory) along the line-of-sight. The angle θ_k is then to be thought of as a bearing, and the components of total horizontal vector $(H_x\hat{\mathbf{x}} + H_y\hat{\mathbf{y}})$ will be related to $H_{\|}$ by:

$$H_{\|}(\theta_k, \phi) = H_x\sin(\theta_k) + H_y\cos(\theta_k) \tag{4.10.4}$$

[2]See the discussion in Section 4.10.1, below.

where the positive x and y directions are chosen to represent the zonal and meridional directions respectively. The zonal and meridional components of $H_{\|}(\theta_k, \phi)$ can be approximated by their first-order Taylor expansion about the zenith:

$$H_x = u_0 + \left(\frac{\partial u}{\partial x}\right) x(\theta_k, \phi) + \left(\frac{\partial u}{\partial y}\right) y(\theta_k, \phi) \tag{4.10.5}$$

$$H_y = v_0 + \left(\frac{\partial v}{\partial x}\right) x(\theta_k, \phi) + \left(\frac{\partial v}{\partial y}\right) y(\theta_k, \phi) \tag{4.10.6}$$

For observations of airglow emitted at a mean altitude h, the variables $x(\theta_k, \phi)$ and $y(\theta_k, \phi)$ in the above represent the horizontal distance from the origin to the point (θ_k, ϕ), and are given in terms of radial distance $R = h\tan(\phi)$ by:

$$\begin{aligned} x(\theta_k, \phi) &= h\tan(\phi)\sin(\theta_k) \\ &= R\sin(\theta_k) \qquad (4.10.7) \\ y(\theta_k, \phi) &= h\tan(\phi)\cos(\theta_k) \\ &= R\cos(\theta_k) \qquad (4.10.8) \end{aligned}$$

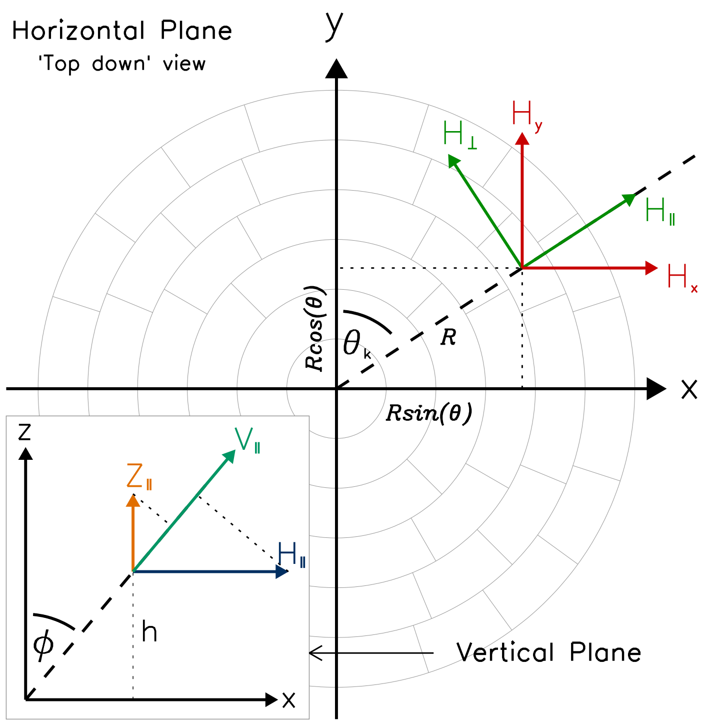


Figure 4.10.1: Angles and vector components relevant to the vector fitting algorithm. The main diagram shows a zone map (2007 configuration) in the horizontal plane, viewed from above. Inset shows the wind components in the vertical plane. Angles and vector components are discussed in-text.

Combining these with Equations 4.10.4 - 4.10.6 gives:

$$\begin{aligned} H_{\|}(\theta_k, \phi) &= u_0 \sin(\theta_k) + \left(\frac{\partial u}{\partial x}\right) R \sin^2(\theta_k) + \left(\frac{\partial u}{\partial y}\right) R \cos(\theta_k) \sin(\theta_k) \\ &+ v_0 \cos(\theta_k) + \left(\frac{\partial v}{\partial x}\right) R \sin(\theta_k) \cos(\theta_k) + \left(\frac{\partial v}{\partial y}\right) R \cos^2(\theta_k) \end{aligned} \tag{4.10.9}$$

The trigonometric identities:

$$\begin{aligned} \sin^2(x) &= \frac{1}{2} - \frac{1}{2}\cos(2x) \\ \cos^2(x) &= \frac{1}{2} + \frac{1}{2}\cos(2x) \\ \sin(x)\cos(x) &= \frac{1}{2}\sin(2x) \end{aligned}$$

allow Equation 4.10.9 to be expanded as:

$$\begin{aligned} H_{\|}(\theta_k, \phi) &= u_0 \sin(\theta_k) + \left(\frac{R}{2}\right)\left(\frac{\partial u}{\partial x}\right)(1 - \cos(2\theta_k)) + \left(\frac{R}{2}\right)\left(\frac{\partial u}{\partial y}\right)\sin(2\theta_k) \\ &+ v_0 \cos(\theta_k) + \left(\frac{R}{2}\right)\left(\frac{\partial v}{\partial x}\right)\sin(2\theta_k) + \left(\frac{R}{2}\right)\left(\frac{\partial v}{\partial y}\right)(1 + \cos(2\theta_k)) \end{aligned} \tag{4.10.10}$$

If, for a given zenith angle ϕ, the wind field is sampled at n independent azimuths, then $H_{\|}$ can also be expressed as a Fourier series expansion in θ_k:

$$H_{\|}(\theta_k, \phi) = a_0 + \sum_{m=1}^{\frac{n}{2}-1} (a_m \cos(m\theta_k) + b_m \sin(m\theta_k)) \tag{4.10.11}$$

with coefficients a_0, a_m and b_m given by:

$$\begin{aligned} a_0(\phi) &= \frac{1}{n}\sum_{k=0}^{n-1} H_{\|}(\theta_k, \phi) \\ a_m(\phi) &= \frac{2}{n}\sum_{k=0}^{n-1} H_{\|}(\theta_k, \phi)\cos(m\theta_k) \\ b_m(\phi) &= \frac{2}{n}\sum_{k=0}^{n-1} H_{\|}(\theta_k, \phi)\sin(m\theta_k) \end{aligned} \tag{4.10.12}$$

Equating the coefficients of Equations 4.10.10 and 4.10.11 gives the relations between the

Fourier coefficients and the partial derivatives in the Taylor expansion:

$$a_0 = \frac{R}{2}\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right) \tag{4.10.13}$$

$$a_1 = v_0 \tag{4.10.14}$$

$$b_1 = u_0 \tag{4.10.15}$$

$$a_2 = \frac{R}{2}\left(\frac{\partial v}{\partial y} - \frac{\partial u}{\partial x}\right) \tag{4.10.16}$$

$$b_2 = \frac{R}{2}\left(\frac{\partial u}{\partial y} + \frac{\partial v}{\partial x}\right) \tag{4.10.17}$$

The values of the Fourier coefficients are calculated directly from the set $\{H_{\|}(\theta_k, \phi)\}$, from which the values of the background wind field, v_0 and u_0, are given directly by Equations 4.10.14 and 4.10.15 respectively. The only requirement is that the number of wind samples within a given annulus is ≥ 6 for the calculation of both partial derivative and background terms, or ≥ 4 to compute only the background terms. In the former case, what remains after solving for the two background wind terms is three equations in *four* unknowns, and hence an assumption must be made about one of the four partial derivatives if this set of equations is to be solved.

The assumption of Burnside et al. (1981) was that the meridional wind field is stationary in local time, and that any variation observed in this field is due to the rotation of the Earth moving the station underneath this wind field, and allowing it to be sampled at multiple points along the zonal direction. Under this assumption it is possible for a single station to obtain information regarding the vorticity of the wind field by sampling it from multiple longitudes. This is not possible with a single observation (exposure) from a single station, as only the divergent component is directly measured.

Experience has shown that the gradient calculated using the method of Burnside et al. is often very small in comparison with the other gradients, however wind fields derived under this assumption can often be unrealistic when the local wind field is rapidly time-varying, as for example often occurs during substorms. Therefore the assumption which is made in this work is that during a single exposure the geomagnetic meridional wind does not vary along the geomagnetic zonal direction, which requires setting $\frac{\partial v}{\partial x} = 0$. In the auroral zone, the zonal wind field (in magnetic coordinates) is likely to be the most variable, as this is the direction along which the ionospheric plasma is driven by the convection electric field, a driving force which is highly variable on short time and spatial scales. Auroral heating is likely to affect the meridional gradient of the meridional wind since it is in this direction that appreciable pressure gradients would be expected to develop. There, the component which shows the least spatial variation for the majority of the time is the meridional component along the zonal direction, and experience has indeed shown that this is the most satisfactory way to deal with the limitations of measuring the wind from just a single site. By making this assumption, the remaining partial derivatives can be solved for, and are given by:

$$
\begin{aligned}
v_0 &= a_1 & u_0 &= b_1 \\
\frac{\partial u}{\partial x} &= \left(\frac{a_0 - a_2}{R}\right) & \frac{\partial v}{\partial x} &= 0 \\
\frac{\partial u}{\partial y} &= \frac{2b_2}{R} & \frac{\partial v}{\partial y} &= \left(\frac{a_0 + a_2}{R}\right)
\end{aligned}
$$

Once the above partial derivatives are known, these in conjunction with the background wind terms yield a two-component fit to the horizontal vector field. However, since the line-of-sight component of the vector field is measured directly, it is appropriate to resolve the fitted vector field into components parallel and normal to the line-of-sight direction, and then to merge the *measured* parallel component and the *fitted* normal component to give an estimate of the wind field which includes all of the features of the measured line-of-sight wind field, but which may omit or distort some features of the field perpendicular to the line-of-sight. The normal component is given by:

$$H_{\perp}(\theta_k, \phi) = H_y(\theta_k, \phi)\sin(\theta_k) - H_x(\theta_k, \phi)\cos(\theta_k) \tag{4.10.18}$$

The components of the wind field given by Equations 4.10.5 and 4.10.6 are then:

$$H_x(\theta_k, \phi) = H_{\parallel}(\theta_k, \phi)\sin(\theta_k) - H_{\perp}(\theta_k, \phi)\cos(\theta_k) \tag{4.10.19}$$

$$H_y(\theta_k, \phi) = H_{\parallel}(\theta_k, \phi)\cos(\theta_k) + H_{\perp}(\theta_k, \phi)\sin(\theta_k) \tag{4.10.20}$$

Thus the vector fit provides an estimate of the Taylor series background terms in each annulus containing ≥ 4 wind samples, and the partial derivative terms in each annulus containing ≥ 6 wind samples. It is assumed that the gradients of the wind field across the entire field-of-view can be represented by a single set of four partial derivatives (instead of a separately computed set of four derivatives within each annulus), and so the partial derivatives within each annulus (including those for which no derivatives were calculated) are replaced with their mean values over all annuli. When computing the final wind field, however, the individual background wind terms calculated in all annuli with ≥ 4 samples are used. Note that no fitting can be done in the central zone. In practice, the mean horizontal vector components across the entire field-of-view are assigned to the central zone, which then contains an estimate of the average or background wind flow over the field-of-view. Figure 4.10.2 shows an example of the observed line-of-sight component, fitted normal component, and the vector wind field obtained by combining the two perpendicular components. This vector field was fitted to the data shown in Figure 4.8.1.

4.10.1 Vertical Wind Assumption

Traditionally, two approaches have been used to subtract the Doppler-shift component due to a vertical wind from an off-zenith line-of-sight wind measurement. In the simplest case,

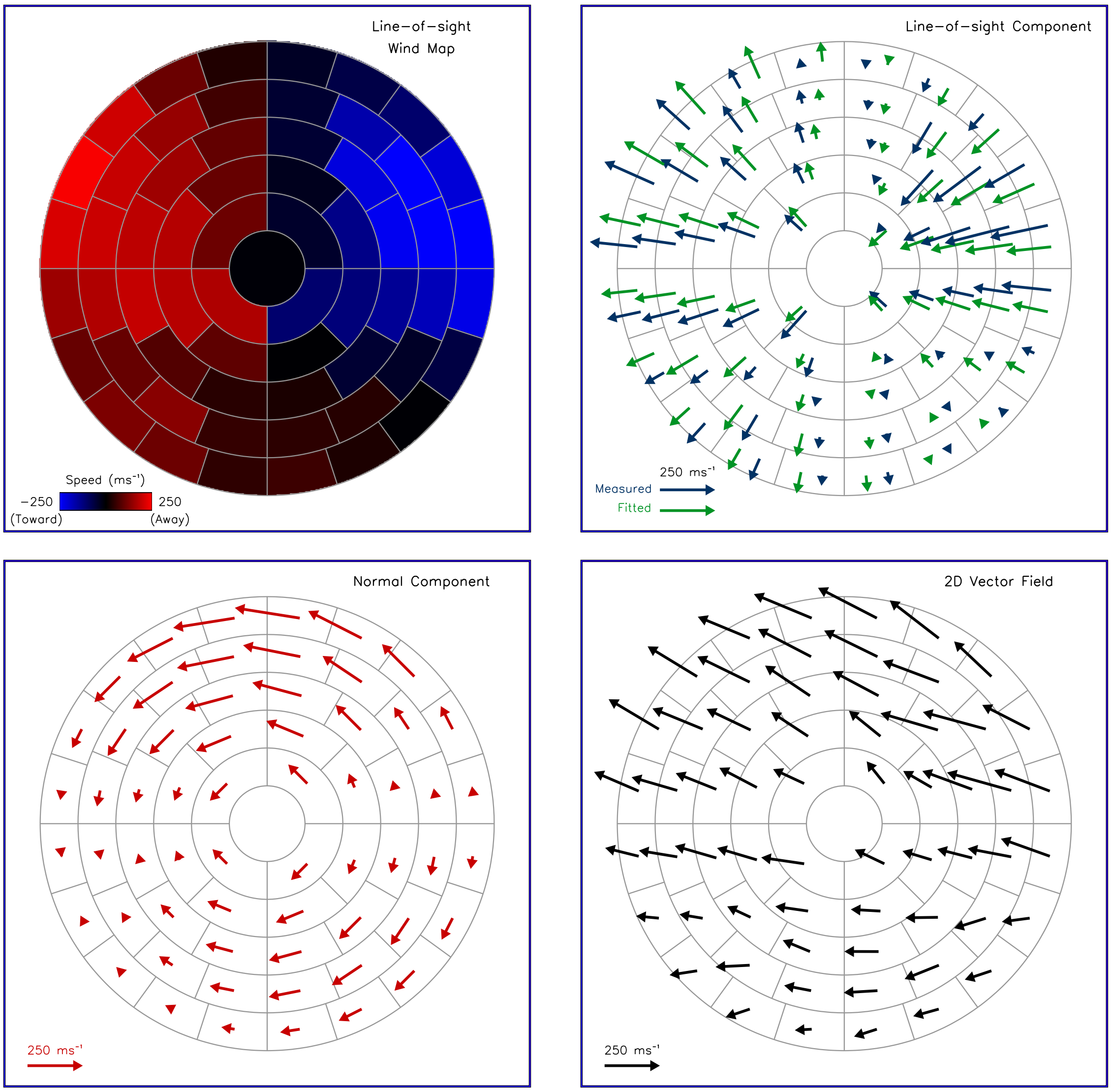


Figure 4.10.2: Example vector fit, for the data shown in Figure 4.8.1. Top left: line-of-sight velocity wind map (with central zone excluded), as measured directly by the instrument, where red hues (positive) indicate wind directed away from the observatory, and blue hues (negative) indicate motion toward the observatory (scale in bottom left corner). Top right: the measured (blue) and fitted (green) line-of-sight component. Bottom left: fitted component normal to the line-of-sight. Bottom right: the full two-dimensional wind field produced by combining the line-of-sight (divergent) and normal (rotational) component wind fields. Magnetic north is to the top of these images, magnetic east to the right.

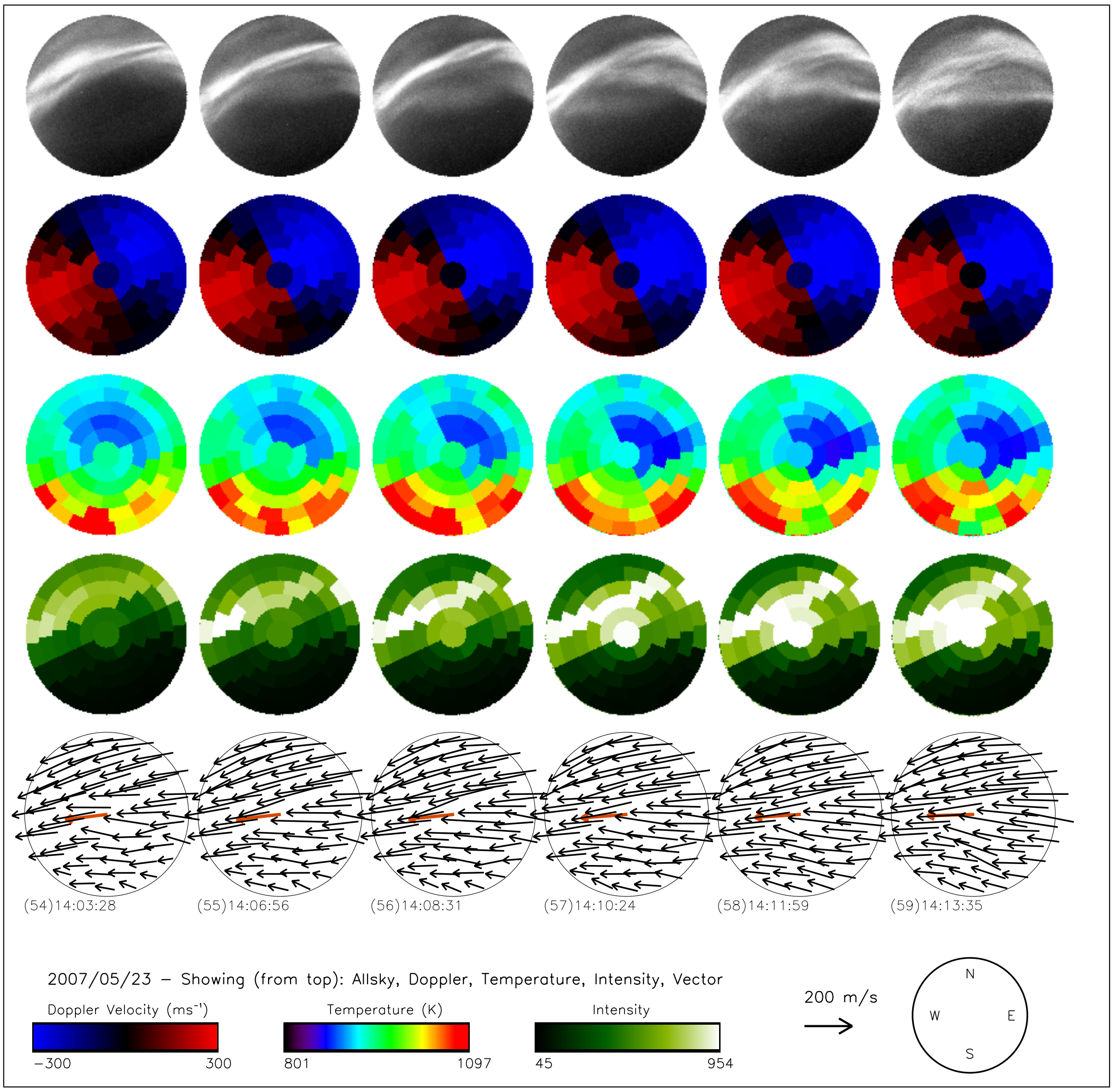


Figure 4.10.3: Routine analysis output from the Mawson spectrometer. Shown from the top (where each column represents a single exposure) are the recorded all-sky images, the horizontal component of the line-of-sight Doppler velocity (blue hues indicate winds directed toward the observatory, red hues indicate flow away from the observatory), temperature, emission intensity, and the fitted vector field (orange vectors represent the average wind over the whole field-of-view). The universal times at which each exposure was made are shown underneath the vector wind field. Doppler velocity, temperature and intensity scales are shown at the bottom of the figure, as are the vector scale and the field-of-view orientation in geomagnetic coordinates.

no subtraction is actually done, since it is assumed that the vertical velocities are much smaller than the horizontal velocities. On average, at least, this assumption may be valid. Vertical wind speeds are often on the order of tens of meters per second (speeds > 30 m.s^{-1} are not frequently observed), compared to horizontal winds which are on the order of hundreds of meters per second. Combined with the fact that many narrow-field Fabry-Perot spectrometers generally view the sky at zenith angles of around 45°, means that the component of the line-of-sight velocity due to a vertical wind ($V_z \cos(\phi)$) will generally be negligible under most conditions. Another approach, and the one presented in the previous section (Equation 4.10.2) has been to use the vertical wind directly measured in the zenith as an approximation of the vertical wind field over the instrument field-of-view, in which case $V_z \cos(\phi)$ would give the component of the line-of-sight wind due to the vertical wind when viewing at a zenith angle ϕ.

If the assumption of small vertical wind speeds held, then the latter approach would at best be a minor correction to the derived horizontal wind. If the assumption of a uniform vertical wind field overhead of the station did not hold, then the vertical wind 'correction' may in fact produce unwanted errors (in the worst case the assumed vertical wind may be of opposite sign to the 'true' vertical wind in the measurement volume). In Chapter 7, results are presented of a study into the spatial variation of the vertical wind field at multiple points between Mawson and Davis stations. These results clearly demonstrate that the vertical wind field can be non-uniform over scales as small as 160 km, comparable to the width of an observing zone of the Mawson instrument. In addition, there have been a number of observations of strong vertical winds (greater than 100 ms^{-1}) above Mawson (and elsewhere, see for example Spencer et al. (1982); Crickmore et al. (1991); Innis et al. (1996, 1999) and the discussion of Section 2.4.2), which, particularly during the current solar minimum conditions, makes them at least comparable to horizontal wind speeds, violating the first assumption above. Thus it was decided that no correction for vertical wind would be made to off-zenith wind measurements in the present analysis, since the available data shows no consistent uniformity over the scales necessary to confidently apply the directly measured zenith wind to off-zenith zones, and that if such a correction were made then the large vertical wind speeds which are sometimes measured could introduce correspondingly large errors into the analysis.

4.11 Software Design

A new suite of control software was written for the Mawson SDI, both during the initial stages of construction and testing at La Trobe and also after the instrument's installation at Mawson station. The majority of the software was written in the Interactive Data Language (IDL), with some hardware-communications support developed in C++ and compiled into an IDL-callable dynamic link library (DLL). Much of the control software was based on early versions of software written by Dr. Mark Conde, which controlled a similar instrument at Poker Flat, Alaska (see for example Conde and Smith (1997, 1998); Conde et al. (2001); Holmes et al. (2005)). While based on the older code, the new software was completely re-written using an Object-Oriented approach. Thus individual tasks such as scan-range calibration (Section 4.4), phase mapping (Section 4.5), and spectral acquisition (Section 4.7) were implemented as individual IDL "objects" (and referred to as "plugins").

A single top-level object, called the "Console", provides an interface between the plugins and the instrument hardware (for example the camera, etalon, and servo-motors). The software was designed in such a way that plugins could be written and added as needed, while the Console automatically searches for compatible plugins and provides the interface for initiating them. The Console is thus responsible for starting plugins, providing them with the most recent images from the camera (or simply a timer "event", in the case that the plugin does not require images, to allow the plugin to update), and also for ensuring there are no clashes between plugins attempting hardware-sensitive operations, such as etalon scanning; the Console ensures that only one plugin is performing such operations at a given time.

4.11.1 Operating Modes and Scheduling

The Console allows for two basic modes of operation, called "manual" and "auto" modes. Manual mode, as the name implies, requires that all operations be initiated manually. In this mode all plugins must be initiated by a human operator. This mode is useful for setting-up and testing the instrument. However for most of the time the instrument operates in auto mode. When operating in this mode, the Console consults a "schedule file" whenever it is not performing another task. The schedule file thus controls all aspects of operation, removing the need for an operator. These schedule files, which are simply text files written in a language which the control software can interpret, contain directives for filter selection, calibration/airglow source selection, spectral acquisition at multiple wavelengths and using different zone maps, as well as scan-range calibration and phase mapping. These directives are placed within 'blocks' of solar elevation angle, which instruct the control software when to execute a certain directive (for example, airglow observations were only scheduled for times when the Sun was more than 10° below the horizon). The schedule thus allows for autonomous observations, and the schedule files can easily be modified remotely and uploaded in order to change the observing schedule. An example block of schedule file commands is given below.

```
ifsea: [-100, -10] [cont]
  ## Record calibration spectra...
     cameraset: [0.2, 200]
     mirror: [drive_cal]
     cal_switch: [1]
     spectrum: [632.8, default_zones.txt,
               'fslaser_cals_Y$_doy$_date_0n$_0d$_632.8nm.las']
  ## Now do interleaved red/green airglow observations...
     cameraset: [0.2, 240]
```

```
    mirror: [drive_sky]
    filter: [2]
    spectrum: [630.0, default_zones.txt,
              'red_sky_Y$_doy$_date_0n$_0d$_630.0_nm.sky']
    filter: [3]
    spectrum: [557.7, default_zones.txt,
              'gre_sky_Y$_doy$_date_0n$_0d$_557.7_nm.sky']
    filter: [2]
    spectrum: [630.0, default_zones.txt,
              'red_sky_Y$_doy$_date_0n$_0d$_630.0_nm.sky']
    filter: [3]
    spectrum: [557.7, default_zones.txt,
              'gre_sky_Y$_doy$_date_0n$_0d$_557.7_nm.sky']
ifsea: [-100, -10] [loop]
```

The `ifsea` (read 'if solar zenith angle') command is used to determine whether or not to execute a command block. The Console first checks to see if the current solar zenith angle lies between the [min, max] range supplied. If it does, and the `ifsea` command ends with a `[cont]` directive, then the command block is entered, and all commands within it are executed in order. If the `ifsea` command is followed by a `[loop]` directive, and the solar zenith angle condition is met, the Console returns execution to the start of the given command block, and loops through again. If the solar zenith angle condition is not met, the Console searches for the next command block.

The `cameraset` command sets the exposure time and gain of the camera. The `mirror` directive tells the console to drive a small mirror out of (for sky observations, `[drive_sky]`) or into (for calibration, `[drive_cal]`) the path of the sky light. `cal_switch` selects a calibration source (up to four sources are selectable, however only two calibration sources were available during 2007). The `filter` command similarly selects the interference filter to use for sky observations (filters were numbered 1-6, corresponding to the filter wheel position, filters 1-3 were available during 2007-2008). The `spectrum` command directs the console to start a spectrum plugin for the specified wavelength. This plugin, once started, remains open until the command `shutdownspex` is issued. The plugin remains passive until another command to take a spectrum at that wavelength is received. The schedule allows a format string to be specified which sets the format of the file name under which the spectra and associated information are stored. The schedule also supplies the name of a zone map configuration file to be used when acquiring spectra. Other tasks such as scan range calibration and phase mapping can also be run from the schedule file; it is also possible to specify a maximum amount of elapsed time between such calibration tasks, after which time the console schedules them to be updated automatically.

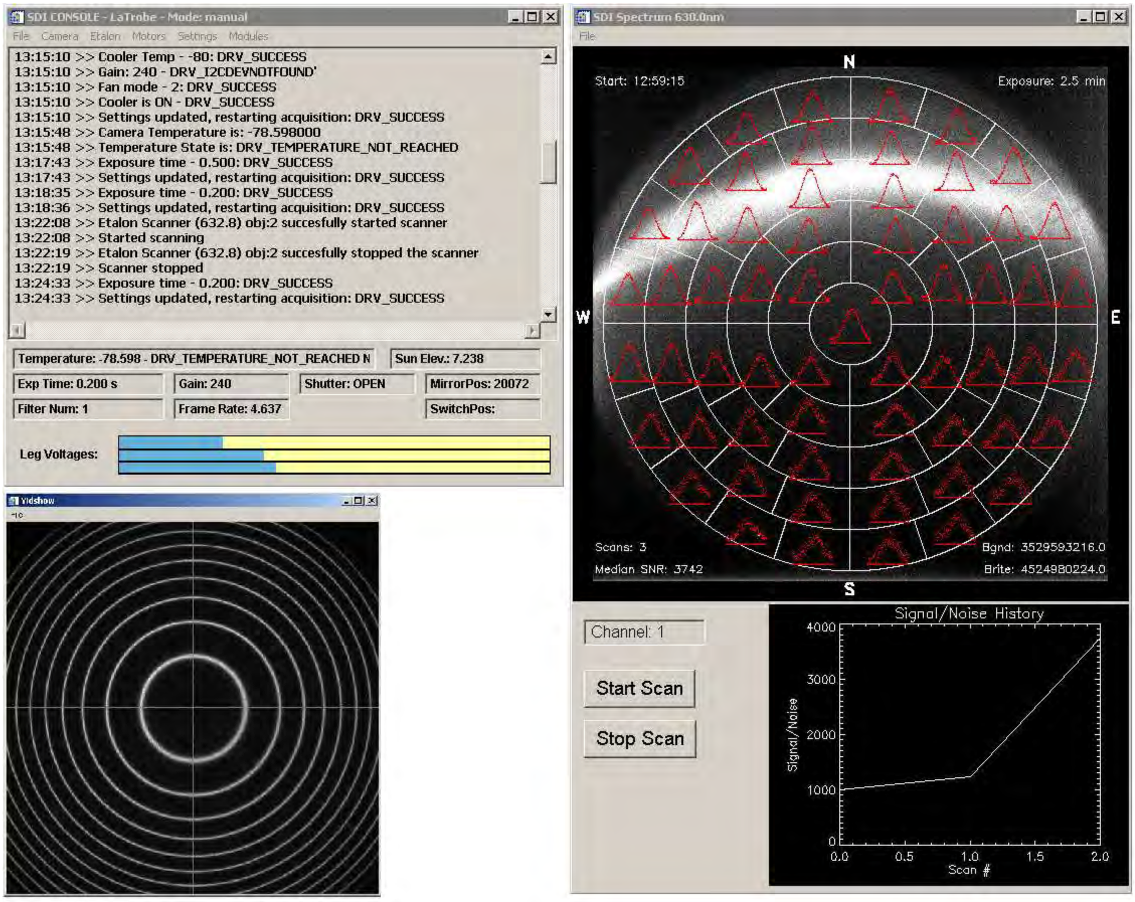


Figure 4.11.1: Console and plugin screenshots. The top left image shows the Console object, below it is the Vidshow plugin which displays the current camera image (a calibration laser fringe in this example), and on the right is the Spectrum plugin, used for acquiring spectra, in this case airglow spectra at a wavelength of 630.0 nm. Displayed behind the acquired spectra is the simultaneously acquired all-sky image, while in the bottom right corner of the spectrum plugin is shown a history of the signal-to-noise ratio for the current exposure.

Chapter 5

The Davis Fabry-Perot Spectrometer

5.1 Introduction

The Davis instrument is a fixed-gap, narrow-field, imaging Fabry-Perot spectrometer (FPS). A schematic is shown in Figure 5.1.1. A periscope mounted beneath an observing dome is used to admit light from a 6° (full-angle) field-of-view. This periscope is steerable to any part of the sky above an elevation angle of 10° (in which position the lower edge of the field-of-view is 7° above the horizon). A small centrally mounted mirror allows for rapid switching between azimuthally opposed fields of view by selecting light from one of two oppositely directed baffle tubes.

After passing through the periscope, the optical beam is collimated before it enters a filter chamber. A 6-position filter wheel currently holds three filters which can be inserted into the path of the beam. Light then passes through a 130 mm (working aperture) etalon, the plates of which are separated by a 12.948 mm air gap, before being focused by a 150 mm, two-element achromatic fringe-forming lens (1000 mm focal length) onto a third generation Gallium Arsenide image intensifier which is part of a Pulnix ICCD camera (768 x 576 pixels, physical size 14.5 x 10.8 mm). Digitised fringe images are passed to a computer where they are stored along with the rest of the night's observations.

Airglow fringes are commonly recorded from the four geographic cardinal directions (at an elevation angle of 60°) plus the zenith. With the installation of the Mawson spectrometer, additional look-directions were added to the observing sequence, directed along the Mawson-Davis great-circle. Calibration fringes, imaged prior to each cycle of airglow observations, are recorded by passing the light from a frequency stabilized laser (at 632.8 nm) through an opal glass screen before it is imaged by the spectrometer. The opal glass screen creates a diffuse light designed to fill the field-of-view of the periscope.

Exposure times are computed dynamically by the control software, and vary depending on the prevailing level of signal in the previous exposure. Typical values are 200 seconds for low signal levels and around 15 seconds during very active aurora. Minimum and maximum exposure times of 1 and 220 seconds are enforced by the control software.

5.2 Data Acquisition

The Davis FPS, in comparison to its Mawson counterpart, is less complicated in its operation. For airglow exposures, the periscope is aligned along a chosen azimuth and elevation angle, and a two-dimensional interference fringe image is recorded by the CCD detector, with an exposure time set by the control software based upon the prevailing level of signal in the previous exposure. For calibration, the periscope is directed downwards toward an opal glass screen which is illuminated by laser light. Since the raw data are the two-dimensional interference fringes (examples of which are shown in Figure 5.3.3), spectra are derived from the spatial structure of the fringes in the radial direction. A simple example of such a scheme would involve locating the centre of the interference fringes (the pixel

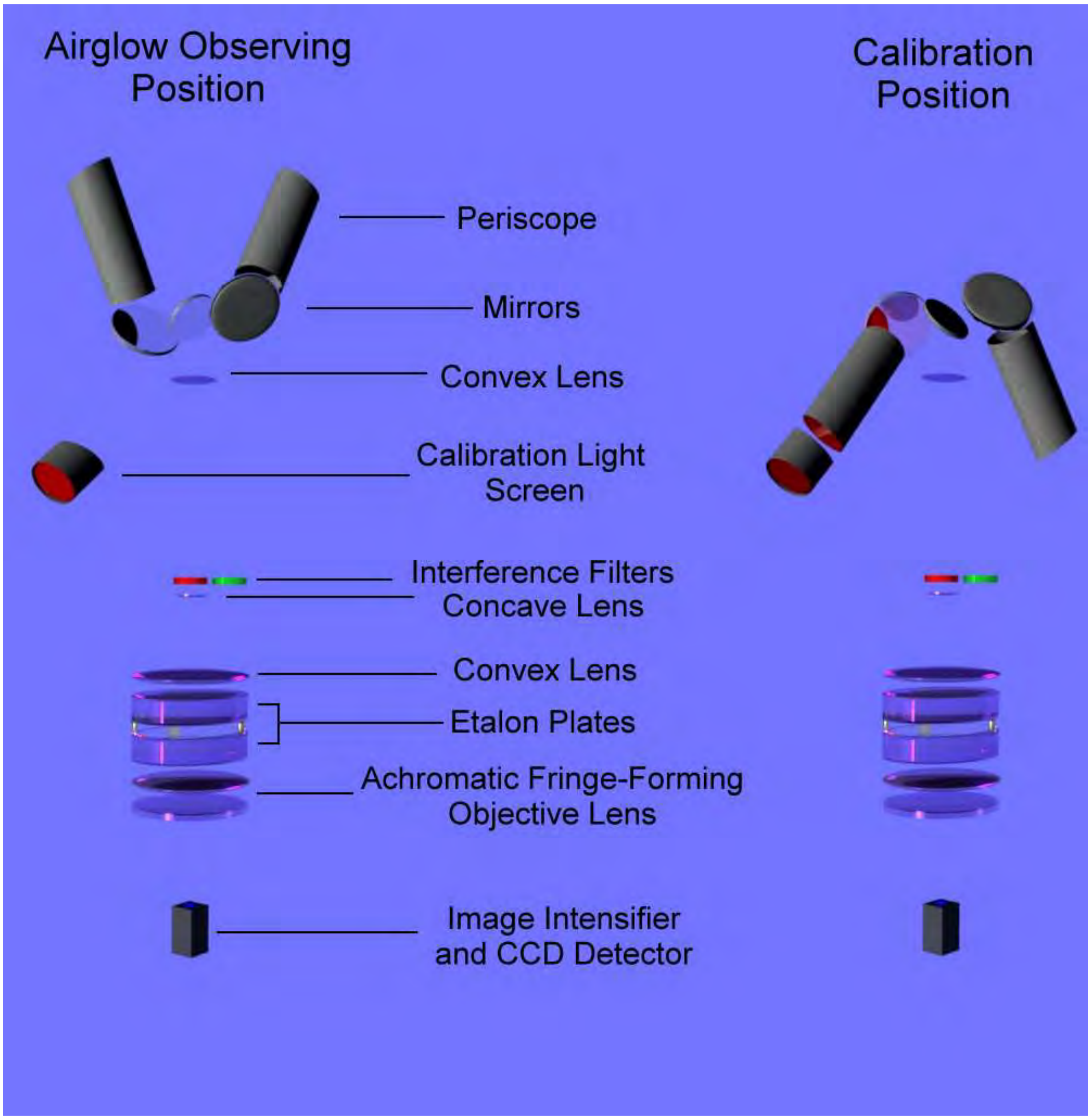


Figure 5.1.1: Optical layout of the Davis spectrometer. The image on the left shows the instrument in position for airglow observing, the image on the right shows the periscope position for observing the calibration laser, which is incident on the calibration light screen (an opal glass screen for producing a diffuse illumination).

corresponding to normal incidence, m_0), binning the image pixels into uniform increments of the square of the radial distance relative to this centre (r^2), and summing the signal within each radial distance bin. This amounts to summing the recorded signal around a circular annulus of finite width, centered on the pixel corresponding to normal incidence, for different values of the annular radius. To a very good approximation, interference order decreases in proportion to r^2, thus a spectrum acquired in this way will be linear in m, or equivalently, λ.

This simple reduction of a two-dimensional fringe image to a one-dimensional spectrum is however not the optimal method, since it assumes that contours of constant interference order on the image are in the form of perfect and mutually concentric circles. Experience has shown that optical aberrations often cause this assumption to be violated, to a degree that can noticeably degrade the recovered spectra (compared to spectra recovered using a method that does account for the aberrations). Any departure from perfect circularity will artificially broaden the resulting spectrum, since pixels sharing the same value of r^2 will no longer necessarily share the same value of m. These departures will inevitably arise in any real instrument, if for example the plane of the focusing lens or detector is not perpendicular to the optical axis. Either of these defects will result in interference fringes whose radii vary with azimuthal angle around the fringe. The preferred reduction scheme then is one which allows for non-circular fringes, and therefore will not artificially broaden spectra derived from such fringes. Such a scheme has been developed by Conde (2002), and has been employed in this work with slight modification in order to apply it to images containing partial fringes (i.e. those not completely sampled in azimuth because they are bigger than the detector). The full reduction procedure is described in the following sections.

5.3 Image Linearisation

The first step in the current analysis scheme is to correct recorded fringe images for a known non-linearity in the response of the detector. Each pixel on the detector responds to incident light by recording a value between 0 - 255 (herein referred to as 'grey levels'). Tests carried out on the detector (T. Davies, private communication, 2008) revealed that for a given source of incident light, the pixel grey level was not directly proportional to the product of exposure time and calibration source intensity as would be expected of a linear detector. The result is that for a given fringe image, the ratio of grey levels returned by two different pixels will *not* in general be an accurate representation of the ratio of the incident intensities. This non-linearity is believed to be related (at least to a good approximation) to the total integrated brightness (recorded grey level) recorded by the detector.

From the observed detector response (grey levels) as a function of illumination level (exposure time multiplied by source intensity), as shown in the upper panel of Figure 5.3.1, an illumination value can be calculated for a given value of the grey level by inverting the response curve and scaling the illumination axis such that the maximum value returned is equal to the maximum grey level of 255. This curve then describes the mapping between raw grey levels and 'linearised' grey levels, which corresponds to a fringe image which has been corrected, to a good approximation, for the detector non-linearity (this mapping is shown in the lower panel of Figure 5.3.1). The polynomial which describes this mapping

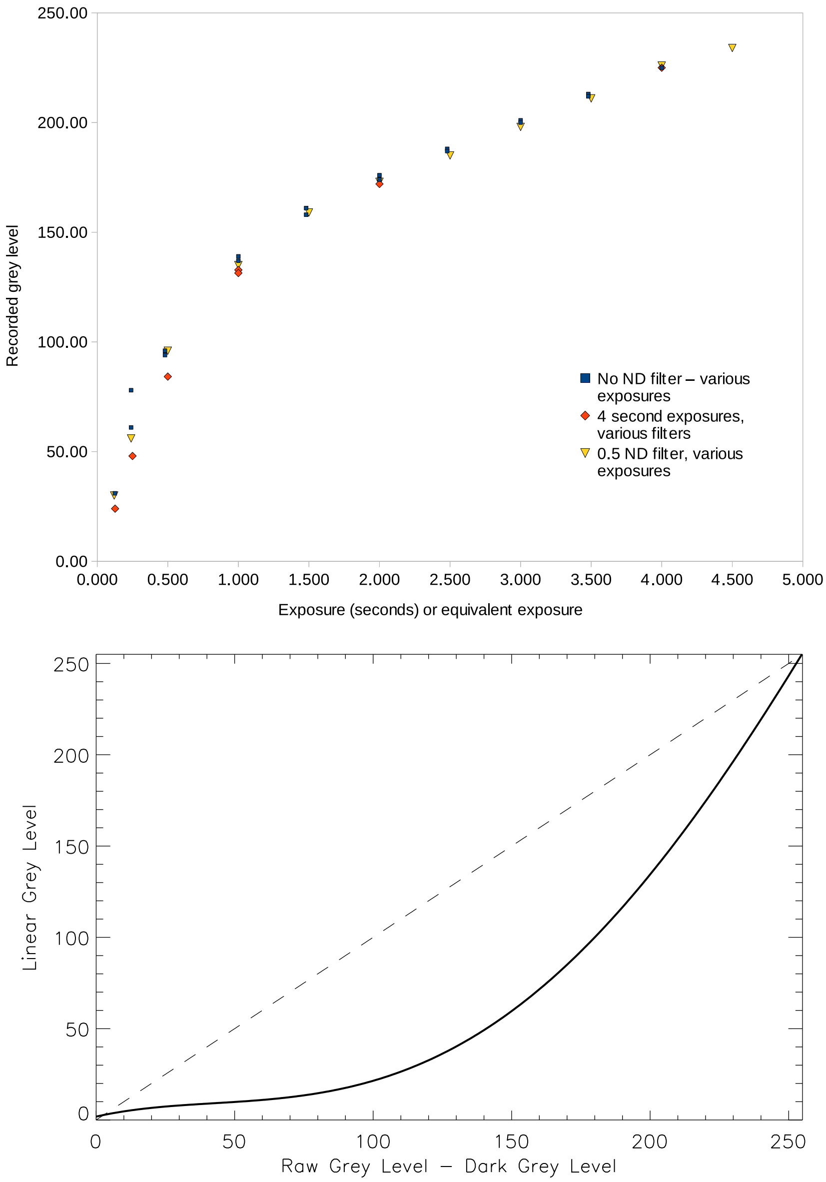


Figure 5.3.1: Upper panel: the detector response measured during three experiments: without neutral density (ND) filters, using various exposure times; 4 second exposure times, using various ND filters; 0.5 ND filter, using various exposure times. Lower panel: the mapping between raw grey levels and linearised grey levels for the Davis CCD detector. Figures reproduced from T. Davies, private communication, 2008.

is given by:

$$y = 1.8440 + ax + bx^2 + cx^3 + dx^4 \tag{5.3.1}$$

where:

$$a = 3.4169 \times 10^{-1}$$
$$b = -6.2306 \times 10^{-3}$$
$$c = 5.6317 \times 10^{-5}$$
$$d = -8.5645 \times 10^{-8}$$

with x and y corresponding to the raw (dark level subtracted) and linear grey levels respectively. Dark level subtraction is necessary because the dark grey level (grey levels recorded with no illumination, denoted by C_{dark}) were found to respond linearly to the exposure time T_{exp} (in contrast to the case where illumination was present), according to the equation:

$$C_{\text{dark}} = 11.398 + 0.1373\, T_{\text{exp}} \tag{5.3.2}$$

Thus, in order to correct a fringe image for the non-linear response of the detector, the value of C_{dark} (for which the exposure time must be known) is first subtracted from the raw grey levels. Grey levels are then mapped to their linear values using Equation 5.3.1, and the value of C_{dark} is added back to the linearised image.

Figure 5.3.1 (lower panel) shows the mapping between dark subtracted grey levels and linearised grey levels. The curve is approximately linear for low values of illumination, but

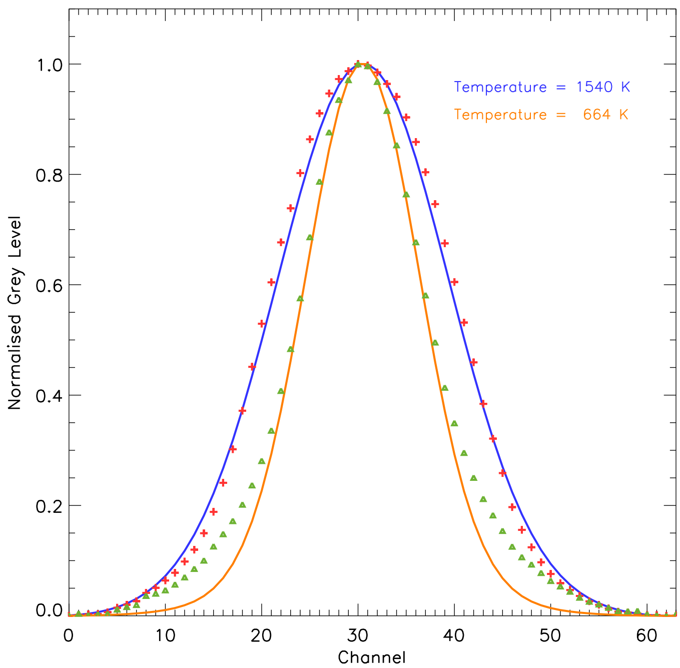


Figure 5.3.2: Sky spectra before (blue curve, red crosses) and after (orange curve, green triangles) linearisation. The solid curves are the emission spectra fitted to the derived spectra represented by the plotting symbols. The temperatures estimated from each of these spectra are shown.

departs significantly from linearity above a grey level of approximately 100. The linearity correction is therefore most significant for fringe images recorded when intensities are high (and also for laser calibration images). In terms of derived spectra, linearisation only affects the widths (temperatures) of derived spectra, and not their peak wavelengths. Figure 5.3.2 shows an example sky spectrum obtained with and without the linearity correction. The sky image from which these spectra were derived was a particularly bright image, and as such the effect of linearisation is very clear. In this case linearisation reduced the estimated temperature by approximately 57%.

When the linearisation scheme outlined above is not implemented, returned temperatures are unreasonably large (often exceeding 2000 K, which, during the current solar minimum conditions, is unrealistic) and also show significant scatter. Implementing the linearisation scheme results in temperatures which are more coherent (showing significantly less scatter), but also lower than would be expected under the current solar conditions (compared with temperature measurements from nearby Mawson station, for example, and by comparison with the empirical NRLMSISE-00 model). It is also evident from Figure 5.3.2 that the linearised spectra are not well fitted by the convolution of the instrument function with a Gaussian source spectrum, which suggests that the linearisation scheme may not be properly correcting for the non-linearity at low grey levels. The sensitivity of the linearisation scheme to the accuracy of the fitted polynomial has not been tested, and thus it is likely that with further investigation reliable temperatures could be obtained. However, in the present study, no use has been made of the temperatures returned by the Davis FPS. It is stressed again that wind estimates returned by this instrument remain unaffected by the camera non-linearity (and also the linearisation scheme).

5.4 Image Cleaning

Linearised calibration and airglow fringe images are then passed through an IDL routine which attempts to remove any noise due to cosmic rays striking the detector. Figure 5.3.3 shows recorded calibration and airglow fringe images both before and after this cleaning step. Bleeding of charge from cosmic ray hits can sometimes affect a large area on the detector, and may distort derived spectra and fitted model fringes. At best they will increase spectral noise. The image cleaning step is thus important for improving signal-to-noise and reducing any distortions which may be caused by such localised image artifacts.

5.5 Model Fringe Fitting

Cleaned and linearised calibration fringe images are then fitted with a model fringe pattern. The purpose of the model is to describe the variation of interference order m with pixel location on the detector (x, y), allowing for various distortions from the perfect circles of an ideal etalon. It is understood that strong local distortions or discontinuities will not

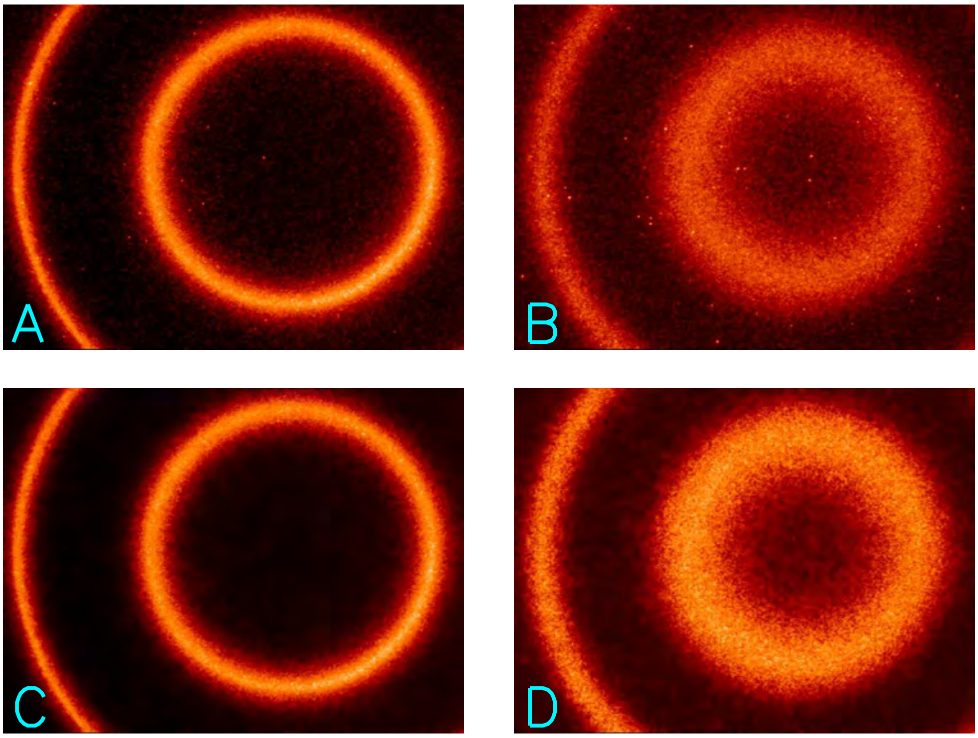


Figure 5.3.3: False-colour examples of interference fringes recorded by the Davis spectrometer. Original (A) and cleaned (C) laser calibration fringe at 632.8 nm. Original (B) and cleaned (D) airglow (sky) fringe at 630.0 nm. The apparent change in brightness between the raw and cleaned images is a result of scaling, and not due to an actual increase in pixel values. These images demonstrate the effect of cleaning images with typical levels of cosmic ray noise.

be described by the model fringe[1], and only those distortions that vary slowly across the fringe image can be characterised in this way. Note that this process of fitting a model fringe to a calibration fringe image is directly comparable to the phase mapping procedure required for the Mawson spectrometer. In the case of Mawson, since the phase map is built up directly from calibration laser fringes, it will automatically account for any fringe distortions. Below, an outline is given of the fringe-fitting method originally described by Conde (2002), and this is followed by a discussion of the modifications to this method which were required to properly fit the asymmetric fringe images recorded by the Davis FPS.

Following Conde (2002), the actual image of the fringe pattern is formed by passing the light emerging from the etalon through a positive-power lens before it reaches the detector. In this way there exists a unique mapping between pixel locations (x, y) on the detector and incidence angles (θ, ϕ) through the etalon. This mapping is defined by the functions:

$$\theta = L_\theta(x, y) \tag{5.5.1}$$

$$\phi = L_\phi(x, y) \tag{5.5.2}$$

where θ is the axial angle (measured form the plate normal direction) and ϕ is the az-

[1]Fortunately, strong local distortions are highly unlikely for the optical configuration used.

imuthal angle. The interference order given by Equation 3.1.5 then becomes:

$$m(x,y) = \frac{2\mu t \cos(L_\theta(x,y))}{\lambda} \tag{5.5.3}$$

In the case of the Davis etalon we are dealing with small θ angles, and thus the cosine term can be expanded as $\cos(x) \simeq 1 - \frac{x^2}{2!}$. Equation 5.5.3 is then approximated by:

$$m(x,y) \simeq m_0 - \frac{\mu t L_\theta^2(x,y)}{\lambda} \tag{5.5.4}$$

where m_0 represents the interference order at normal incidence. In a distortion-free imaging system the L_θ function will map incidence angle θ (for small angles) linearly to radial distance from the fringe centre, with some magnification α, i.e.:

$$L_\theta(x,y) = \alpha\sqrt{(x-x_0)^2 + (y-y_0)^2} \tag{5.5.5}$$

Substituting Equation 5.5.5 into Equation 5.5.4 gives:

$$m(x,y) \simeq m_0 - \beta[(x-x_0)^2 + (y-y_0)^2] \tag{5.5.6}$$

where β incorporates the magnification α and is given by $\beta = \frac{\alpha^2 \mu t}{\lambda}$. Substituting 5.5.6 into the Airy function (equation 3.1.11) will produce a model fringe pattern described by the 4 free parameters:

$$\begin{aligned} m_0 &= \text{interference order at normal incidence} \\ \beta &= \text{magnification} \\ (x_0, y_0) &= \text{fringe centre coordinates (at normal incidence)} \end{aligned}$$

These fringes will form perfect circles since no allowance has been made for asymmetric magnification or warping. Such warping can occur if for example the objective lens is not centered on the optical axis, if the plane of the lens is not perpendicular to the optical axis, or if the plane of the detector is not perpendicular to the optical axis. To incorporate the possibility of a fringe radius that varies with azimuthal angle around the fringe, two parameters are introduced to replace the x and y distances from the centre:

$$\begin{aligned} u &= (x-x_0) + \gamma_x|x-x_0| \\ v &= (y-y_0) + \gamma_y|y-y_0| \end{aligned}$$

where $\gamma_{x,y} = 0$ corresponds to no warping, and $\gamma_{x,y}$ can take both positive and negative values. Ellipticity is allowed for by setting:

$$\psi(x,y) = \beta_x u^2 + \beta_y v^2 + \beta_{xy} uv \tag{5.5.7}$$

Equation 5.5.6 thus becomes:

$$m(x,y) = m_0 - \psi(x,y) + \epsilon\psi^2(x,y) \tag{5.5.8}$$

where a barrel aberration (magnification decreasing outward from the centre) is achieved with $\epsilon < 0$ and a pincushion abberation (magnification increasing outward from the centre) with $\epsilon > 0$. The final fringe model is therefore given by Equation 5.5.8 and it responds to variations in 9 different parameters: m_0, x_0, y_0, γ_x, γ_y, β_x, β_y, β_{xy}, and ϵ. Figure 5.5.1 shows four examples of the effect of these parameters on the model fringe. Synthetic fringes have been used in this figure to illustrate the effects of varying selected model fringe parameters, however these fringes contain more orders, and are better centered, than the fringe images produced by the Davis spectrometer. As a result, the Davis fringe images cannot uniquely constrain all of the parameters described in the model so far; the required adjustments to the model and fitting procedure for the case of the asymmetric Davis fringe images are described in Section 5.6 below.

The model defined by Equation 5.5.8 describes the variation of interference order across a 2-dimensional image (compare this with the phase map of Section 4.5). The result of substituting Equation 5.5.8 into the Airy function (Equation 3.1.11) is a set of model fringes, denoted $A_M[m(x, y)]$, which are no longer necessarily perfectly circular. The choice of reflectance R in the Airy function determines the thickness of the model fringes.

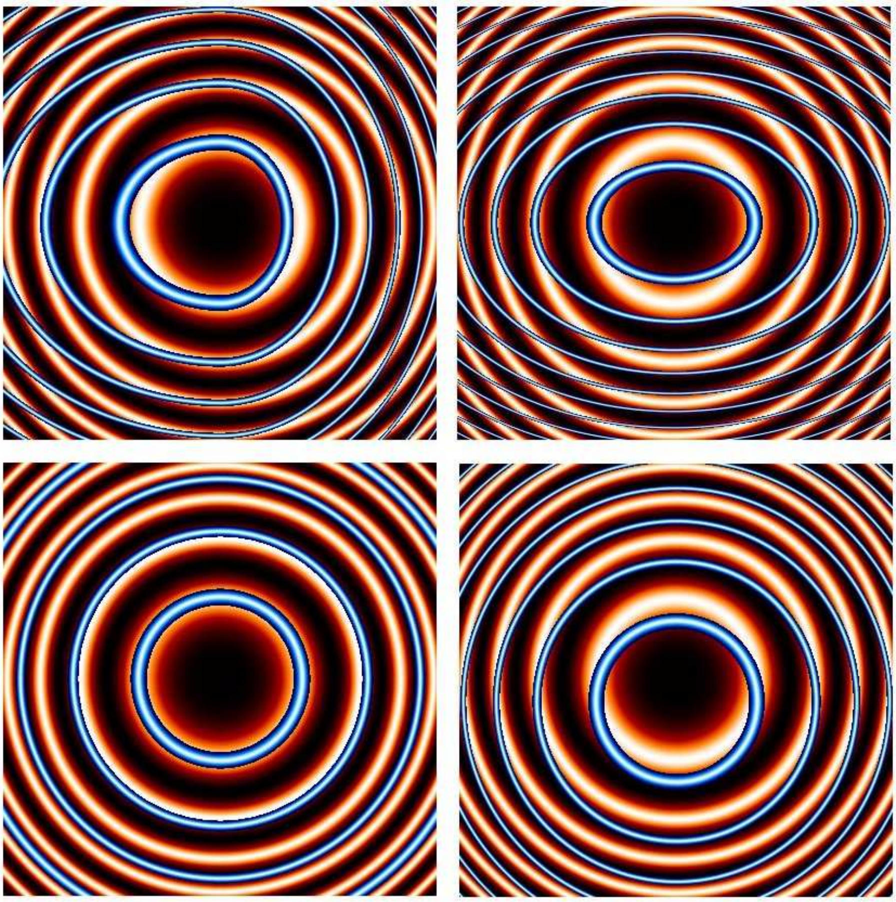


Figure 5.5.1: Exaggerated effects of various incorrect parameter choices on the model fringe. The four figures show model fringes (blue) superimposed on simulated laser fringes (orange), where the model fringes have been generated with incorrect choices of the parameters (clockwise from top left) γ_x, β_y, ϵ, and y_0.

In the current analysis scheme the reflectance is chosen to produce model fringes which are narrow relative to the sky fringes. If an observed calibration fringe image is denoted by $B_c(x, y)$, and the size of each fringe image is $N_x \times N_y$ pixels, then a quantity which describes how well the model fringe represents the observed calibration fringes (the 'goodness of fit') is their discrete 2-dimensional cross-correlation value, κ, given by:

$$\kappa = \sum_{x=1}^{N_x} \sum_{y=1}^{N_y} \{\overline{A_M}[m(x, y)] \times \overline{B_c}(x, y)\} \tag{5.5.9}$$

$$\text{where :} \tag{5.5.10}$$

$$\overline{A_M}[m(x, y)] = A_M[m(x, y)] - \min\{A_M[m(x, y)]\} \tag{5.5.11}$$

$$\overline{B_c}(x, y) = B_c(x, y) - \min\{B_c(x, y)\} \tag{5.5.12}$$

and min{} is used to represent the function which returns the minimum value of the item within the braces.

With all other parameters held constant, the optimal value for a particular parameter is taken to be that which maximises κ. The general procedure is then to perform this cross-correlation iteratively for a range of values of each free parameter in turn (sometimes called a 'grid-search'), starting from some reasonable initial values. It has been found that for symmetric fringe images (those containing at least three orders of interference completely sampled in azimuth) there is little interdependence between the various free parameters in the model, so that varying one parameter to maximise κ has little effect on the cross-correlations of the other parameters. As such, acceptable convergence can be achieved after a fixed number of iterations. For fringe images not satisfying this condition (as the Davis fringe images do not), a slightly modified fitting procedure is required. This is described in the following section.

5.6 Application to Asymmetric Fringe Images

The model fringe fitting outlined in Section 5.5 above was designed to work with images which contain at least three complete orders of interference (a minimum of three orders is required to fit the quadratic term parameterised by ϵ). The Davis interference images contain one complete order and an azimuthal fraction (approximately $\frac{1}{4}$) of a second. Thus this parameter (which describes the magnitude of the barrel/pincushion distortion) is not suitable for analysis of Davis fringe images, and must be omitted. Also, the presence of the partial fringe causes the fitting algorithm to incorrectly warp the model fringe (that is, it produces incorrect choices for the γ coefficients) while searching for the best overall fit to both fringes. In this case the maximum cross-correlation value κ does not necessarily represent the best fit between the model and observed fringes. If two full fringes were present, the signal contained within each should be comparable, however since only a fraction of the second fringe is imaged, the majority of the airglow signal is contained within the central fringe. It is thus important that the central fringe be fitted accurately, since the greatest contribution to the derived spectrum will come from it.

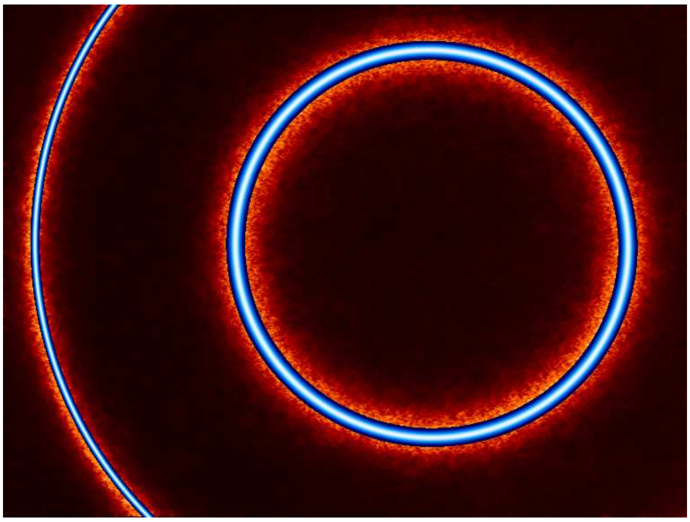


Figure 5.6.1: Cleaned laser fringe from Figure 5.3.3 (red) with fitted model fringe (blue) superimposed.

The fitting procedure outlined in the previous section has therefore been modified to apply it to the special case represented by the Davis fringe images. The ϵ parameter is not used, as an insufficient number of fringes are imaged to allow this parameter to be uniquely determined. The fitting procedure itself is split into three stages. First, the parameters corresponding to the fringe centre, x_0 and y_0, and the order at normal incidence m_0, are fitted using both the complete central fringe and the partial outer fringe to ensure that the model fringe is centered on the observed fringes, and of the right size. Secondly, the free parameters (excluding ϵ and m_0) are fitted to the central fringe *only*. This ensures that the model fringe accurately characterises the central fringe which will contribute most to the derived spectrum. Finally, a small correction is made to the m_0 term by fitting it to both the central and outer fringes. Considerable experience[2] has shown that the result is a model fringe which best represents the Davis interference images, biased as it is toward the central fringe. A model fringe fitted to an actual calibration fringe image, using this procedure, is shown in Figure 5.6.1.

5.7 Fringe Reduction

The result of fitting the free parameters in the model fringe is an empirical relation describing the variation of interference order m (relative to the order at normal incidence) across a laser calibration image, recorded at the wavelength of the laser, λ_0. This effectively characterises the response of the etalon and fringe-imaging lens and detector assembly at the time the laser fringe was imaged. In order to apply this fringe model to an interference image of the sky (described by $B(x, y)$), imaged at the wavelength λ, Equation 5.5.4 gives

[2]The technique has now been applied to a total of 93,385 laser fringes, and thereby used to generate a total of 536,643 sky spectra, from six years of instrument operation.

m at this new wavelength as:

$$m(x, y, \lambda) = m(x, y, \lambda_0)\frac{\lambda_0}{\lambda} \tag{5.7.1}$$

This is equivalent to wavelength-shifting a phase map from the Mawson SDI. Since the usual observing sequence at Davis is such that the spectrometer records one calibration fringe image for each cycle of airglow fringe images (in multiple look directions), some interpolation in time is usually performed before a fitted phase map is applied to an airglow fringe image. Equation 5.7.1 was encountered previously in the description of the Mawson phase map (Section 4.5). In this case, however, a continuous phase map results from the fringe fitting procedure directly, and so no 'unwrapping' is needed before the Davis phase map is scaled to a new wavelength. From a single interference fringe image, it is the spatial variation of m which allows a spectrum to be derived. As before, the phase map is used to define the spectral channel q_{spec} to which the value of $B(x, y)$ will be added. For a spectrum containing Q channels, this spectral channel is given by:

$$q_{\text{spec}}(x, y) = [m(x, y, \lambda) \bmod 1] \times (Q - 1) \tag{5.7.2}$$

In a symmetrical fringe image, for which the number of pixels contributing to each spectral channel will be comparable, a spectrum $Y(q)$ is obtained by simply adding the value of $B(x, y)$ to $Y(q_{\text{spec}})$ for all pixels in the image. For the asymmetrical Davis fringes, however, it is clear from Figure 5.7.1 that the number of pixels contributing to each spectral channel is *not* comparable, and therefore simply adding $B(x, y)$ to $Y(q_{\text{spec}})$ would result in a distorted spectrum. The unequal distribution of spectral channels can be accommodated by weighting the contribution to each spectral channel by the number of pixels contributing to that channel. Let $N(q_{\text{spec}})$ denote the function which describes the number of pixels contributing to spectral channel q_{spec} (an example of which is shown in the right panel of Figure 5.7.1). A spectrum is obtained by looping through all the pixels in the fringe image, and for each pixel (x_i, y_i) performing the following steps:

1 Calculate $q_{\text{spec}}(x_i, y_i)$ using Equation 5.7.2.

2 Calculate the contribution-weighted signal recorded by the pixel:

$$B_w(x_i, y_i) = \frac{B(x_i, y_i)}{N(q_{\text{spec}}(x_i, y_i))}$$

3 Add the weighted signal to the appropriate spectral channel:

$$Y(q_{\text{spec}}) = Y(q_{\text{spec}}) + B_w(x_i, y_i)$$

After performing the above steps for all pixels in the image (in this case $0 \leq x_i \leq 767$, $0 \leq y_i \leq 575$), $Y(q)$ contains the recorded airglow spectrum (convolved with the instrument function). An example spectrum, derived in this way, is shown in Figure 5.3.2.

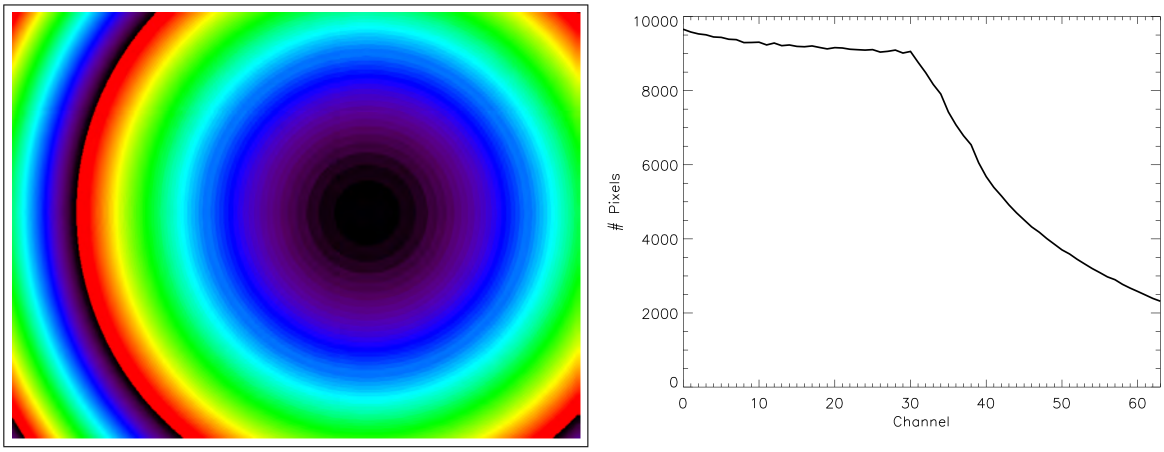


Figure 5.7.1: Left: The phase map (wrapped) produced by the model fringe fit to the laser image shown in Figure 5.6.1. Color represents channel number in the range 0 (black) - 63 (red). Right: Histogram showing the number of pixels contributing to each spectral channel, $N(q_{spec})$, in the phase map on the left.

5.8 Deriving Line-of-Sight Wind Speed

Unlike the Mawson instrument, for which spectral acquisition is an inseparable part of its operation, the Davis instrument simply outputs recorded fringe images (along with some housekeeping information), and the reduction of these images to one-dimensional spectra is left to the analysis software. The following summary, which describes how fringe images are reduced to time-series' of winds and temperature, will assume for simplicity that an entire night of observational data is available, however it should be kept in mind that this analysis is currently applied in near real-time, with only slight modification. The analysis proceeds as follows:

1. Fringe images are linearised (Section 5.3) and cleaned (Section 5.4), before being separated into calibration and airglow images.

2. Each calibration fringe image is fitted with a model fringe as described in Sections 5.5 and 5.6. The fitted model fringe for each calibration image is used to derive a spectrum from that image, as outlined in Section 5.7. Davis spectra are commonly derived using $Q = 64$ channels. Laser spectra are later used as instrument functions and for tracking etalon drift.

3. Parameters describing each model fringe (or each phase map) are linearly interpolated to the times at which airglow fringe images were recorded. The phase maps defined by these interpolated parameters are scaled from the laser wavelength (λ632.8 nm) to the airglow wavelength (λ630.0 nm) using Equation 5.7.1.

4. Spectra are derived from each airglow fringe image, using the interpolated and scaled phase maps, as described in Section 5.7.

5. Derived spectra are fitted with an emission spectrum using the algorithm described in Section 4.8 (i.e. the same spectral fitting algorithm as is used for the Mawson SDI).

The first laser spectrum of the night is used as the instrument function. Estimates of peak wavelength (in channels), temperature, background and emission intensity are returned by this algorithm.

6. As described in Section 4.9, etalon drift, measured by the change in peak position of the laser spectra over the course of the night, is subtracted from the airglow peak positions. In the present analysis scheme, all sky and laser spectra are derived using the m_0 parameter of the first fitted laser fringe image of the night. Thus all spectral peak shifts are essentially relative to this first laser fringe, and any instrumental drift will be tracked by the time-variation of the derived laser peak positions (an equivalent scheme could involve using the time-varying m_0 parameter to derive spectra, in which case instrumental drift would already be accounted for). A zero-velocity reference is then calculated from the time-series of the drift-corrected zenith peak positions, and this reference is subtracted from all airglow peak positions.

7. Peak positions are converted to line-of-sight velocities using the Doppler formula (Equation 4.9.3). Using an etalon gap value of 12.948 mm, and $Q = 64$ channels, the conversion from peak position to line-of-sight velocity is given by:

$$\begin{aligned}\frac{\Delta v}{Q} &= \frac{(3 \times 10^8)(1.53 \times 10^{-11})}{64 \times (630.0 \times 10^{-9})} \\ &= 114 \text{ ms}^{-1}\text{channel}^{-1} \qquad (5.8.1)\end{aligned}$$

8. The horizontal component $H_{\|}$ of the line-of-sight wind sampled at zenith angle ϕ is then given by Equation 4.10.3, where V_z is set equal to zero, in keeping with the discussion of Section 4.10.1.

An example of the analysis output from the Davis FPS is shown in Figure 5.8.1. The figure shows the north and south-looking estimates of the geographic meridional wind (top panel), the east and west-looking estimates of the geographic zonal wind (second panel from top), the vertical wind (third panel from top) and the intensities recorded in each of the cardinal and zenith look directions (lower panel). As mentioned in Section 5.3, temperatures returned by the Davis FPS analysis scheme were not used in the present study, and thus are not displayed in Figure 5.8.1. However, recent work has shown that the detector non-linearity is now sufficiently well characterised that the Davis temperature data are generally good.

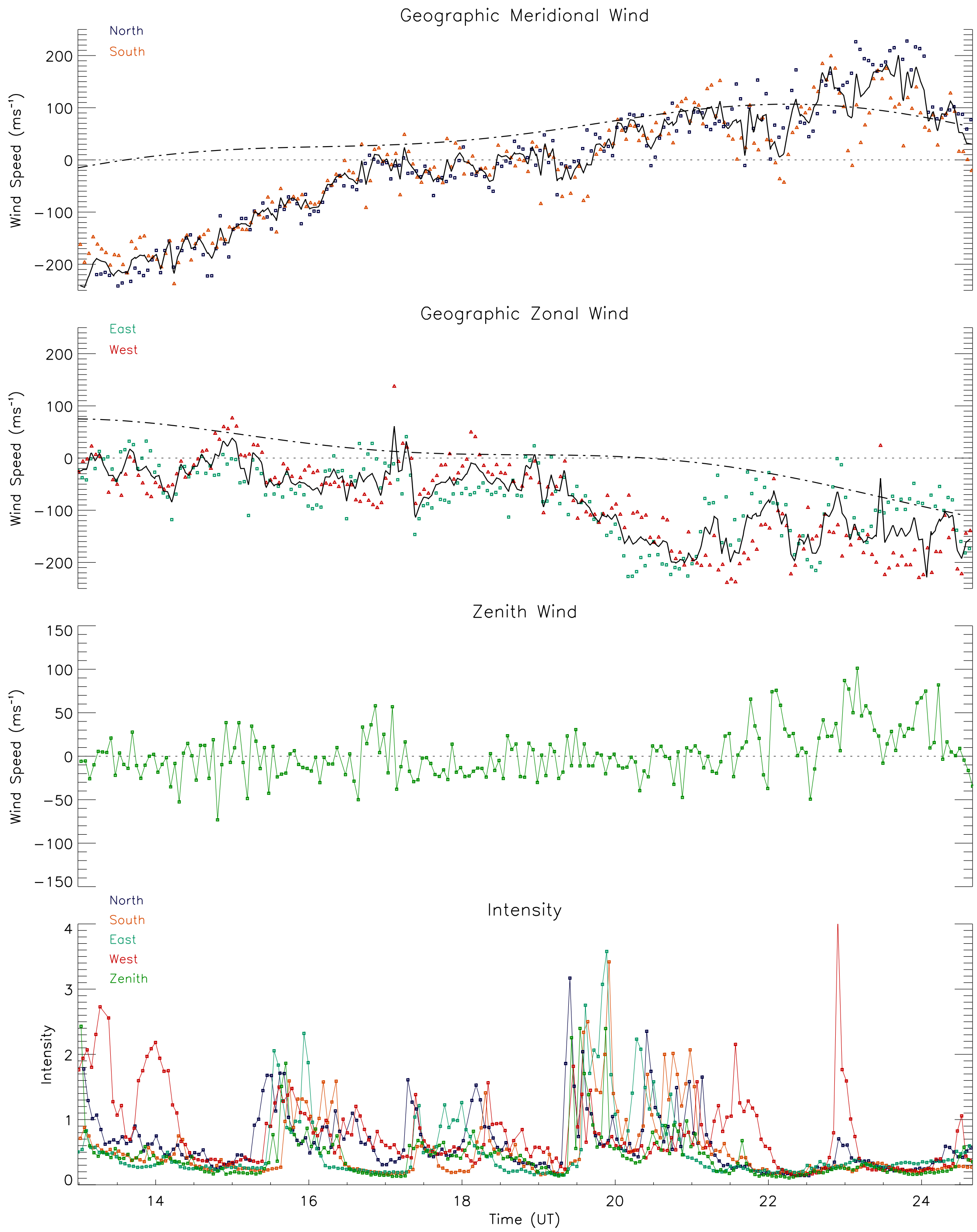


Figure 5.8.1: Analysis output from the Davis FPS, for April 12, 2005. Top panel: geographic meridional winds derived from north- and south-looking observations (symbols, colours indicated top left of panel), meridional wind derived from the average of (interpolated) north and south observations (solid black line) and the meridional winds predicted by HWM93 (black dash-dot line). Second panel from top: same as top panel, except for the east and west (geographic zonal) look directions. Third panel from top: zenith wind. Lower panel: intensities recorded in each look-direction (colours indicated top left of panel). Geomagnetic activity on this day was relatively high, with a mean *ap* of 30 and mean Dst of approximately -47 nT. Solar activity however was low, with daily $F_{10.7} \sim 85.3 \times 10^{-22}$ Js^{-1}m^{-2}Hz^{-1}.

Chapter 6

Daily Variability and Climatological Behaviour

6.1 Introduction

This chapter presents measurements of winds and temperatures derived from observations of the 630.0 nm airglow line of atomic oxygen. The first part of this chapter focuses on the daily variability observed in the large-scale ($\sim$ 1000 km) winds and temperatures above Mawson, and how such variability is affected by the presence of aurora and increasing levels of geomagnetic activity. The second part presents results of climatological analyses which made use of all available data from both the Mawson and Davis spectrometers to investigate the average behaviour of temperatures above Mawson and winds above both Mawson and Davis under two ranges of geomagnetic activity.

The deep solar minimum period during which these data were collected were extremely quiet in terms of solar and geomagnetic activity. Therefore, in this work, conditions for which $ap \leq 10$ are considered 'quiet', and 'active' conditions are defined as those for which $ap > 10$. While 'active' conditions as defined here would not be considered active under normal circumstances, for the current data set the division serves to capture the difference in average behaviour under two different forcing regimes, while ensuring that sufficient data were present in each division to give meaningful results. These average data fill-in the large-scale picture of wind behaviour within both the polar cap and auroral oval regions, and provide a context for the local-scale dynamics to be presented in Chapters 7 and 8. The initial data reduction procedures, used to derive horizontal wind vectors in the case of Davis and temperature and 2-dimensional vector wind fields in the case of Mawson have already been described (Chapters 5 and 4). Further analysis procedures are described in the relevant sections below.

6.2 Daily Variability

The following series of figures summarise four individual nights of observation of the 630.0 nm airglow line above Mawson, and give a clear indication of the types of behaviour observed in derived winds and temperatures under different levels of geomagnetic activity. The first two nights (day numbers 163 and 110, Figures 6.2.1 and 6.2.2) represent 'quiet' times, where the *ap* index was less than 10 for the entire night, Dst was generally of small magnitude ($-10 < \mathrm{Dst} < 10$), and IMF B_z was close to zero or showed relatively little variability. The second two nights (day numbers 165 and 143, Figures 6.2.3 and 6.2.4) occurred during more active geomagnetic conditions, with elevated *ap*, strongly negative Dst, and highly variable or consistently southward IMF B_z. The figures have been ordered approximately in order of increasing geomagnetic activity, based on the indicators just mentioned.

In addition to plotting the wind components (zonal and meridional), temperatures and intensities estimated from each zone in each exposure (plotted as coloured dots), an all-sky median value of each of these parameters is also plotted as a solid black curve. This median was calculated from the values of a particular parameter across all zones within an exposure, and is intended to show the average ('background') behaviour throughout the night, in contrast to the zone dependent values, which show how much variability was present within each exposure. Note that since only the central zone measures vertical wind directly, only the values from this zone are shown in the vertical wind plot. Figures 6.2.1 - 6.2.4 are laid-out as follows. The left hand side, from top to bottom, shows:

1. Magnetic meridional component of the fitted vector wind field calculated in each zone (coloured dots, the zone number scale is indicated bottom left) and averaged across the entire field-of-view (solid black line). Also shown is the meridional wind component predicted by HWM93 for a height of 240 km (dashed black line).
2. Magnetic zonal component of the fitted vector wind field calculated in each zone (coloured dots) and averaged across the entire field-of-view (solid black line). Overplotted is the zonal wind component predicted by HWM93 for a height of 240 km (dashed black line).
3. Zenith wind along with error bars indicating the 1σ uncertainty in the wind estimate returned by the spectral fitting algorithm.
4. Temperature derived in each zone (coloured dots) and averaged across the field-of-view (solid black line). The temperature predicted by NRLMSISE-00, at a height of 240 km, is also shown (solid red line).
5. Intensity derived in each zone (coloured dots) and averaged across the field-of-view (solid black line).

On the right hand side, from top to bottom, are plotted:

1. *ap* index value.
2. Dst index value.
3. The B_z component of the IMF as recorded by the Advanced Composition Explorer (ACE) satellite, and delayed to approximately account for travel time between satellite position and magnetopause.

4. The horizontal divergence (or convergence if negative) of the fitted vector wind field. The magnetic north-south (blue line) and east-west (red line) components are shown along with the total horizontal divergence (black line).
5. Exposure time.

6.2.1 Quiet-time Observations

Figures 6.2.1 and 6.2.2 are good examples of the quiet-time behaviour of the winds and temperatures observed above Mawson. Wind flow was magnetically south-eastward (antisunward) in the magnetic afternoon, and remained eastward for most of the night. The beginnings of northward flow were seen around 23:00 magnetic local time (MLT). Typical wind speeds were around 100 m.s^{-1}. Intensities on both nights were quite low prior to the northward-flow transition, and that low level of signal was reflected in the long ($\sim$ 20 minute) exposure times during those periods.

Winds modelled by HWM93 were very similar for both nights. Magnetic meridional winds were overestimated by the model, which predicted an almost linear change from magnetic southward to magnetic northward wind throughout the observation periods, whereas the observed winds tended to maintain a constant southward component until after approximately 20:00 MLT, when the change from southward to northward was observed to occur. Maximum northward wind speed was overestimated by the model by around 50-100 m.s^{-1}. Conversely magnetic zonal wind speeds were generally underestimated by the model. While HWM93 predicted an approximately sinusoidal variation in magnetic zonal wind between 0-50 m.s^{-1} during the period 14:00-00:00 MLT, observed winds were consistently magnetically eastward, with speeds of around 50-100 m.s^{-1}. Observed zonal winds did not begin to decrease until around 01:00-02:00 MLT. On day 163, the decrease in magnetic eastward wind was observed approximately one hour later than predicted by HWM93. Similarly on day 110, when observed zonal winds were actually seen to increase, HWM93 predicted a decrease, however the time at which zonal winds were observed to reach $\sim$ 0 m.s^{-1} was in reasonable agreement with the model.

Vertical wind activity was significantly greater on day 110 compared to day 163, and this activity was seen to coincide with the period of elevated intensities (23:30-02:00 MLT), after a sudden increase in intensity related to diffuse aurora covering most of the field-of-view. During this period very large vertical winds were observed, up to 150 m.s^{-1}. Short (approximately 1 minute) exposure times during this period highlighted the significant variability of the vertical winds on very short time scales, in contrast to the reasonably consistent horizontal winds observed on this day. Vertical winds were much less variable on day 163, where magnitudes were generally less than 50 m.s^{-1}. Total horizontal divergence in the wind field observed on these quiet days was generally very small, in the range $[-0.2, 0.3] \times 10^{-3}$ s^{-1} (note that the divergence scales differ between each of the Figures 6.2.1-6.2.4). Of the two days, the largest divergence was observed on day 110, during the period of increased intensity and vertical wind activity, where two positive divergence peaks were observed.

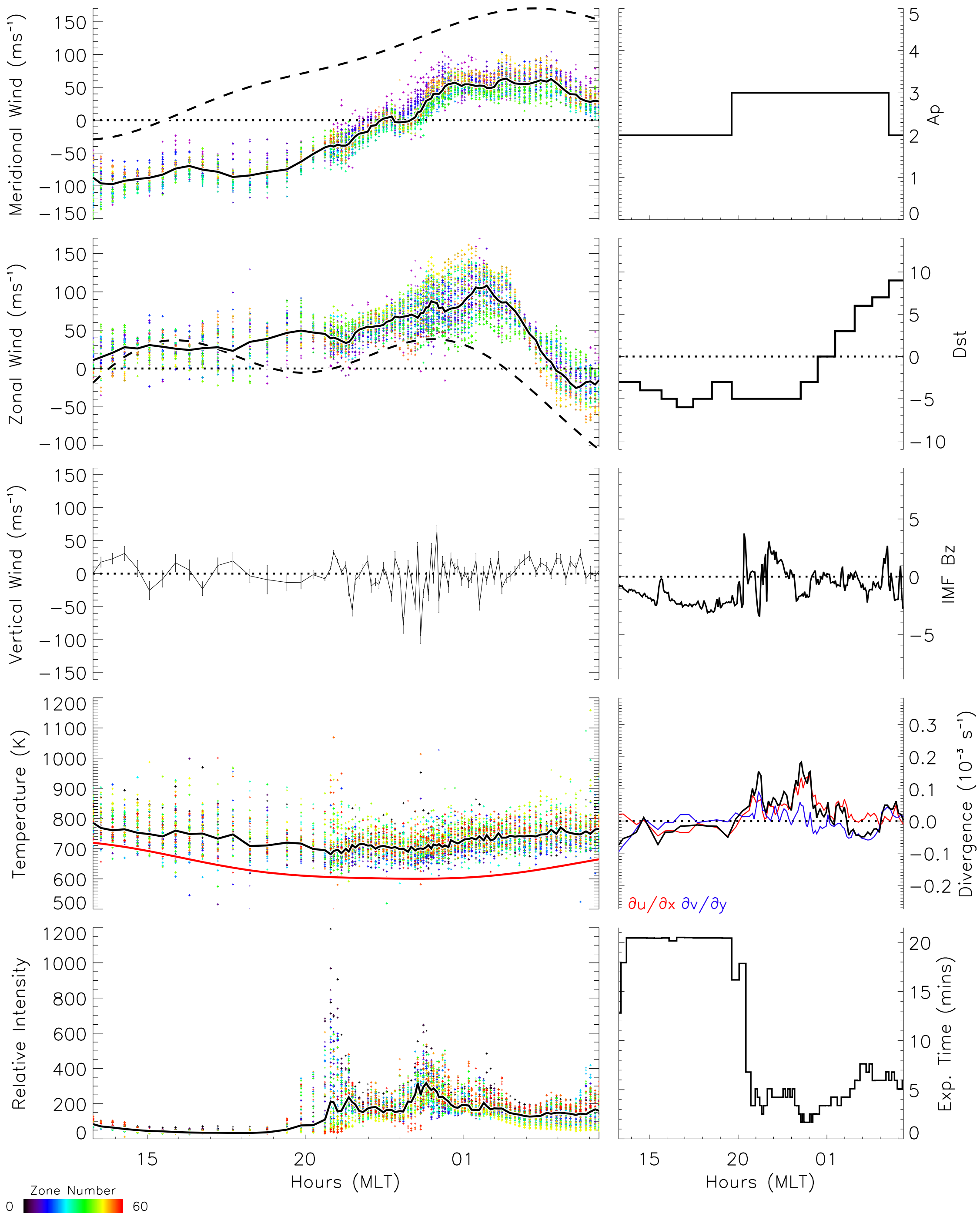


Figure 6.2.1: Results from the night of June 12, 2007 (day 163). The layout of this and the following three figures is described in the text.

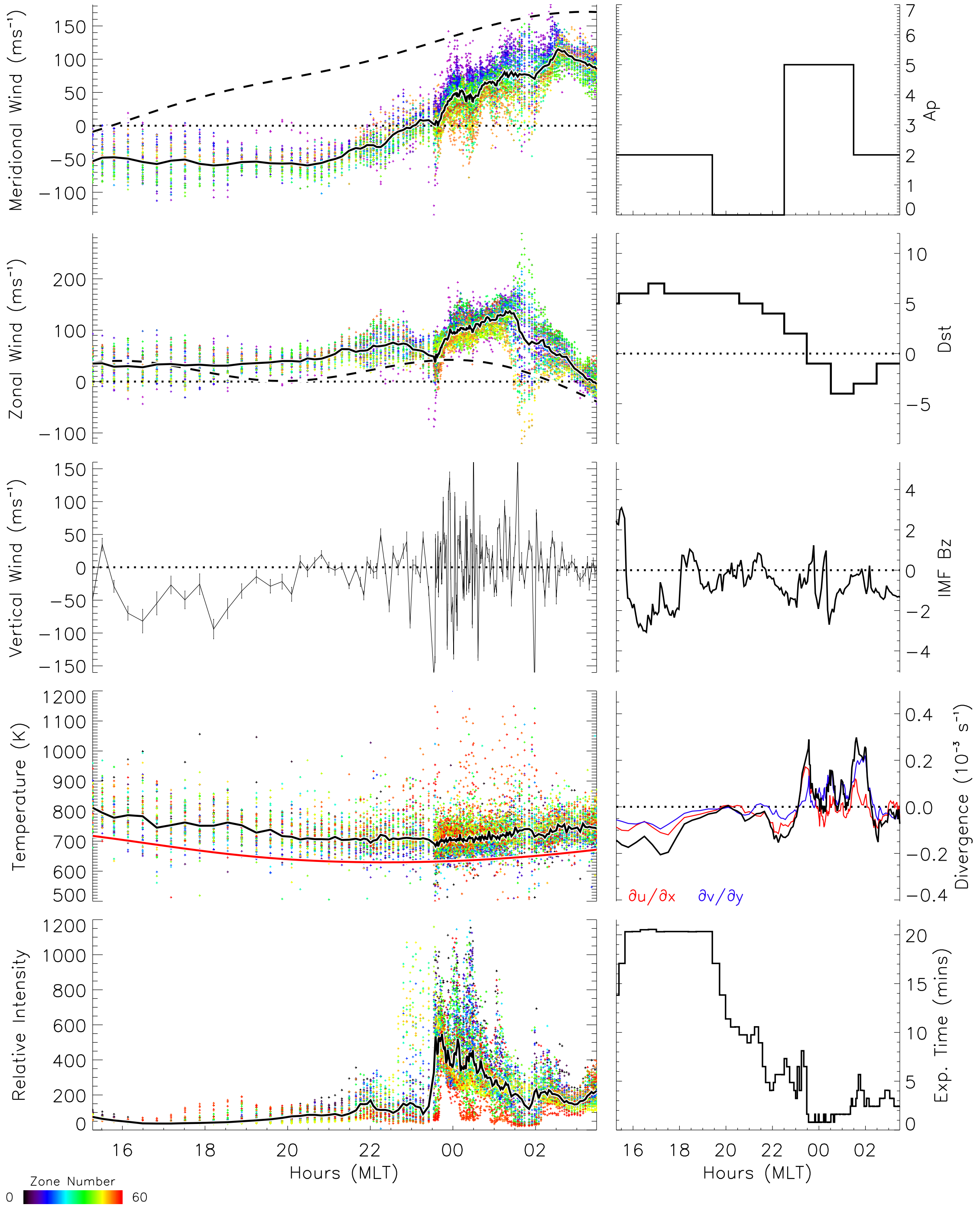


Figure 6.2.2: Results from the night of April 20, 2007 (day 110).

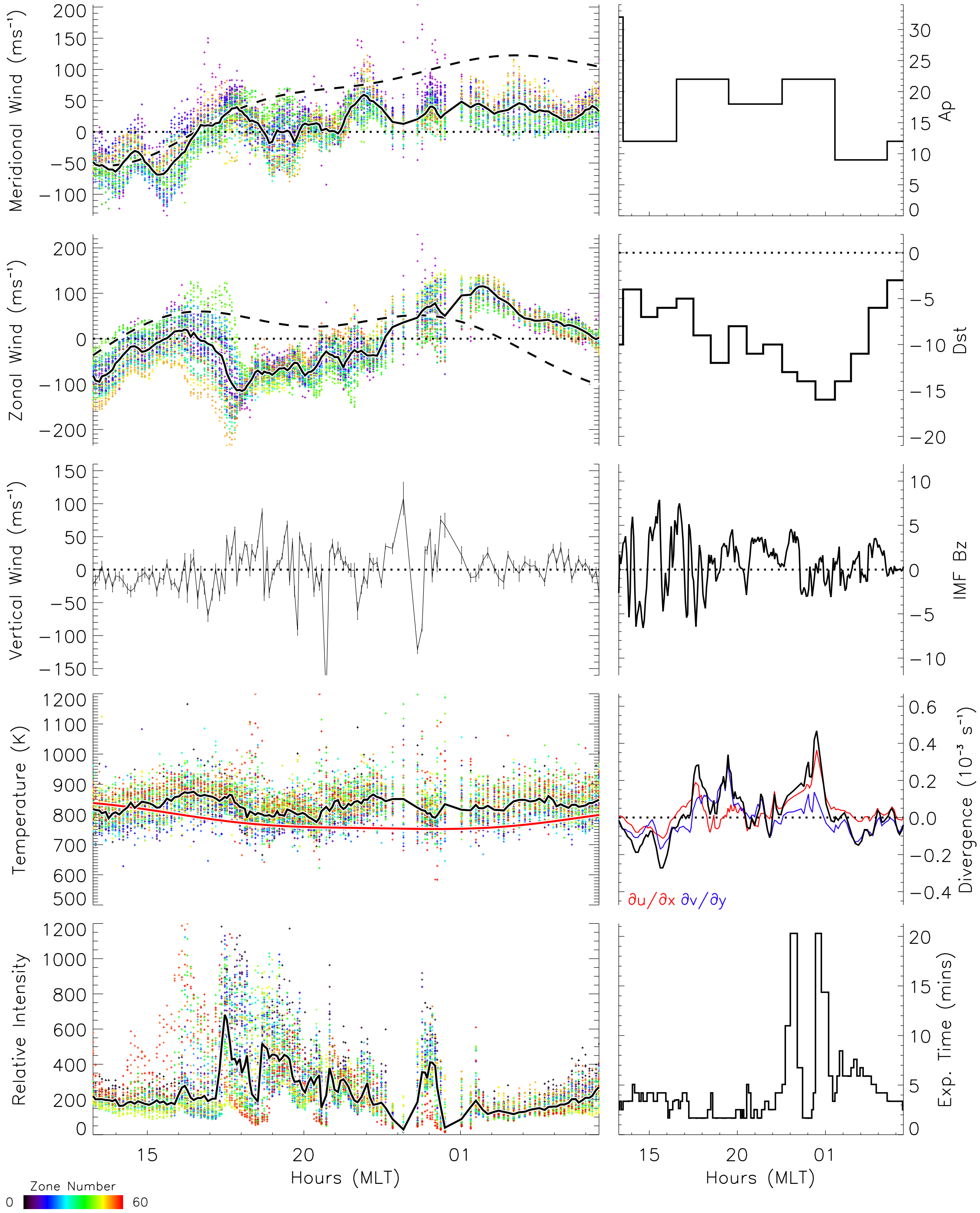


Figure 6.2.3: Results from the night of June 14, 2007 (day 165).

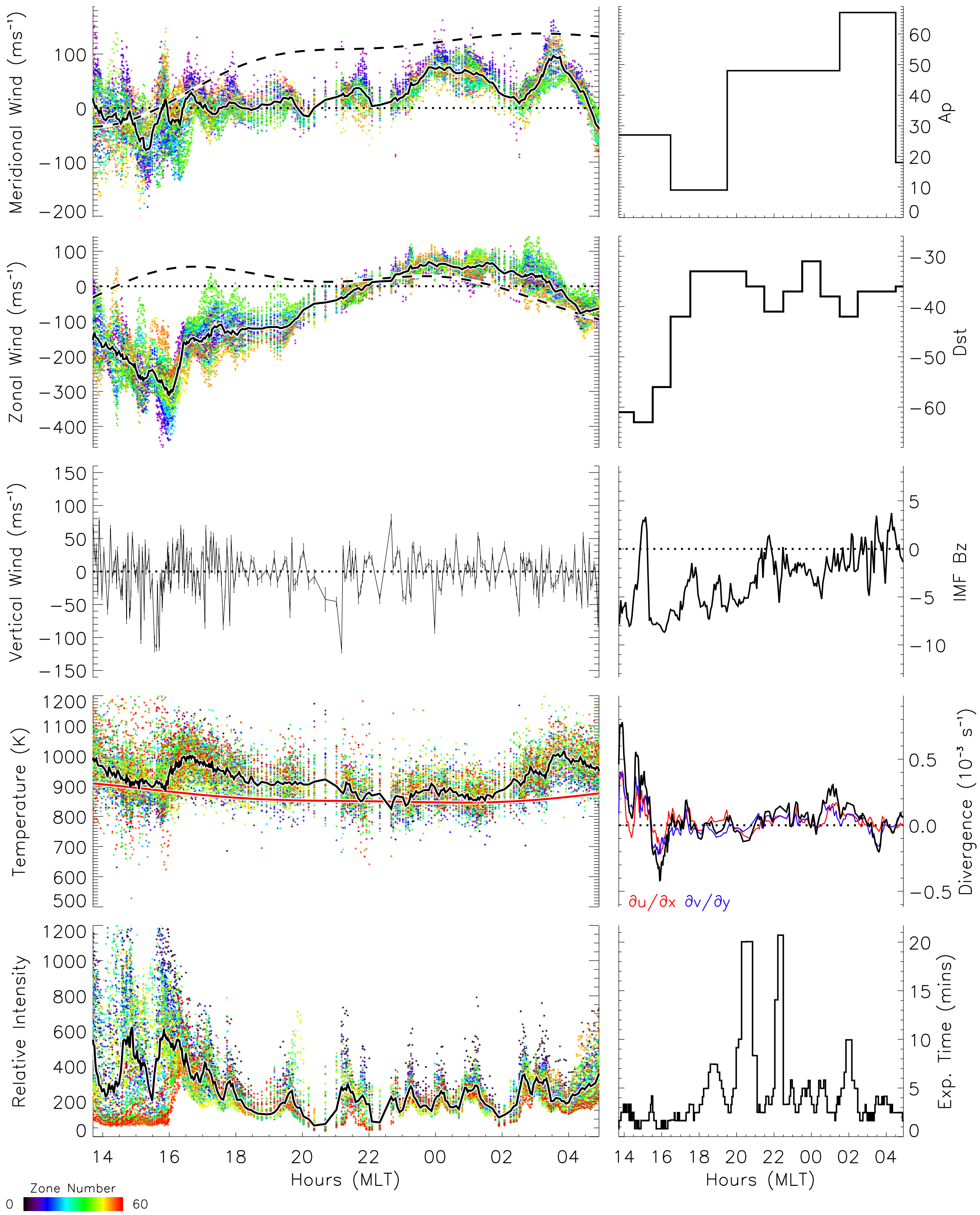


Figure 6.2.4: Results from the night of May 23, 2007 (day 143).

Temperatures on both nights showed only minor variability, with a generally smooth decrease from 800 K at the start of the night, to 680 K around magnetic midnight, followed by a smooth increase to 740-760 K at the end of the night. NRLMSISE-00 was able to reproduce the observed temperature trend, but consistently underestimated the magnitude. The mean difference between model and observed temperatures was around 110 K and 75 K for day numbers 163 and 110 respectively. It will be shown in Section 6.3.2 that on average NRLMSISE-00 was better able to predict observed temperatures under conditions of elevated geomagnetic activity (*ap* greater than approximately 35). NRLMSISE-00 is a large-scale empirical model, and thus relies on observations from point locations to generate model estimates at a given location, for a given level of magnetic and solar activity. Differences between modelled and measured temperatures could arise due to a lack of spatial coverage in the underlying model dataset near Mawson, and/or an underlying dataset that was compiled predominantly under conditions unlike those which prevailed for the measurements in the current work (i.e. an unusually deep, long-lasting solar minimum). In addition to this Mawson's proximity to the auroral oval, and the consequently highly variable heat sources, would make it difficult for a large scale model to accurately predict the intraday temperature variations.

6.2.2 Active-time Observations

Days 165 and 143 were periods of increased geomagnetic activity. In particular the second of these nights was very active, with *ap* reaching a maximum of 67, IMF B_z consistently southward between 15:00-20:00 MLT, and strongly negative Dst. Intensities on both nights were high, and quite variable, associated with bright aurora. Correspondingly, exposure times were generally short.

In contrast to the steady antisunward flow observed on the quiet days, meridional winds on these active days showed much more variability, both in magnitude and direction. Magnetic meridional winds on day 165 were southward at the beginning of the night, while on day 143 meridional winds early in the night showed a large amount of variability across all zones. The transition to magnetically northward flow was observed after approximately 21:00 MLT on both nights, around 2 hours earlier than was observed on the quiet nights. Meridional wind speeds after 18:00 MLT were again overestimated by the model, particularly at the end of the night, after around 01:00 MLT, when the observed (average) meridional wind decreased to less than ~ 30 m.s^{-1} when HWM93 predicted a continued magnetic northward flow of 100-150 m.s^{-1}. Superimposed on the average wind trends were fluctuations on the order of 50 m.s^{-1} and lasting around 30 minutes. Significantly more of these short time-scale fluctuations were observed on the active nights than were observed on the quiet nights.

Observed magnetic zonal winds were quite different between the two active nights. On day 165, prior to 16:00 MLT, observed zonal winds were in reasonable agreement with modelled winds returned by HWM93 (however they were offset from the model winds by approximately 50 m.s^{-1} westward), with early westward flow changing to eastward flow by

16:00 MLT. However, after this time, observed zonal winds again turned westward, with speeds reaching 100 m.s^{-1} after around two hours, and this westward flow continued until around 23:00 MLT. Throughout this period HWM93 predicted a minimal (< 50 m.s^{-1}) eastward flow. Around magnetic midnight HWM93 predicted the beginning of a transition to westward zonal flow, however this change was not observed until approximately 1-2 hours later. Zonal winds observed on day 143 were different again. Instead of the 0-50 m.s^{-1} eastward flow predicted by HWM93 prior to 22:00 MLT, very strong westward flow was observed, with average speeds reaching approximately 300 m.s^{-1}, and this flow continued with gradually decreasing magnitude until around 22:00 MLT, at which time a transition to eastward flow was observed. Similar to day 165, the predicted transition to westward flow around 01:00 MLT was not observed until approximately 03:00 MLT, two hours later.

Interestingly, while vertical wind activity on the two active nights was high, vertical wind speeds were generally less than were observed on day 110. However the periods of increased variability were concurrent with increased intensities, as was observed on day 110. This effect was clearly seen on day 165, where intensities before 17:00 MLT and after 01:00 MLT were quite stable (and relatively low), and the corresponding vertical winds were considerably less erratic. On both active days the total horizontal divergence was greater on average than on the quiet days, with a range of $[-0.4, 0.8] \times 10^{-3}$ s^{-1}. As on the quiet days, these periods of increased divergence were concurrent with periods of increased intensity and vertical wind activity. The link between vertical wind and horizontal divergence will be investigated in detail in Chapter 8.

Temperatures also showed more structure during these geomagnetically active days. Mean temperatures were higher by around 100 K (day 165) and 200 K (day 143) than the quiet day means of around 720 K. However mean differences between observed and model temperatures (predicted by NRLMSISE-00) were only around 50 K for both days, which was less than the corresponding differences for the quiet days. Temperature fluctuations of around 50-100 K were observed on both nights, the most obvious of these was observed on day 143, where a period of around 40 minutes saw temperatures increase by around 100 K. This heating was presumably caused by Joule and particle heating related to auroral activity.

6.2.3 Discussion

The effect of increasing magnetic activity on thermospheric winds can be more clearly seen in Figures 6.2.5 - 6.2.8. In these figures, fitted wind vectors from each zone are plotted in geomagnetic coordinates, with magnetic local time increasing clockwise around the dial. The view is that of an observer in space above the south geomagnetic pole (which is shown in the centre of the figure as a cross, with an arrow pointing toward the Sun) and looking down upon it. Thus geomagnetic south is everywhere directed toward the centre of the circle, and geomagnetic north is therefore radially outward. The eastward direction is clockwise, and west anti-clockwise. The arrow in the top left quadrant indicates the scale of the wind vectors. Exposures are displayed at a variable cadence (and at much less than the available time resolution) in order to fill the dial as much as possible while still retaining clarity. The all-sky image recorded by the spectrometer is shown in the background (in false colour), with the inferred horizontal wind fields overlaid. These wind fields have been (minimally) spatially smoothed. Printed radially inward of each exposure is the corresponding magnetic local time and exposure time. Plotted fields-of-view are not to scale, in that the diameters of the field-of-view circles are approximately 3 times smaller than they should be given their apparent distance from the geomagnetic pole.

As with Figures 6.2.1 - 6.2.4, these 'dial plots' are shown ordered (approximately) by the level of geomagnetic activity during the night. Quiet-time winds were mostly southward and eastward at the start of the night. Prior to magnetic midnight there was a change to a predominantly magnetic eastward (antisunward) flow as the meridional wind decreased, followed by a transition to magnetic northward flow after magnetic midnight. With increasing geomagnetic activity, a rotation of the afternoon sector winds to a more magnetic westward flow was observed, along with increased wind speeds. This rotation is consistent with the direction in which ionospheric plasma is forced to move by $\mathbf{E} \times \mathbf{B}$ drift, in response to the magnetospheric electric field (Sections 1.2.1 and 2.4.1).

Figure 6.2.9 shows the direction of F-region ionospheric plasma flow assuming a vertical magnetic field (directed out of the page) and an electric potential pattern shown by the grey contour lines (from the Weimer96 model, for IMF southward conditions, for which the potential contours take on the relatively simple 'twin-cell' convection pattern). Over the polar cap, the plasma flow indicated in this figure is reinforced by the pressure gradient established by solar heating on the dayside. It is therefore not surprising that the magnetic equatorward flow which is seen in Figure 6.2.9 around magnetic midnight was clearly observed on all four nights presented here, since in this time sector pressure-gradient forcing and ion-drag forcing would accelerate the neutral wind in the same direction.

Only on the two more active nights (day numbers 165 and 143) did the wind direction in the 13:00-21:00 MLT sector resemble the sunward flow depicted in Figure 6.2.9. On these more active nights there was presumably sufficient momentum transfer from the convecting ions to reverse the antisunward flow. On day 165 (Figure 6.2.7) a shear in the wind flow was observed in the 16:16 and 17:18 MLT frames, associated with the auroral oval moving (magnetically) equatorward overhead of Mawson. After the passage of the auroral oval the flow became magnetic westward, consistent with the sunward ion convection, while prior to that time the wind flow had been directed approximately poleward (antisunward). Strong latitudinal shears and flow reversals have been observed previously by both ground (Conde and Smith, 1998; Conde et al., 2001) and satellite (e.g. Rees et al., 1983a; Hays et al., 1984; Killeen and Roble, 1988) observations, usually collocated with the duskside optical auroral oval. At Mawson, shears of this type were observed almost exclusively in the presence of bright auroral arcs, and will be examined further in Chapter 8.

Positive correlation has been observed between geomagnetic activity and solar wind parameters such as IMF magnitude and solar wind speed (e.g. Wilcox et al., 1967). Thus increased geomagnetic activity will typically be associated with a higher cross-polar cap

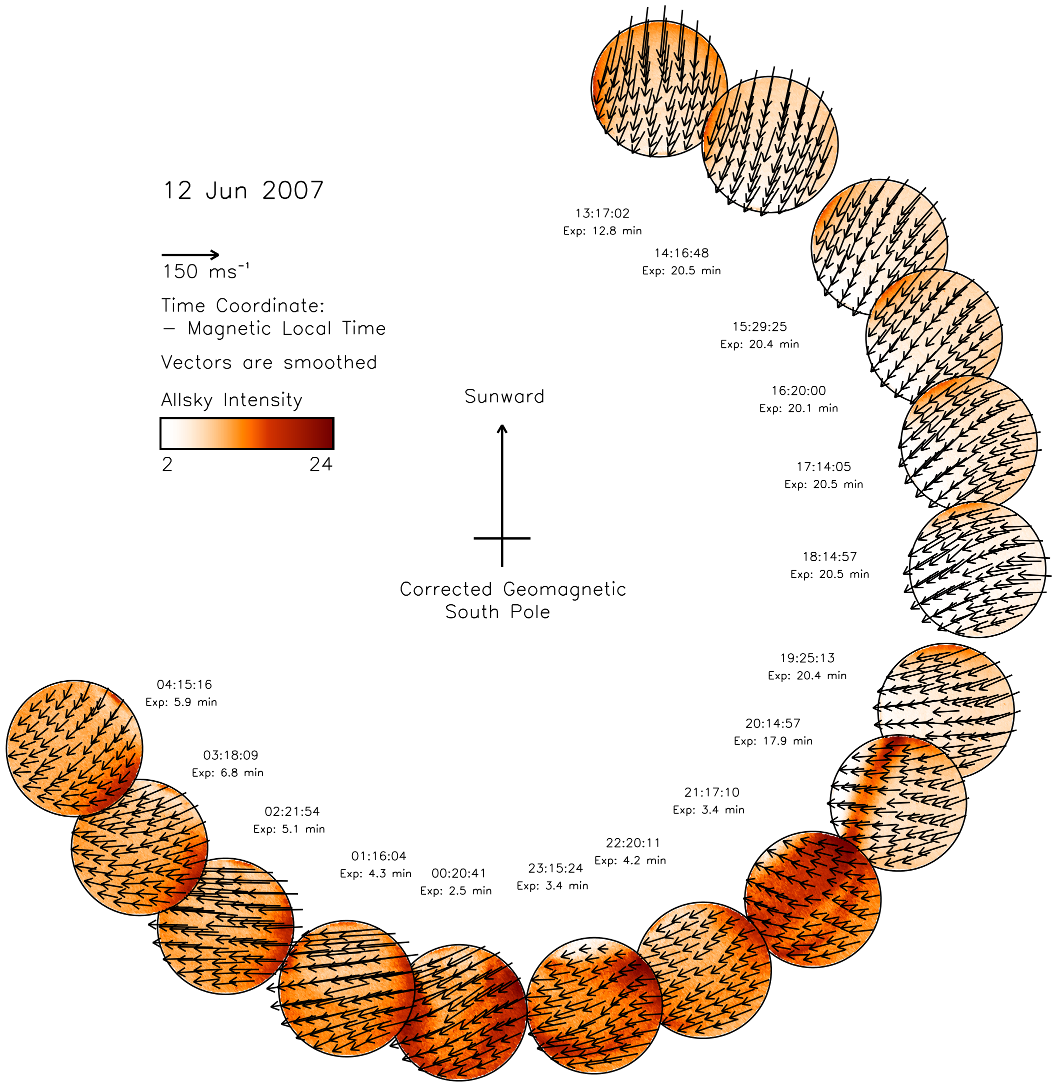


Figure 6.2.5: Dial-plot showing the fitted wind field overlaid onto the (false colour) all-sky images, in geomagnetic coordinates, for June 12, 2007 (day 163). Plot format is described in the text.

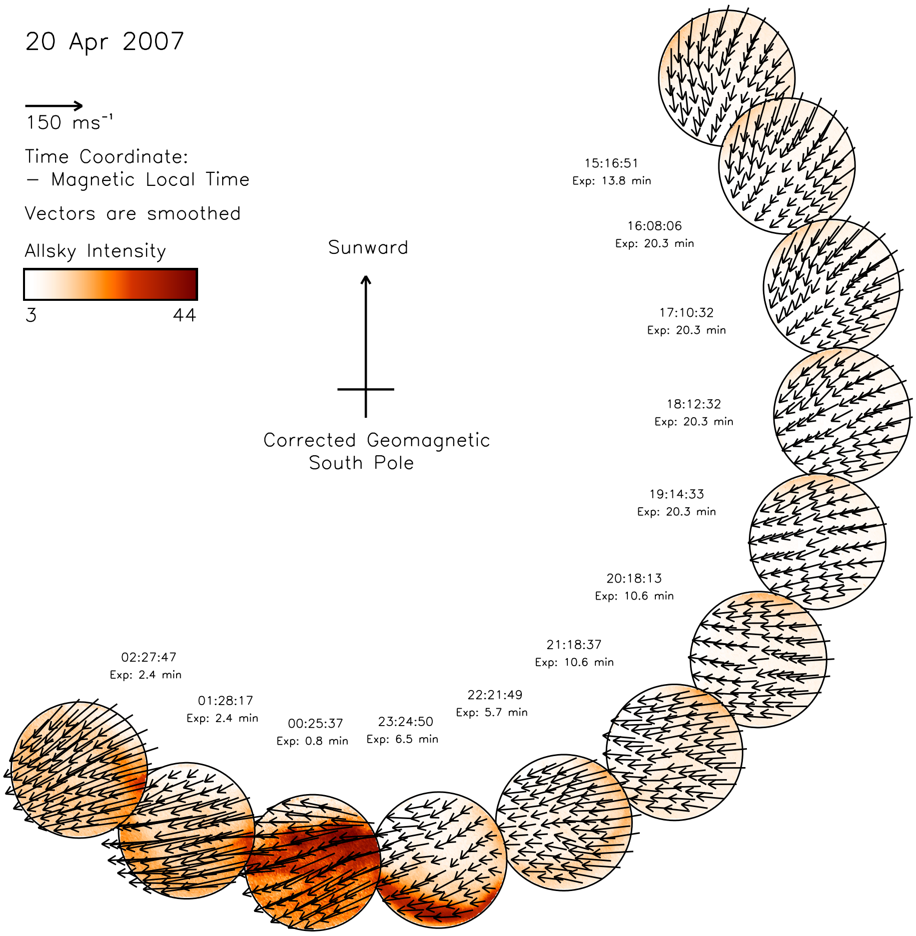


Figure 6.2.6: Same format as Figure 6.2.5, except for April 20, 2007 (day 110).

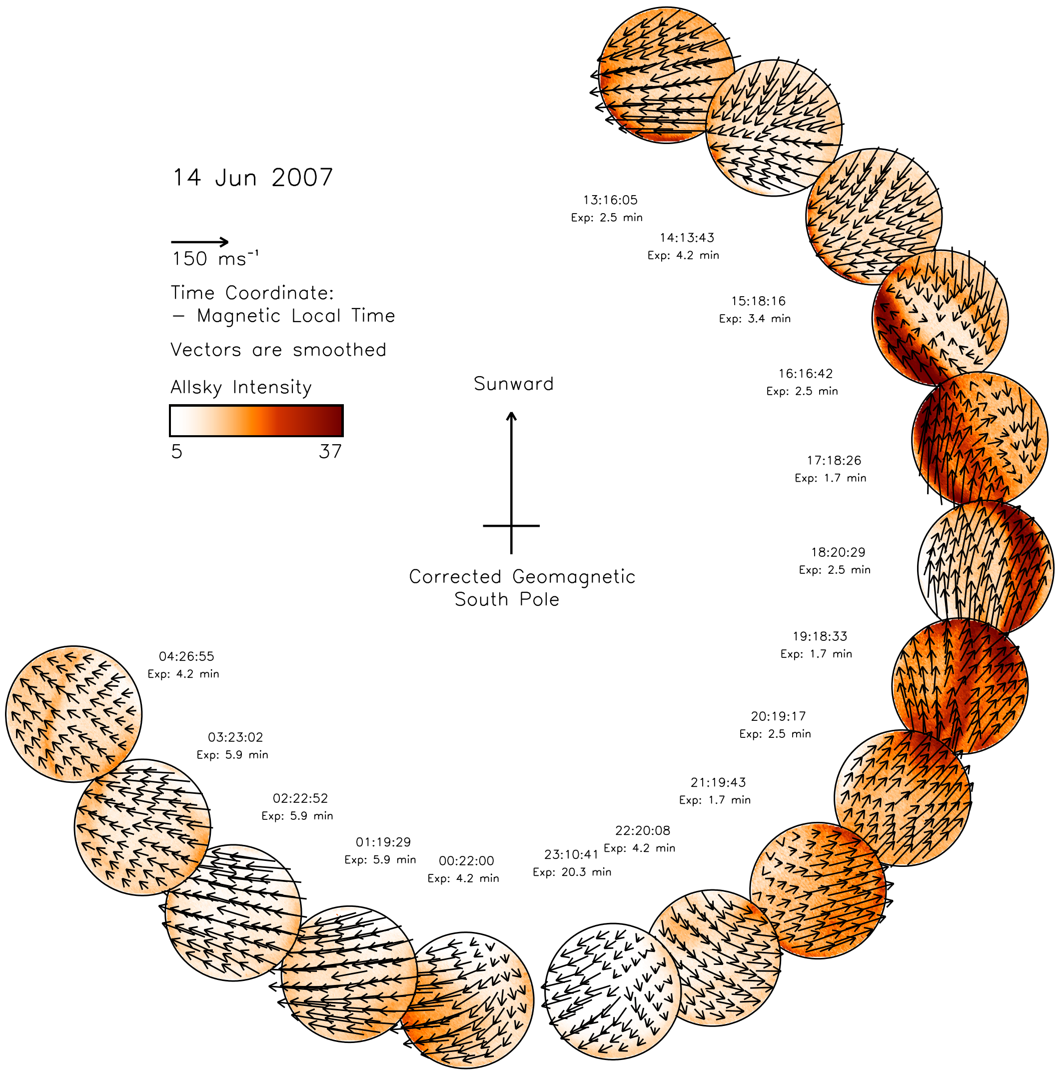


Figure 6.2.7: Same format as Figure 6.2.5, except for June 14, 2007 (day 165).

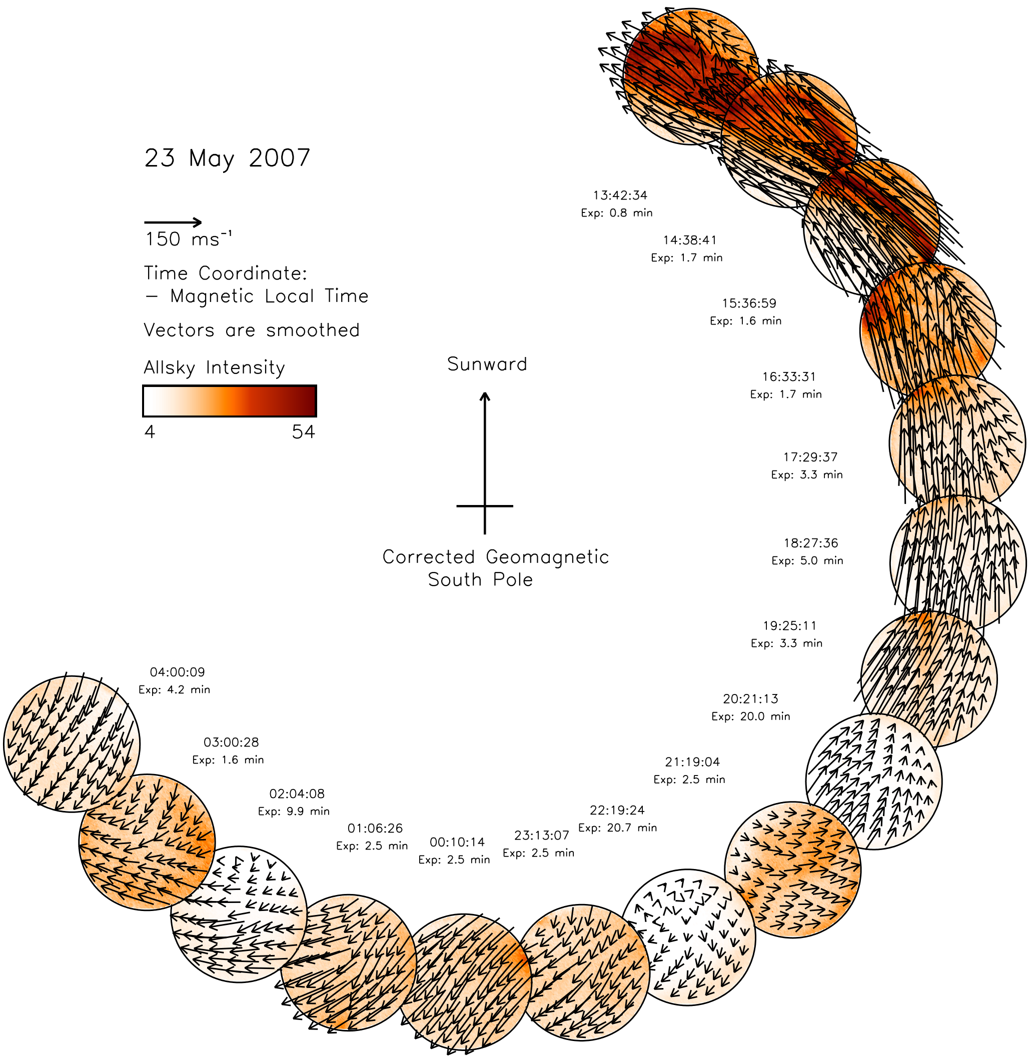


Figure 6.2.8: Same format as Figure 6.2.5, except for May 23, 2007 (day 143).

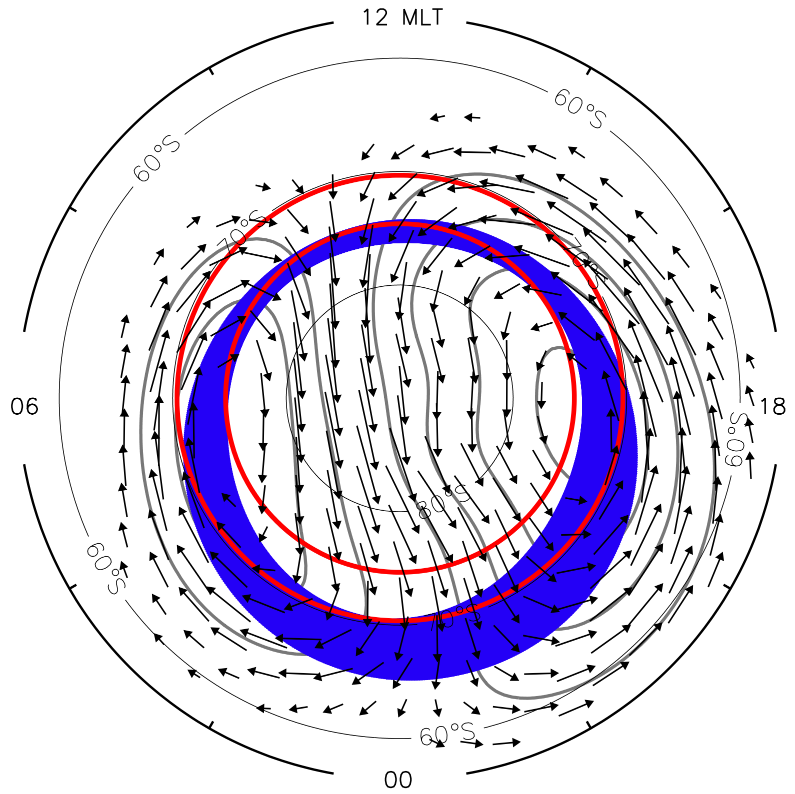


Figure 6.2.9: F-region ionospheric plasma drift in the $\mathbf{E} \times \mathbf{B}$ direction (arrows), assuming a magnetic field directed out of the page ($\mathbf{B} = B_0\hat{\mathbf{z}}$) and an electric potential shown by the grey contour lines (from the Weimer96 model). The view is the same as for the dial plots (Figures 6.2.5 - 6.2.8), with corrected geomagnetic latitude marked. The latitude of Mawson and Davis stations are shown by the outer and inner red circles respectively. A nominal auroral oval, corresponding to moderately active geomagnetic conditions (Q=2), is shown in blue (Holzworth and Meng, 1975).

potential (Equation 1.2.3), leading to faster plasma convection and therefore increased ion-drag forcing. In addition to faster plasma flows, particle precipitation can increase the local plasma density by ionizing neutral species. Higher plasma density implies larger ion-neutral collision frequencies, and therefore greater ion-neutral coupling. It is presumably this increased coupling between convecting ions and the neutral atmosphere which results in the observed sunward evening sector winds during active conditions. Of course convecting ions can transfer heat as well as momentum to the neutral atmosphere (through Joule heating, Section 2.3.2). Particle precipitation similarly heats the neutral gas, albeit less efficiently than Joule heating (Section 2.3.3). Thus the signature of enhanced ion-neutral coupling is also seen in the increased temperatures observed on these two active nights, which were 100-200 K higher on average than the temperatures observed on the quiet nights.

The winds presented here are in good agreement with results presented by other authors. Conde and Dyson (1995a) showed two examples of thermospheric wind data above Mawson recorded during 1992. These examples demonstrated the extremes of the wind behaviour which was typically observed. The first of these nights showed antisunward wind flow throughout the night, similar to the two quiet nights presented in Figures 6.2.5-6.2.6, however the direction of the winds presented by Conde and Dyson were more properly antisunward than those in the present work. The other extreme presented by these authors showed winds directed strongly magnetic westward (sunward) in the early magnetic afternoon, which rotated to become antisunward around 21:00 MLT. Again this is entirely consistent with the active period winds, particularly as shown in Figure 6.2.8. For both examples presented by Conde and Dyson the wind magnitudes were higher (in the active period they were considerably so, with magnitudes approaching 400 m.s^{-1}), than the corresponding winds observed in the present work, and this is presumably attributable to the more active solar conditions during which these earlier data were recorded.

The vertical wind response to enhanced emission intensity was apparent in all four of the nights presented in the current work. Wardill and Jacka (1986) presented similar results of vertical winds above Mawson which showed a rapid response to local geomagnetic field perturbations. Strong (50-100 m.s^{-1}) upward vertical winds were observed in the evening sector auroral oval, associated with strong negative perturbations in the H component of the local magnetic field. Vertical wind oscillations of > 50 m.s^{-1} were observed on one night which persisted for several hours following the main geomagnetic disturbance. Sica et al. (1986a) reported large vertical wind variations above College, Alaska, during periods of auroral activity, particularly during pulsating aurora. These authors suggested that changes in the magnetospheric-ionospheric current system, changes in the energy and flux of precipitating particles, or thermospheric gravity waves propagating upward from lower-altitude heating sources were all possible sources of these vertical wind variations.

6.3 Average Behaviour

In addition to the individual nights presented in the previous section, the average daily behaviour of winds and temperatures was investigated by combining data from many nights of observation. For this analysis, the all-sky median values described in Section 6.2 were calculated for every available exposure satisfying the condition that the cloud level[1] was $\leq$ 3/8. These exposures were sorted by geomagnetic and solar activity, as parameterised by the ap and $F_{10.7}$ indices respectively. Selection criteria were then applied to these exposures, which corresponded to ranges of ap, $F_{10.7}$, day number and year (specific selection criteria are described where appropriate). Exposures which satisfied these criteria were then binned into time-bins of 30 minutes width, and the median value of each parameter (wind, temperature, intensity, etc.) within each time bin was calculated. Model values were calculated for every exposure in the data set, and averaged in the same way as observed winds and temperatures. Each time bin then contained, along with the average of the observed wind components and temperatures, the average of the returned model estimates corresponding to each of those parameters, for the levels of geomagnetic and solar activity prevailing at the time those observations were made.

[1]Meteorological observers traditionally report the fractional cloud cover over the sky in increments of eighths.

6.3.1 Wind

Figures 6.3.1 and 6.3.2 show the winds averaged in this way above Mawson and Davis during 2007 and 2008. In this case the only selection criterion was the year, so that only observations from the given year contributed to the averages. Vectors are plotted in the dial-plot format described in Section 6.2.3. Corrected geomagnetic coordinates are used, so that the centre of each plot represents the south geomagnetic pole, with south directed everywhere radially inward and eastward pointing clockwise around the dial. Magnetic local times are indicated inside the circle. In these plots wind vectors have been calculated using observations from four geomagnetic look directions (north, south, east and west). At Davis these correspond to interpolates from the routine geographic look directions, whereas at Mawson they were calculated from zones corresponding to these directions, and averaged over all rings containing more than 4 zones. The wind estimates from these look directions were then combined (north with east, north with west, south with east, south with west) to produce the vectors shown. Northward-looking estimates are plotted in blue, southward-looking estimates in red (note that the figure is not drawn to scale in that the locations of the vectors do not correspond to their true projected locations). Between the Mawson and Davis latitude circles are shown the vector winds predicted by HWM93 for the position and conditions above Mawson (averaged as described above). Since the model estimates for the two stations were very similar, this suffices to demonstrate the model estimates for both locations.

For data recorded during 2007, 84% of Mawson observations were recorded when ap was less than 20, while at Davis the percentage was 92%. During 2008, these values were 93% and 95% for Mawson and Davis respectively. Thus these averages were heavily biased toward a geomagnetically quiet day. There is also a seasonal bias due to the changing day length throughout the year. Relatively fewer observations contribute to the averages in the early magnetic afternoon (near 12 MLT) and early morning (near 6 MLT), and these observations will be biased toward midwinter, when observations are possible at those times. The number of data points contributing to each time sector will be approximately symmetrically distributed around a maximum at $\sim$ 20:30 MLT (local solar midnight).

The wind averages in Figures 6.3.1 and 6.3.2 reflected the low average level of geomagnetic activity, exhibiting wind flow consistent with a predominantly solar pressure-gradient driven wind. That is, the wind was directed magnetically southward in the early magnetic evening, toward the polar cap and presumably feeding into the cross-polar jet (the dayside-to-nightside cross-polar cap wind, see Section 2.4.1). This is consistent with the thermosphere's in-situ forced solar diurnal tide. Wind speeds were observed to decrease until around magnetic midnight, at which time both stations observed the nightside emergence of the cross-polar jet as a predominantly magnetic northward (equatorward) flow.

No strong convection signatures were present in the magnetic afternoon or evening winds above either station, although wind directions were not completely antisunward at that time either. Mawson, which was usually northward (equatorward) of or underneath the auroral oval, observed a reduction and almost complete disappearance of the wind during both years between about 19:00 and 22:00 MLT, while Davis, which was often southward (poleward) of the auroral oval at that time, saw a reduction in wind speed to less than 50 m.s^{-1}, but not a complete abatement. While a reduction in wind speed was often observed in this time sector on individual nights, the low average wind speeds may also reflect a balance (on average) between days which exhibited sunward (i.e. ion-drag forced) flow in this sector and days which showed antisunward (pressure-gradient driven) flow. The high wind velocities observed on days with sunward flow may have been enough to offset the relatively small number of these days contributing such flow to the average.

Around magnetic midnight both stations during both years saw wind vectors directed magnetic northward with magnitudes of 100-200 m.s^{-1}, where pressure-gradient and ion-drag forcing would be expected to reinforce each other and drive the wind toward the equator. At this time observed average winds were in good agreement with model wind vectors. The maximum magnetic northward flow occurred around 02:00 MLT. Also observed during this period of northward flow at both stations was a tendency for the magnetic northward-looking wind speeds to be greater than those looking magnetic southward, resulting in a magnetic meridional divergence. This divergence was larger in 2007 than 2008, and larger above Davis than above Mawson. There was also a period of magnetic zonal divergence evident above Davis, beginning around 21:30 MLT and continuing until around 02:00 MLT. This magnetic zonal divergence was much larger in 2007 than in 2008, and was really only evident above Davis.

Greet et al. (1999) compared results from narrow-field Fabry-Perot spectrometers operating at Mawson and Davis. Five nights of simultaneous observations from these two instruments were examined. While these authors presented only single-night observations, the wind vectors they derived were in general very similar to those shown here, with magnetic southward flow in the magnetic afternoon sector, gradually decreasing to a period of minimal wind speeds between 18:00 MLT and magnetic midnight (a period that Conde and Smith (1998) have termed the "doldrums"). After magnetic midnight the wind turned magnetic northward. For both the most quiet and most active day they presented (with average daily *ap* of 3 and 12 respectively), Greet et al. observed faster (magnetic) midnight sector wind speeds when observing magnetically northward of Davis station compared to observations made looking southward. This trend can be seen clearly above Davis in both the 2007 and 2008 dial plots presented here. It was also observed to a lesser degree above Mawson, although Greet et al. observed this only on their quietest day. Also on this quietest day, Greet et al. observed larger magnitude winds above Davis than those above Mawson, which is consistent with the present work. However during their most active day (which in this study would be considered quiet), observed speeds were comparable after magnetic midnight, and before magnetic midnight larger speeds were observed above Mawson than above Davis.

Average winds from both quiet (K-sum $\leq$ 27) and active (K-sum $>$ 27) geomagnetic conditions[2] above Mawson were presented by Conde and Dyson (1995a), from data

[2]K-sum is the sum of the local K indices (as opposed to the planetary K indices, Kp) for the preceding 24 hours, 12 UT-12 UT.

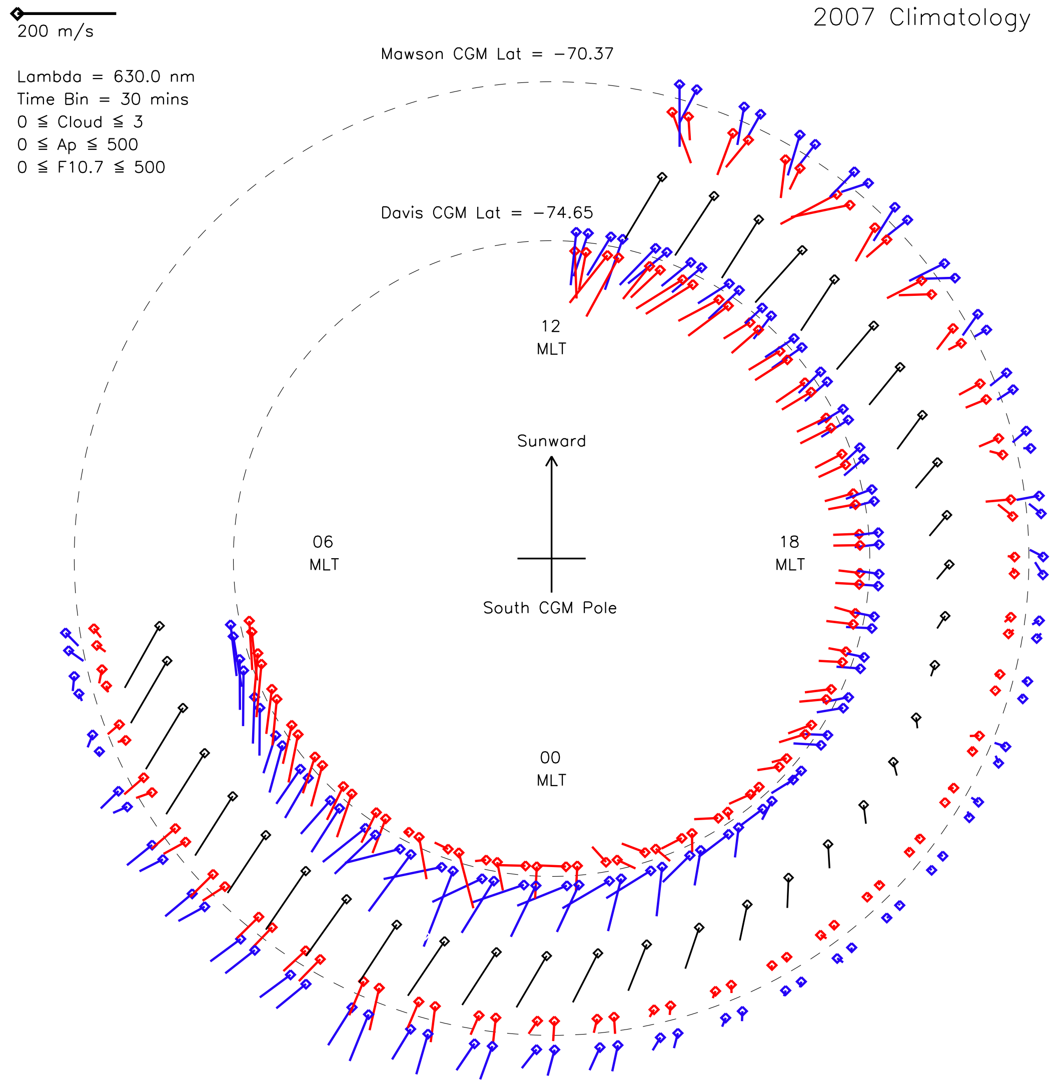


Figure 6.3.1: Average winds above Mawson and Davis during 2007. Northward-looking vectors are shown in blue, southward-looking vectors in red (the base locations of the vectors are not to scale). HWM93 wind vectors are plotted in black, and correspond to the location and conditions prevailing at Mawson. The wind speed scale is shown in the top left corner, along with selection criteria. Vectors are plotted in corrected geomagnetic coordinates, with the circle symbol plotted at the base of each vector.

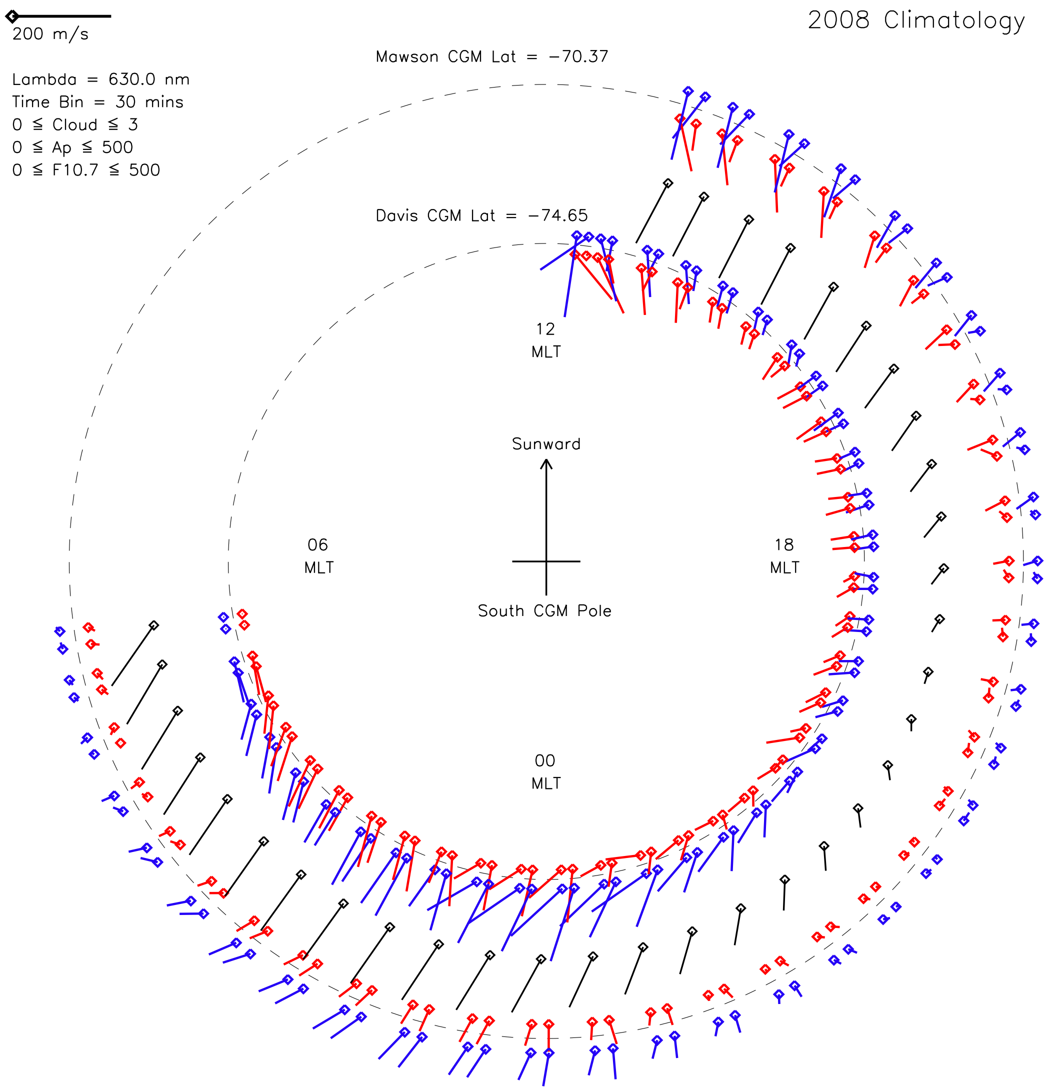


Figure 6.3.2: Average winds above Mawson and Davis during 2008. The plot format is the same as Figure 6.3.1.

recorded during the Austral winter of 1992. The quiet-time winds these authors observed around midwinter showed very similar characteristics to the winds presented in the current work. Winds were generally antisunward for the entire average night, and directed approximately along the 14:00-02:00 MLT line, as was observed in the present work. Whereas the average northward wind vectors presented by Conde and Dyson did not show a complete reduction in speed in the 19:00-22:00 MLT sector, the southward looking winds did show such a reduction, to very small wind speeds, in general agreement with the present work.

Quiet-time average winds were also reported by Smith et al. (1998), who averaged Fabry-Perot data from 8 days at Mawson and 9 days at South Pole which exhibited quiet to moderate magnetic activity. South Pole station is at a similar magnetic latitude (-74.18°) to Davis. These data were recorded during 1992, when solar activity was still high (during the downward phase of solar cycle 22), and hence at a time when plasma densities should be increased, thereby increasing the efficiency of ion-neutral momentum coupling and the ability of the high-latitude ionospheric convection pattern to drive a similar neutral circulation. Smith et al. observed the cross-polar jet to emerge with a direction towards 02:00 MLT, consistent with the magnetic northward flow seen in Figures 6.3.1 and 6.3.2 around this time. They also observed a period of doldrums in the northward-looking Mawson data, however this period of wind abatement did not last as long as in the present study. In contrast to the data presented here, these authors observed slower wind speeds magnetically northward of the station compared with those to the south, at both Mawson and South Pole, in the magnetic midnight/magnetic morning sectors.

Average winds above Mawson were calculated by Conde (1990), and broadly similar features were observed. For example, faster flows were observed near magnetic noon (12:00-15:00 MLT) and magnetic midnight. Also the period of slower winds around 20:00 MLT was observed, however this period of 'doldrums' was again confined to a narrower time interval than in the present study (approximately 2 hours compared to 5 hours observed here). Conde (1990) ascribed this wind abatement to the station passing beneath a region where ion-drag effectively cancelled the pressure-gradient forcing.

Wardill et al. (1987) averaged 41 nights of wind observations made above Mawson during the 1983 winter (using the same instrument as Greet et al. (1999)) which corresponded to quiet-moderate magnetic conditions ($ap < 25$). These authors observed winds which closely followed the pattern expected from a predominantly convection-driven flow, similar to the active-time winds presented in Section 6.2.2 (and particularly Figure 6.2.8). Winds were directed south-westward in the magnetic afternoon and evening and magnetic northward around magnetic midnight, with a period of low wind speeds 20:00-22:00 MLT. From this pattern they concluded that ion-drag forcing was the principal momentum source for the high-latitude thermosphere. The Sun at the time of this study was passing through the downward phase of solar cycle 21, thus solar conditions were similar to those of the Greet et al. study, and similar conclusions apply. The direction of flow in the early magnetic afternoon and in the post magnetic midnight sector observed in the current work is also entirely consistent with the results of Thayer and Killeen (1993), who observed that the irrotational (divergent) component of the wind field was representative of the solar-driven component, directed primarily along the 14:00-02:00 MLT plane.

Dependence on Geomagnetic Activity

From the discussion of Section 6.2.3 it is evident that the winds and temperatures observed above Mawson were rather strongly dependent upon the prevailing level of geomagnetic activity. It is therefore useful to separate the data by a suitable indicator of magnetic activity (in this case the *ap* index) and to examine the average behaviour under different levels of activity. Since the data set was biased toward very quiet conditions (Section 6.3), it was not possible to average over many narrow *ap* bins. Thus the data have been split into two categories of magnetic activity: $ap \leq 10$ and $ap > 10$. These categories will be referred to as 'quiet' and 'active' conditions respectively, although it should be kept in mind that while the 'quiet' category is indeed representative of very magnetically quiet conditions, the 'active' category can at best be described as only moderately so when compared to the full range of activity typically observed over a normal solar cycle. Figure 6.3.3 shows the result of averaging both the Mawson (2007-2008) and Davis (2004-2008) data in this way. Blue curves show 'quiet time' average behaviour, for which $ap \leq 10$, while red curves show 'active time' behaviour, for which $ap > 10$. The average geomagnetic meridional and zonal components of the wind are plotted in the top two panels (along with HWM93 winds shown as dashed lines), beneath which are shown the average vertical wind and the number of observations contributing to each time bin. The time axis shows magnetic local time.

Even under this modest separation by activity level, differences in average wind behaviour above both stations were evident. At Davis the nightly variation of both the meridional and zonal components of the wind appeared to be shifted toward earlier magnetic local times. For example the transition from magnetic southward to magnetic northward flow was observed at approximately 21:00 MLT during quiet conditions, whereas during active conditions this transition was observed to occur approximately 2 hours earlier, at 19:00 MLT. The peak northward wind was also observed approximately 2 hours earlier under active conditions. Similarly, the peak eastward zonal wind was observed approximately 2 hours earlier on average when conditions were active, as was the transition from a magnetic eastward to a magnetic westward wind after 22:00 MLT. There was also significant contrast between zonal flow prior to 17:00 MLT. Under active conditions the zonal flow was around 100 m.s^{-1} westward just after magnetic noon, and decreased to approximately 0 m.s^{-1} over the next 5 hours. Under quiet conditions zonal flow was all but absent during this time period.

At Mawson the peak northward meridional wind was also observed earlier in the night during active conditions, however the general shift toward earlier magnetic local times which was observed above Davis was not apparent in the Mawson data. Before 20:00 MLT, meridional winds under both levels of activity were very similar at Mawson, while after this time the active-time meridional wind became northward earlier in the night and reached a larger peak magnitude than was observed during quiet times. For the zonal component, the reverse was true: before approximately 20:00 MLT, magnetic westward

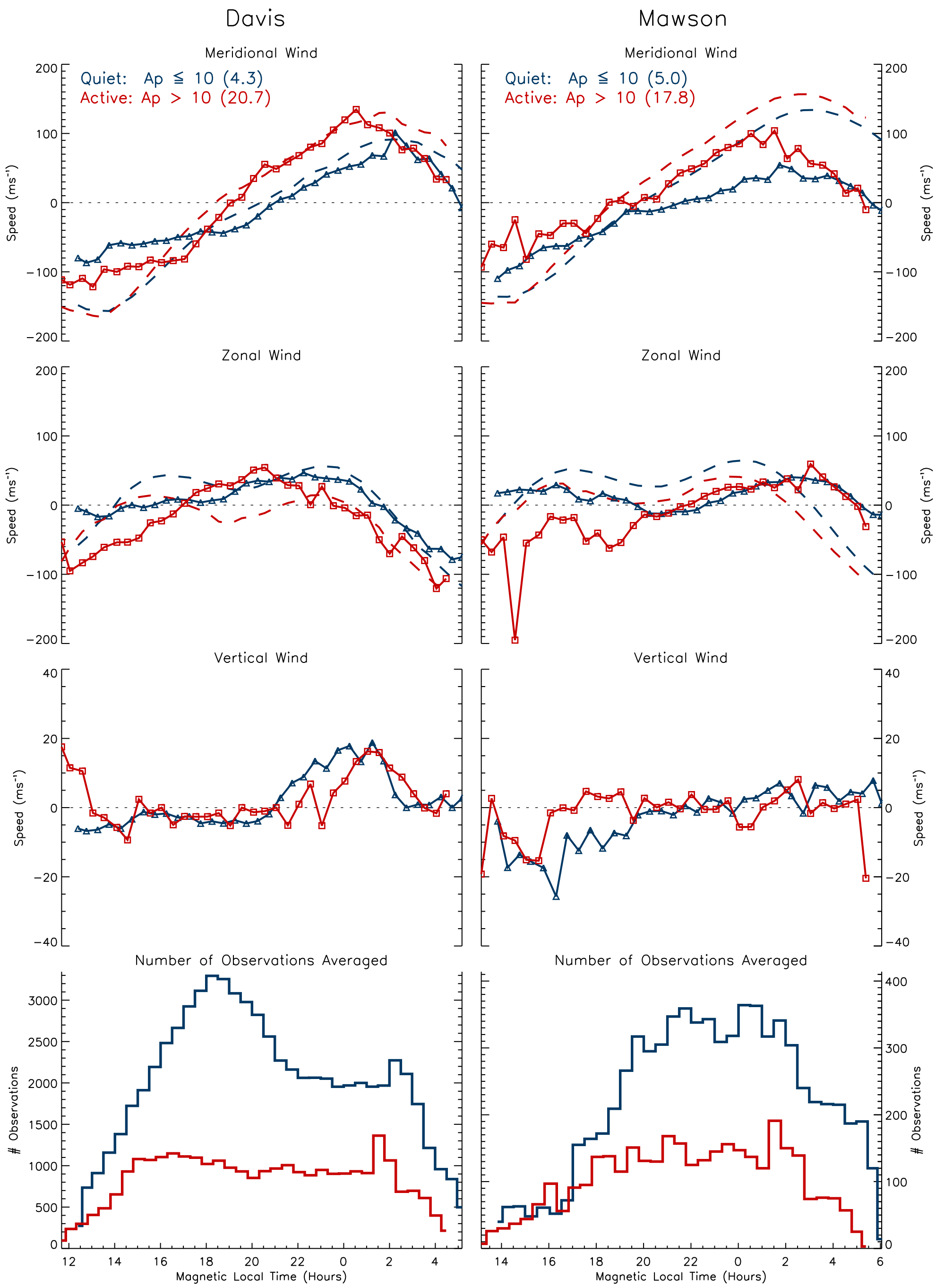


Figure 6.3.3: Average winds above Davis (left) and Mawson (right) for quiet and active geomagnetic conditions. Davis results are averaged over 5 years of data (2004-2008), Mawson results are averaged over 2 years (2007-2008). Averages under quiet conditions (corresponding to $ap \leq 10$) are shown in blue, averages obtained during active conditions ($ap > 10$) are shown in red. Mean ap for the given level of activity is indicated in parenthesis after the ap range. Winds predicted by HWM93 are shown as dashed lines. Meridional and zonal components are in geomagnetic coordinates.

flow on the order of 50 $m.s^{-1}$ was observed under active conditions, while during quiet times a weak ($\simeq$ 20 $m.s^{-1}$) eastward flow was observed at this time. After 20:00 MLT, both active and quiet-time winds were almost identical.

In general, the nightly trend of the average meridional wind was very similar between the two stations, under both quiet and active conditions. Southward (poleward) flow in the early magnetic afternoon, changing to magnetic northward (equatorward) flow sometime before magnetic midnight, peaking around 00:00-02:00 MLT, before decreasing to zero by around 05:00 MLT at Davis and 06:00 MLT at Mawson. Identifying the region of magnetic equatorward flow around magnetic midnight with the emergence of the cross-polar jet on the nightside, the flow region was observed to be broader (cover a larger range of magnetic local times) on average under active conditions, with the dusk-side edge of the region being encountered earlier under such conditions.

Interestingly, at each station, the dawn-side edge of the flow region was encountered at approximately the same magnetic local time under both levels of activity (∼05:00 MLT at Davis and ∼06:00 MLT at Mawson). The delay between Davis observing the edge of the flow region and Mawson observing the edge is consistent with the geographical separation between the two stations, which is equivalent to approximately 40 minutes of magnetic local time. It would thus appear that while the dusk-side edge of the antisunward flow region expanded westward under active conditions, the dawn-side edge remained relatively stationary (in a Sun-fixed coordinate frame). This might imply that the centre of the region shifted westward, along with a general broadening of the region (in magnetic local time), to account for the lack of a shift in the dawn-side edge. A broadening of the antisunward flow region would be consistent with an increase in the size of the high-latitude convection pattern under increased levels of geomagnetic activity. In terms of Figure 6.2.9, this can be pictured by placing Mawson at a geomagnetic latitude higher than 70°. The width of the antisunward flow region would then be seen to cover a larger range of magnetic local times, effectively broadening the region.

As was observed in Section 6.2, average active-period winds in the early magnetic afternoon showed a stronger westward component than they did during quiet times, above both stations. This is consistent with increased forcing from ions convecting in the sunward direction in this time sector. Above both Mawson and Davis, the magnitude of the average horizontal winds was greater when conditions were active, which is also to be expected from the relative increase in the importance of momentum coupling from ions under these conditions.

This response to geomagnetic activity has been observed in many previous studies. In the northern hemisphere, Killeen et al. (1995) examined the dependence of thermospheric winds at high magnetic latitudes (74° and 86°) on solar and geomagnetic activity, and the B_y component of the IMF. The data they presented covered a complete solar cycle, and they observed a moderate increase in wind speed under increasing levels of geomagnetic activity at solar minimum. Much greater speed increases were observed during solar maximum, indicating that the polar thermosphere was much less responsive to changes in geomagnetic forcing during solar minimum, a finding that had been reported previously by Aruliah et al. (1991). In this earlier study, westward (sunward) flow in the evening sector was only observed (at solar minimum) for Kp > 4 ($ap > 27$). Weak antisunward flow was observed at lower levels of geomagnetic activity in that time sector. The data presented by Aruliah et al. were for Kiruna, at a magnetic latitude of 64.7°, and thus at a lower magnetic latitude than Mawson (by approximately 5°). In the present work, a change from very weak zonal flow to moderate sunward flow in the evening sector was observed at both Davis and Mawson for mean $ap \sim 20$, and this may be partly explained by the closer proximity of these stations to the auroral oval.

The average solar minimum winds presented by Aruliah et al. (1991) did not show a significant increase in the magnitude of the antisunward flow near magnetic midnight with increasing geomagnetic activity, as was observed above both Davis and Mawson in the present work. However, the width of the antisunward flow region (in magnetic local time) *was* observed by Aruliah et al. (1991) to increase with increasing activity, and this broadening was seen as a movement of the dusk-side edge of the antisunward flow region toward earlier magnetic local times, a finding consistent with the southern hemisphere results presented in this work.

Average winds reported by Sica et al. (1986b) at College, Alaska (magnetic latitude 65°) showed a moderate increase in antisunward flow with increasing magnetic activity, however this was much smaller than the zonal wind response in the early evening. Indeed the character of the zonal flow observed by these authors was similar to that observed in the present work, with very weak zonal flow in the evening sector for $ap \leq 11$, and sunward flow of approximately 100 m.s^{-1} for $12 < ap \leq 38$. The larger speeds observed by Sica et al. were likely due to the higher average levels of solar activity under which these data were recorded. This would be consistent with the findings of Aruliah et al. (1991) and Killeen et al. (1995) that the auroral thermospheric wind response to enhanced geomagnetic activity is much more pronounced under higher levels of solar activity.

Average vertical winds are typically expected to be small, since they often appear as random fluctuations in individual night time-series and thus would not be expected to show correlation over many nights. However there did appear to be persistent features in the vertical winds above both Mawson and Davis. Significant upward vertical winds were seen above Davis in the broad time period between approximately 21:00 and 03:00 MLT. On average this period of upwelling reached a maximum of around 20 m.s^{-1}. The effect of increased geomagnetic activity was to confine the upwelling to a narrower range of magnetic local times, which under active conditions was observed to begin approximately 2 hours later than it did during quiet conditions. Under both quiet and active conditions the upwelling had abated by approximately 03:00 MLT. Above Mawson there were significant downward winds observed in the early magnetic evening during quiet times, which had abated by 20:00 MLT. The magnitude of these winds was similar to the magnitude of the upwelling observed above Davis, approximately 20 m.s^{-1}. Under active conditions this downward wind was only observed before ∼16:30 MLT.

Both of these features (the average upward wind above Davis and average downward wind above Mawson) have been observed by Greet et al. (2002), who combined four years of data to investigate average vertical winds above Mawson and Davis between 1997 and 2000, and examined the effect of increasing magnetic activity. At Davis, Greet et al. observed increasingly larger downward winds in the early evening as Kp increased, which was not observed in the present work. However, Greet et al. also observed the average upward wind around magnetic midnight, of 10-20 m.s^{-1}, the beginning of which was observed at earlier magnetic local times as magnetic activity increased, consistent with the observations reported here. At Mawson, the average downward wind observed by Greet et al. in the early evening showed little variation with magnetic activity, whereas in the present work the average quiet-period vertical winds above Mawson remained downward for longer than did the active-period winds. This is in qualitative agreement with average vertical winds reported by Conde and Dyson (1995b), who observed the same average downward wind in the early evening, which increased in magnitude from 2-3 m.s^{-1} under quiet conditions to 6-7 m.s^{-1} under active conditions.

Sica et al. (1986b) also reported average vertical winds above College, Alaska. College is approximately 5° lower in magnetic latitude than Mawson, and while these data also showed structure in the average vertical wind, it was quite different at each of the low ($ap \leq 11$), moderate ($13 < ap \leq 38$), and active ($ap \geq 39$) levels of geomagnetic activity. Subject to a number of assumptions (see for example Section 8.1), vertical winds are linked to horizontal motions by the requirement of mass conservation. The average vertical motions reported by Sica et al. (1986b) did not appear to correspond to the periods of divergence/convergence observed in the average horizontal winds. However, above Davis, the present data showed quite a close relationship between the average vertical winds and the divergence in the average horizontal winds, and this is further investigated in Chapter 8.

6.3.2 Temperature

Figure 6.3.4 shows a time-series plot (upper panel) of all temperature data (averaged over all zones in each exposure) recorded above Mawson during 2007 and 2008. Temperature data from Davis are not yet considered reliable due to the non-linearity of that instrument's detector which was described in Chapter 5 (Section 5.3), thus only temperatures recorded by the Mawson SDI are presented here. In Figure 6.3.4 data recorded under all levels of geomagnetic activity and cloud cover have been shown (cloud cover does not affect mean all-sky temperatures). Pre-filtering to remove obviously bad spectra was performed by requiring positive spectral area (i.e. positive intensity), however beyond this no extra data rejection was carried out.

Data from all exposures from 2007 and 2008 are shown as grey dots, with the black curve showing a 10-day smoothed version of these data. Temperature estimates from the NRLMSISE-00 model (also smoothed with a 10-day sliding window) are shown in blue. The model was driven by the geomagnetic and solar activity indices (ap and $F_{10.7}$) shown in the lower panel, for an altitude of 240 km. Where gaps occurred in the data (as in late 2008 for example) the smoothed measured temperatures are drawn as a straight line

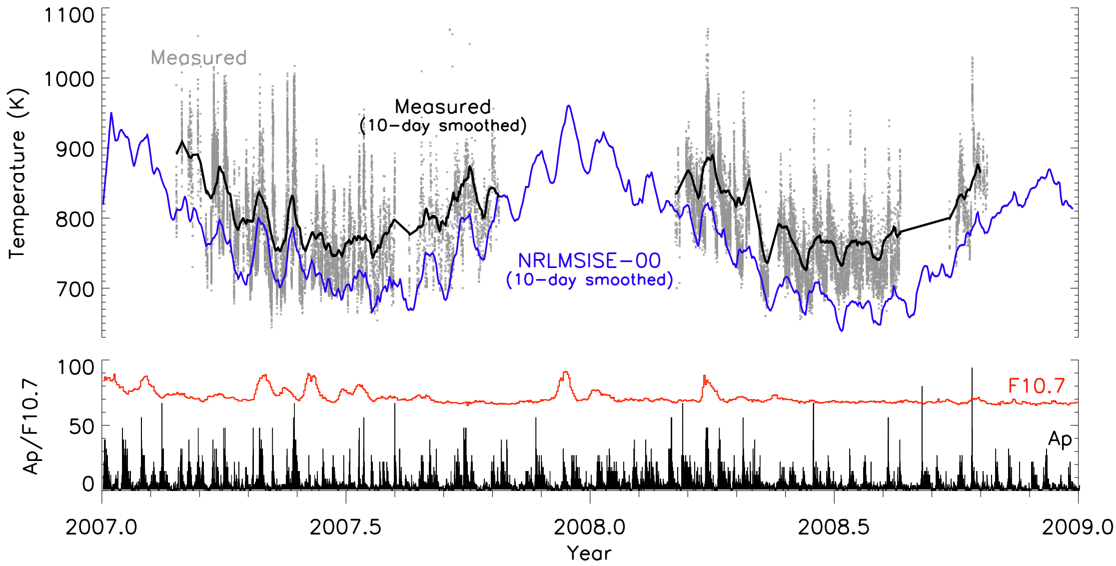


Figure 6.3.4: Upper panel: time-series of temperature data from all exposures from 2007 and 2008. Exposures were included regardless of the level of activity or amount of cloud cover. The grey dots show the raw temperature data, the black solid curve shows the measured temperatures smoothed with a 10-day ('Boxcar') sliding window. The blue solid curve shows the (10-day smoothed) temperature estimates output from the NRLMSISE-00 empirical model. Lower panel: solar (F10.7, red curve) and geomagnetic (*ap*, black lines) indices used to drive the NRLMSISE-00 model.

connecting the data on either side of the gap.

The seasonal temperature variation is clearly visible in the data, with the largest temperatures on average recorded at the start and end of the observing season (i.e. toward summer) and the lowest average temperatures recorded during winter. The smoothed measured temperatures also clearly followed the variations predicted by the model, driven by changes in geomagnetic activity, with a high degree of correlation even down to the smallest visible time variations. Where both measured and model data were available, the Pearson correlation coefficient between the raw (non-smoothed) modelled and measured temperatures was 0.74, which suggests that approximately 55% of the observed temperature variation was predicted by the model. The probability of observing such a correlation purely by chance is less than 1 part in 10^7. Given the variability associated with auroral-region heat sources, the NRLMSISE-00 model was very capable of predicting the average daily temperature variations that were actually observed.

However, while the model was able to predict the temperature variations, there was a systematic offset between observed and modelled temperatures under most conditions. It is clear from Figure 6.3.4 that the model generally underestimated the temperature, however the magnitude of this underestimation was observed to depend on the average level of geomagnetic activity. This dependence is illustrated in Figure 6.3.5. Median measured and modelled temperature (black triangles and blue/red squares, respectively) are plotted as a function of the *ap* index. The data have been separated into two seasonal categories, in order to reduce the effect of the significant average temperature variation during the year. Thus average data are shown for autumn/spring (day numbers 1-110 and

255-365, left column) and winter (day numbers 111-254, right column), with each category containing data from both 2007 and 2008. The data were averaged over *ap* bins of width 1, and the 'error-bars' indicate the 1σ spread in each bin. The lower panels show the (base 10 logarithm of the) number of observations contributing to each *ap* bin.

Both modelled and measured temperatures in both seasonal categories increased with increasing *ap*, as expected from the single-night observations of the same effect (Section 6.2). Both also showed a reduction in their rate of increase at higher levels of *ap*, thus the temperature-*ap* gradient, while positive, showed a decrease with increasing *ap*. The measured average temperatures were usually underestimated by the model, as the offset between the measured and modelled data points makes clear, however the magnitude of the underestimation decreased with increasing *ap*. This is shown more clearly in Figure 6.3.6, where the difference between the measured and modelled temperatures are shown for the same *ap* bins as Figure 6.3.5. The autumn/spring temperature difference (red squares) was generally higher (the model underestimated the temperatures by a larger amount), but followed the same general trend as the winter temperature difference (blue squares). For low levels of geomagnetic activity ($ap < 10$) the observed temperature was 50-100 K higher than predicted by the model. However, for *ap* greater than approximately 35, the average modelled temperatures in both seasonal categories were in quite close agreement with the average measured temperatures, with an absolute difference of $\lesssim 30$ K.

Average daily temperatures during quiet and active conditions are shown in Figures 6.3.7 and 6.3.8. These data have again been separated into autumn/spring and winter seasonal categories, in addition to separation by *ap*. The grey dots show measured temperatures, the red/blue curves show the median temperature in time bins of 1-hour

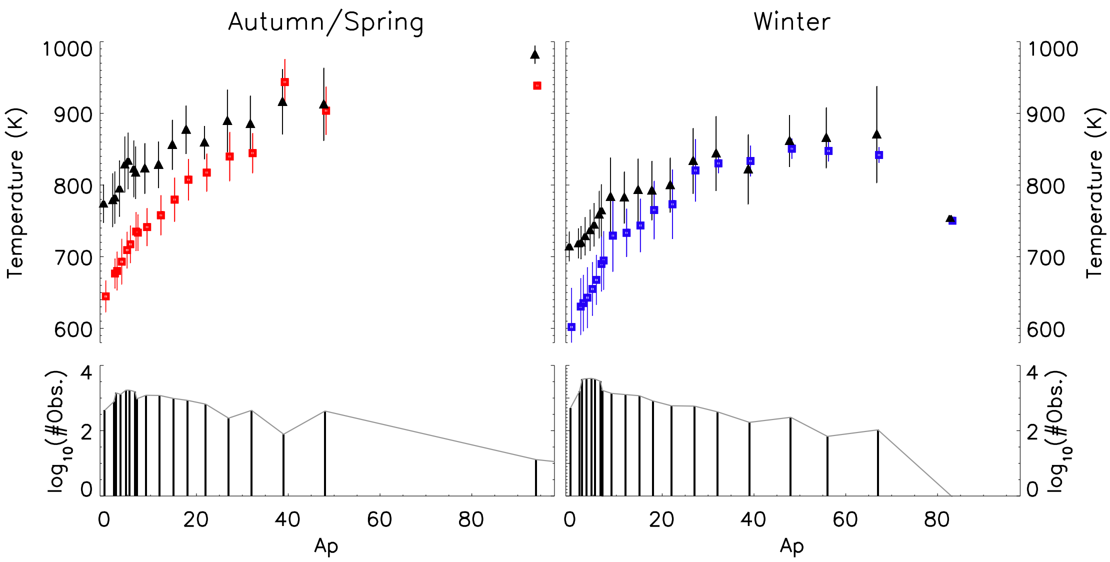


Figure 6.3.5: Average temperatures observed under different levels of geomagnetic activity, averaged over 2007 and 2008. Data are separated by season, autumn/spring (day numbers 1-110 and 255-365, left column) and winter (day numbers 111-254, right column). Observed averaged data are shown in black, NRLMSISE-00 model temperatures in red (autumn/spring) and blue (winter), with 'error-bars' showing the 1σ distribution within each *ap* bin (of width 1).

width during active/quiet conditions, and the solid black curve similarly shows the median modelled temperature in each time bin. The horizontal dash-dot line indicates the mean temperature over the whole time-range, and the vertical bars at the bottom of each panel show the number of observations contributing to each time bin. The mean *ap* level for each average day is shown in the top right corner of each panel. Winter data span a larger range of times than the autumn/spring data due to the longer hours of darkness during winter.

Very similar trends were observed during both years. The highest average temperatures were observed during autumn/spring under active conditions in both years, and the lowest average temperatures were observed in winter during quiet conditions. In 2007 the average active winter temperatures were higher than quiet time temperatures recorded in autumn/spring, while the opposite was observed in 2008 (daily mean temperatures from these figures are summarized in Table 6.1). Mean *ap* was higher in the winter active data set during 2007 than it was in 2008 (23.7 compared to 18.6), while the reverse was true for the autumn/spring quiet data set (4.8 in 2007 compared with 5.5 in 2008), and given the temperature-*ap* relationship in Figure 6.3.5, the different levels of average geomagnetic activity presumably explain the difference between the daily mean temperatures from the two years.

The average nightly temperature trend was similar regardless of season or activity level, with temperatures decreasing from a maximum at the start of the night to a minimum around 21:00-22:00 UT, and then increasing again but generally not reaching as high a temperature as at the start of the night. Exceptions to this trend were seen in 2007, during the autumn/spring active period, where mean temperatures where generally

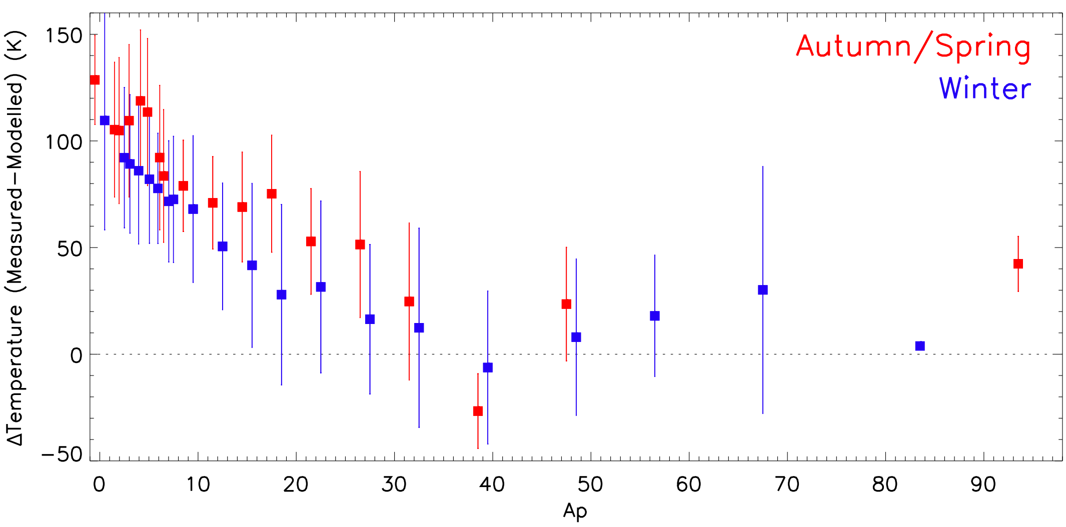


Figure 6.3.6: Difference between median measured and modelled temperatures as a function of *ap*. Positive values indicate that the model underestimated the temperature on average, while negative values indicate overestimation, relative to the average measured temperature within each *ap* bin (*ap* bins are the same as Figure 6.3.5). 'Error-bars' show the 1σ spread in temperature differences in each *ap* bin.

constant throughout the night, and in 2008 during the winter active period, where higher temperatures were observed at the end of the night relative to those at the start. The best agreement between modelled and measured temperature estimates was obtained during the active winter period of 2007, and model temperatures were in general better for the active data than for the quiet data, as expected from Figures 6.3.5 and 6.3.6. For the quiet data the difference between modelled and measured temperatures maximised around 21:00-22:00 UT, when the average nightly minimum temperature was generally observed.

Smith et al. (1998) also reported average temperatures observed above Mawson during three quiet (*ap* 5-15) nights in 1992. These average temperatures showed a very similar trend to the quiet-time data in Figures 6.3.7 and 6.3.8, with the largest temperatures observed at the beginning of the night, and a reasonably smooth decrease to a minimum

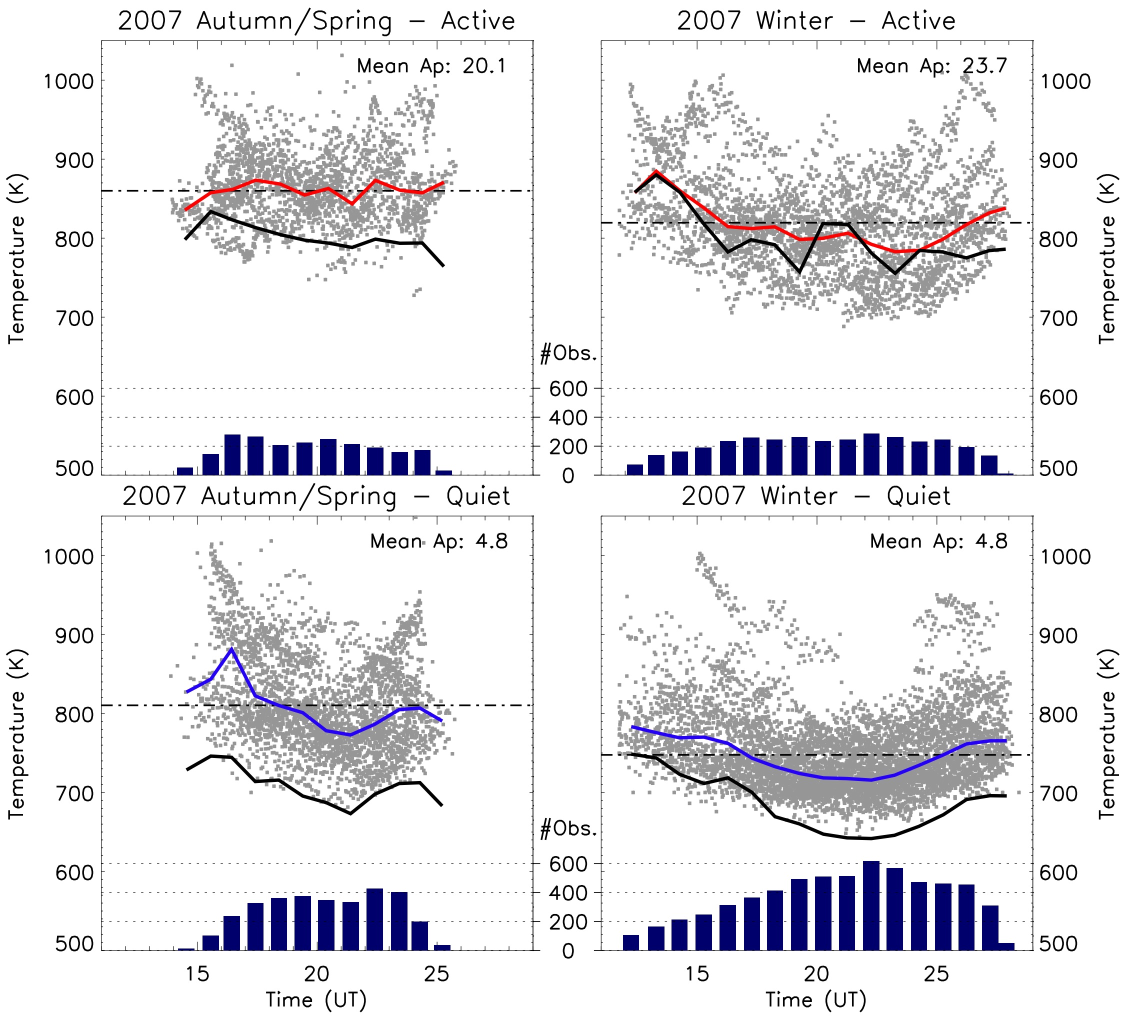


Figure 6.3.7: Median intra-day temperature above Mawson during 2007. Panels are separated according to season (autumn/spring in the left column, winter in the right column) and geomagnetic activity (active in the top row, quiet in the bottom row). Grey dots show all temperature data, red/blue solid curves show active/quiet median temperatures within time bins of width 1-hour. Solid black curve shows the median modelled temperature from NRLMSISE-00 in each time bin, the horizontal dash-dot line shows the mean temperature over the whole time range. Vertical bars at the bottom of each panel show the number of observations within each time bin.

Year - Conditions	Autumn/Spring	Winter
2007 Active	860 K	820 K
2007 Quiet	810 K	745 K
2008 Active	855 K	795 K
2008 Quiet	820 K	745 K

Table 6.1: Mean average daily temperature by season, year, and geomagnetic activity, from Figures 6.3.7-6.3.8.

between 20:00 and 24:00 MLT, followed by an increase toward the end of the night's observations. The temperature range observed by Smith et al. was very similar to that seen in the present work, 800-900 K, with a mean around 840 K.

The nightly temperature trends presented here are also consistent with those reported by Sica et al. (1986b), who showed average temperatures separated by geomagnetic activity, for moderate solar activity ($130 < F_{10.7} < 190$). For low geomagnetic activity, corresponding to the quiet dataset in the present work, mean nightly temperature was 925 K, increasing to $\sim$ 1000 K during moderately active conditions (equivalent to the

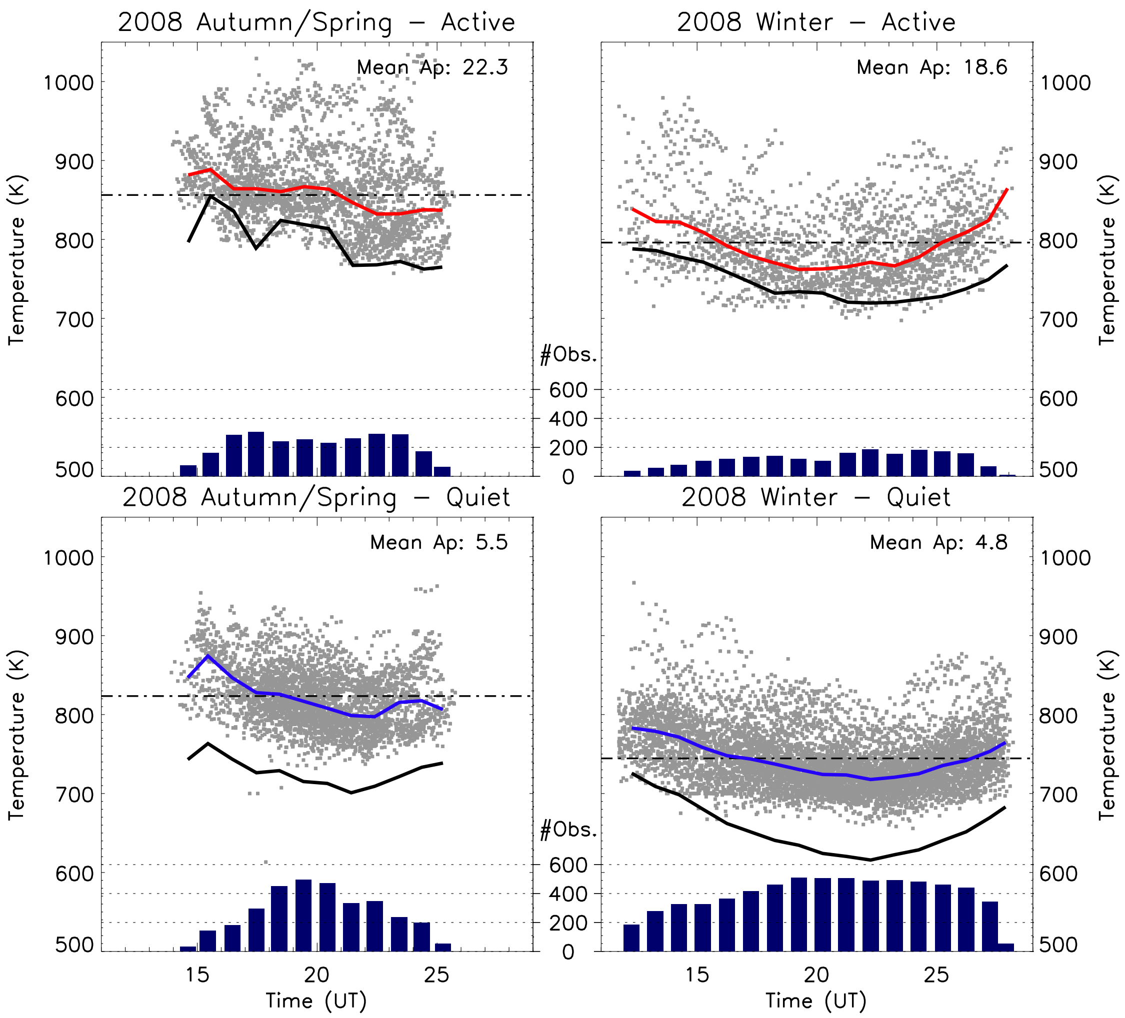


Figure 6.3.8: Same as Figure 6.3.7, for data recorded during 2008.

'active' data in the present work). The larger mean temperatures observed by Sica et al. are explained by the higher levels of solar activity over which these data were averaged compared to the present work. Sica et al. also observed that a change from low to high geomagnetic activity caused as large an increase in mean temperature as did a change from low to high solar activity.

Killeen et al. (1995) reported neutral wind and temperature measurements from Thule Air Base, Greenland (invariant latitude 86°). Although at a higher magnetic latitude than Mawson, these authors also observed an increase in F-region temperatures with increased magnetic activity (as measured by the Kp index), although the correlation between temperature and geomagnetic activity was relatively low, due to the large variability within each Kp bin. For that study the data were restricted to a solar activity range of $150 < F_{10.7} < 180$, approximately twice the level of solar activity (as measured by $F_{10.7}$) available in this present study ($F_{10.7} \sim 80$). Chun et al. (1999) found that the polar cap index, which measures geomagnetic activity in the polar cap, could be used as a proxy for Joule heating. While the *ap* index is a measure of global geomagnetic activity (in contrast to the polar cap index, which is an indicator of high-latitude magnetic activity), it is not surprising that neutral temperatures should show an average increase as the level of geomagnetic activity increases, since Joule heating would presumably be responsible for a significant fraction of that heating, in addition to the contribution from particle precipitation.

Near an auroral zone station like Mawson, the temperature fields often show some local spatial structure, presumably due to the proximity of heating sources associated with the auroral oval. Such structure complicates the comparison between measured temperatures and those estimated by a large-scale empirical model like NRLMSISE-00. In the following section, this average spatial structure is investigated in both the wind and temperature fields.

6.3.3 Magnetic Latitude/Magnetic Local Time Structure

With the imaging capability of the Mawson spectrometer it is possible to investigate the average variation of wind, temperature and intensity with magnetic latitude as well as magnetic local time (MLT), by using the information recorded in every zone (as opposed to averaging over all zones in each exposure, as was done previously). In order to do this, for every exposure, each zone was assigned a magnetic local time based on the current magnetic local time at Mawson station (which corresponds to the central zone) and the difference between the magnetic longitude of that zone and the magnetic longitude of Mawson station (expressed in units of time). Each zone was then assumed to be sampling a different magnetic local time based on the longitudinal separation between the projection of that zone onto the sky (at an assumed altitude of 240 km) and Mawson station. The assumption implicit in this calculation is that Earth's rotation moves the observatory under a wind (or temperature/intensity) structure which is stationary in magnetic local time, allowing magnetic local time sectors to be sampled at different universal times.

For example, consider two zones z_e and z_w which sample the same magnetic latitude, where the subscript indicates an eastward and a westward-looking zone respectively. In a reference frame that is fixed with respect to the Sun, an area of magnetic latitude/magnetic local time (MLAT/MLT) that is sampled by the eastward-looking zone z_e at universal time t_0 will be sampled by the westward-looking zone z_w at time $t_0 + \Delta t$, where:

$$\Delta t = \frac{24}{360} \times (\mathrm{lon}(\mathrm{z}_e) - \mathrm{lon}(\mathrm{z}_w)) \tag{6.3.1}$$

and $\mathrm{lon}(\mathrm{z}_e)$ is the magnetic longitude of z_e, etc. Despite the fact that Δt hours have passed between the two samples, and that the local conditions may have changed in that time, both samples are binned together to produce an average value for that particular MLAT/MLT region. The largest value of Δt will correspond to the largest separation of common-latitude zones, which in the case of the Mawson instrument is equal to approximately one hour. Thus samples of the same MLAT/MLT sector from two different zones will be separated by at most one hour of universal time. The average values calculated in this way will reflect large-scale latitudinal and longitudinal structure, that would for example be associated with the 'steady-state' ionospheric convection pattern. The advantage of this method is that a given magnetic latitude/magnetic local time sector will typically be sampled multiple times, hence there will be many more independent measurements over which to perform the averaging.

Intensity and Temperature

Figures 6.3.9 - 6.3.10 show the result of averaging the intensities and temperatures in this way, where measurements have again been sorted into categories of 'quiet' and 'active' geomagnetic conditions, and pre-filtering for obviously bad spectra was performed as described in Section 6.3.2. In these figures, averages of intensity (upper row) and temperature (middle row), and the number of observations contributing to each MLAT/MLT bin (lower row) are shown for years 2007 and 2008. Averages corresponding to quiet conditions ($ap \leq 10$) appear in the left column, those corresponding to active conditions ($ap > 10$) in the right column. The view in each plot is the same as that in Figure 6.2.9, with the statistical auroral oval (corresponding to the activity level indicated in the top right corner of the plot (Holzworth and Meng, 1975)) plotted in yellow. In order to compare with a typical twin-cell ionospheric convection pattern, electric potential contours from the Weimer96 model are plotted in black on each panel, corresponding to southward IMF ($\mathrm{B_z} = -2$ nT, $\mathrm{B_y} = 0$ nT).

From the lower panel, which shows the number of observations contributing to each MLAT/MLT bin, it is clear that the latitudes closer to Mawson are the most densely sampled, due to the greater number of zones mapping to these latitudes. Another contributing factor is the decrease in the number of observations away from local solar midnight (approximately 20:30 MLT), which is due to the fact that observations in the early afternoon/late morning were only possible close to midwinter, whereas observations around solar midnight were made on all nights the instrument was capable of making measurements. In addition, these numbers should also correlate to some degree with the average intensities shown in the upper panel, since higher levels of signal result in lower exposure times, and hence denser MLT sampling.

One consequence of the sampling is that toward the early magnetic afternoon and early morning, where observations were sparser, there will be some seasonal bias in the results toward observations near midwinter, when the sky was still dark enough to make measurements at those times. In addition, many of the features described below were observed in the early magnetic afternoon (prior to 18:00 MLT) or early magnetic morning, where the number of days contributing to the averages were fewer (particularly during active times). Thus it is possible that some of these features were due to data from a relatively small number of nights, and therefore do not properly represent average behaviour. A much larger data set than is presently available would be required to confirm these as persistent features.

Intensities were much lower in 2008 than during 2007, as is seen from the different scale bars on the intensity plots from these two years. This difference in absolute intensity is likely related to changes in camera settings between the two years, and not necessarily due to a physical difference in sky brightness, however relative changes in intensity would be unaffected by these camera settings. During 2007, quiet-time intensity peaked poleward of Mawson at 18:00 MLT, overhead around magnetic midnight and in a band of higher intensity from approximately 04:00 MLT to the latest time available, 06:00 MLT. This band peaked above and equatorward of Mawson at 04:00 MLT, and was poleward of the station by 06:00 MLT. During 2008, quiet-time intensities were highest in a broad band beginning around 18:00 MLT and continuing until approximately 05:00 MLT, and centred in latitude overhead of Mawson station. During both years the quiet-time intensities corresponded well with the statistical location of the quiet auroral oval.

Intensities were higher and showed more structure during active times. During 2007, active-time intensity showed a large peak equatorward of the station between 14:00 and 16:00 MLT. This peak was part of a larger band of relatively high intensity that continued until approximately 22:00 MLT. A peak was still apparent around magnetic midnight, however it was relatively less significant than during quiet times. The post-magnetic midnight band of high intensity which was present during quiet-times was observed to begin earlier (around 01:00 MLT) and cover a broader range of latitudes during active-times. Active-time intensities during 2008 were quite different to those observed under similar conditions during 2007. Early magnetic afternoon sector intensities were higher poleward of the station, while the peak intensity in this time sector was observed at 18:00 MLT, overhead Mawson. High intensities were observed in a more continuous band of magnetic local times, although there was evidence of the magnetic morning sector peak around 05:00 MLT. Interestingly, while the band of high intensity observed during 2007 in the morning sector could be seen to follow the general shape of the nominal auroral oval, the high intensity region in the early magnetic afternoon clearly did not correspond to the statistical location of the (moderately active) auroral oval. During 2008, intensities showed less MLT structure, and higher intensities were generally observed inside the statistical auroral oval.

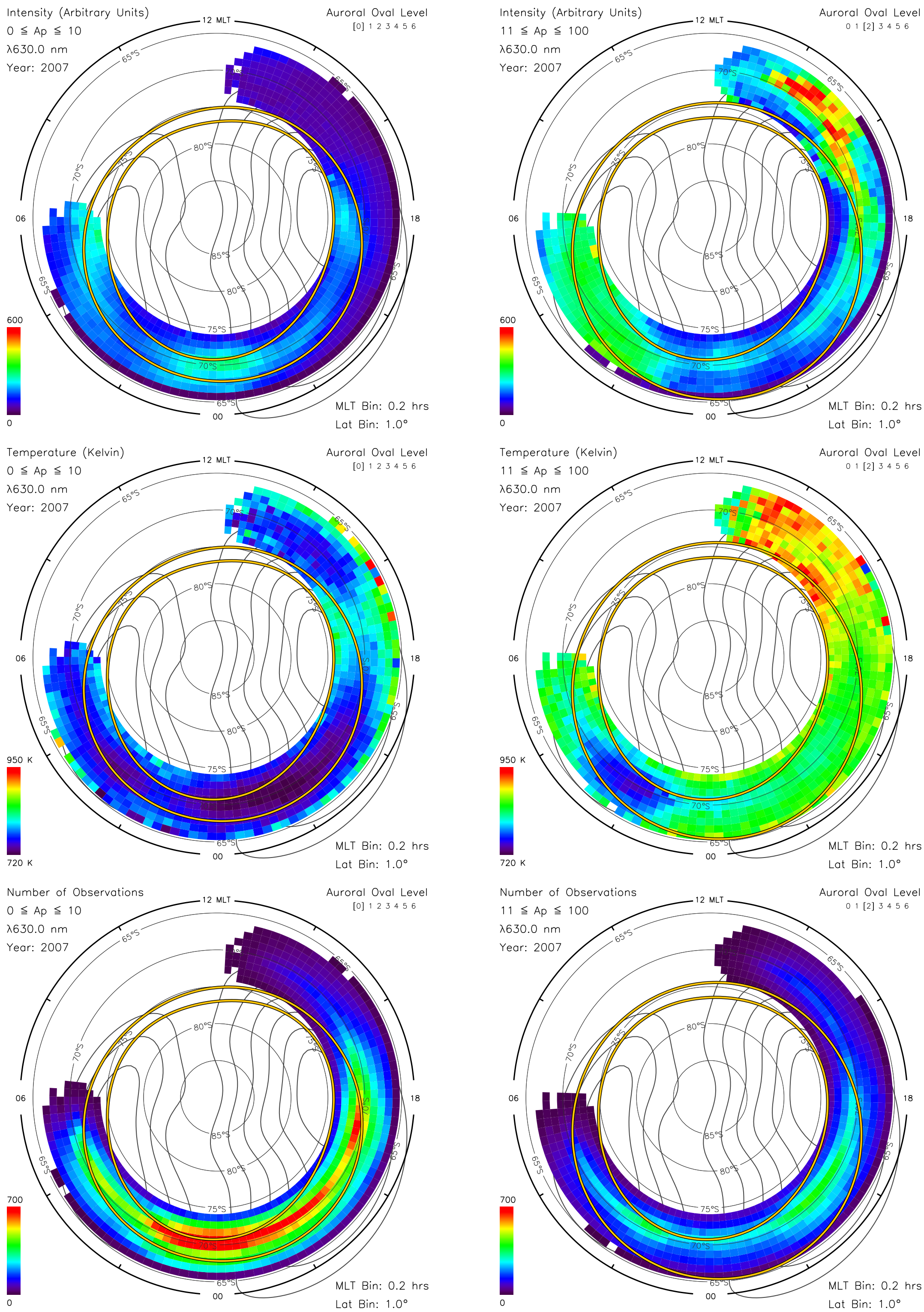


Figure 6.3.9: Average intensities (top), temperatures (middle) and number of contributing observations (bottom), for quiet (left column) and active (right column) conditions during 2007. Contours of electric potential are shown (from the Weimer96 model, corresponding to southward IMF), along with the nominal auroral oval for the activity level indicated in the top right corner of each plot. Color values are indicated by the scale bar in the lower left corner of each plot, latitude and MLT bin widths in the lower right corner.

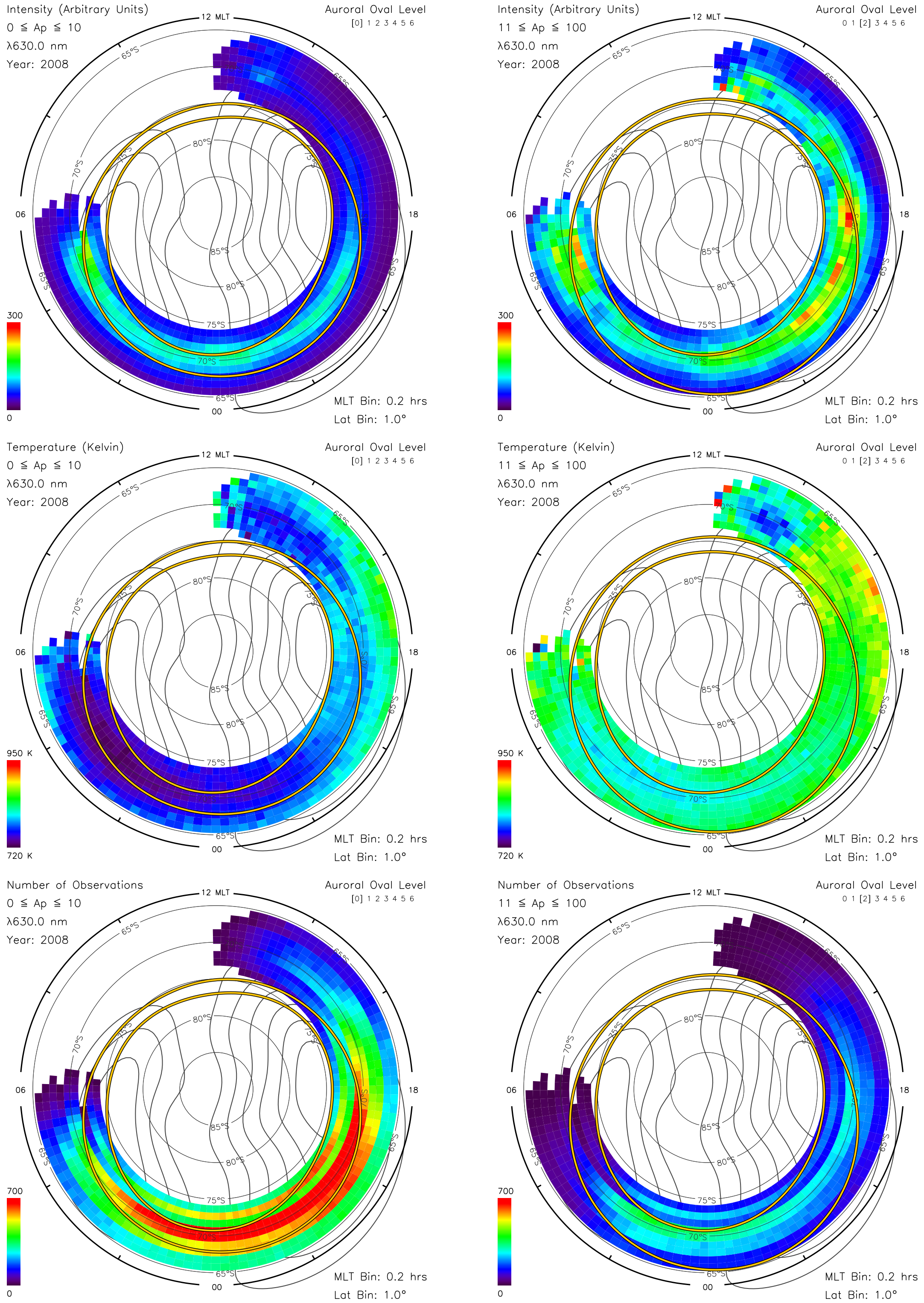


Figure 6.3.10: Average intensity, temperature and contributing observations during 2008. Plot format is the same as Figure 6.3.9. Note the change in the intensity scale.

Quiet-time temperatures were broadly similar in 2007 and 2008. Lower temperatures were observed overhead and poleward of the station in the early magnetic afternoon (12:00-16:00 MLT), with higher temperatures (50-100 K) in a relatively narrow latitude band equatorward of the station between approximately 14:00 and 20:00 MLT. By 17:00 MLT this band had expanded to be overhead of Mawson, at which time a similar region of higher temperature was also observed poleward of the station, forming a saddle-type structure. This structure was more apparent in the 2007 data. Following the saddle-type temperature peak, a broad 'trough' was observed during 2007 centred overhead of the station, and reaching minimum temperatures at approximately 22:00 MLT. This trough covered approximately 4 hours of magnetic local time, and was not apparent in the 2008 data, which showed instead a trough of similar time duration but with minimum temperatures observed around 03:00 MLT. This region of lower temperatures probably corresponds to a much narrower (in magnetic local time) region observed in 2007 around the same time.

Average temperatures observed during active times were also broadly similar between the two years, however there were some interesting differences. In the early afternoon, where quiet-time temperatures showed a minimum, active period temperatures in 2007 showed a broad region of higher temperatures (up to ~950 K), over most of the field-of-view. Before 14:00 MLT, where temperatures were high during 2007, a localised region of relatively low temperatures was observed in 2008. There was also a very pronounced temperature trough observed during 2007 near 02:00 MLT, whereas this region of lower temperatures was not nearly as significant in 2008. Higher average temperatures were observed during 2007 than during 2008, which is probably due to that fact that magnetic activity was greater in 2007.

Horizontal Wind

Average winds from 2007 and 2008, separated by geomagnetic activity and binned by magnetic latitude and magnetic local time, are shown in Figures 6.3.11 and 6.3.12. The format of these figures is similar to Figures 6.3.9 and 6.3.10, where the nominal auroral oval has been omitted, and electric potential is indicated by the orange-green shading (orange representing positive potential, green negative). Blue arrows (all arrows are centred on the location from which the average observation was assumed to be made) show averages of fitted vector winds observed from Mawson, while red arrows are the average winds observed from Davis, which differ from those of Figures 6.3.1 and 6.3.2 because of a change in the widths of the *ap* and MLT bins. While Figures 6.3.1-6.3.2 allowed for direct comparison with HWM93 model wind vectors, Figures 6.3.11-6.3.12 resolve more of the (magnetic) latitudinal structure observed in the wind fields.

Average quiet-time winds observed during both years were very similar to the averages presented in Figures 6.3.1 and 6.3.2 (which were averaged over all levels of activity). Flow was poleward and relatively fast (~100 m.s^{-1}) near magnetic noon, directed across the polar cap. In these figures however there is evidence that wind speed decreased with increasing absolute magnetic latitude in the magnetic afternoon sector, particularly between 12:00 and 16:00 MLT. Winds observed from Davis confirmed this wind deceleration.

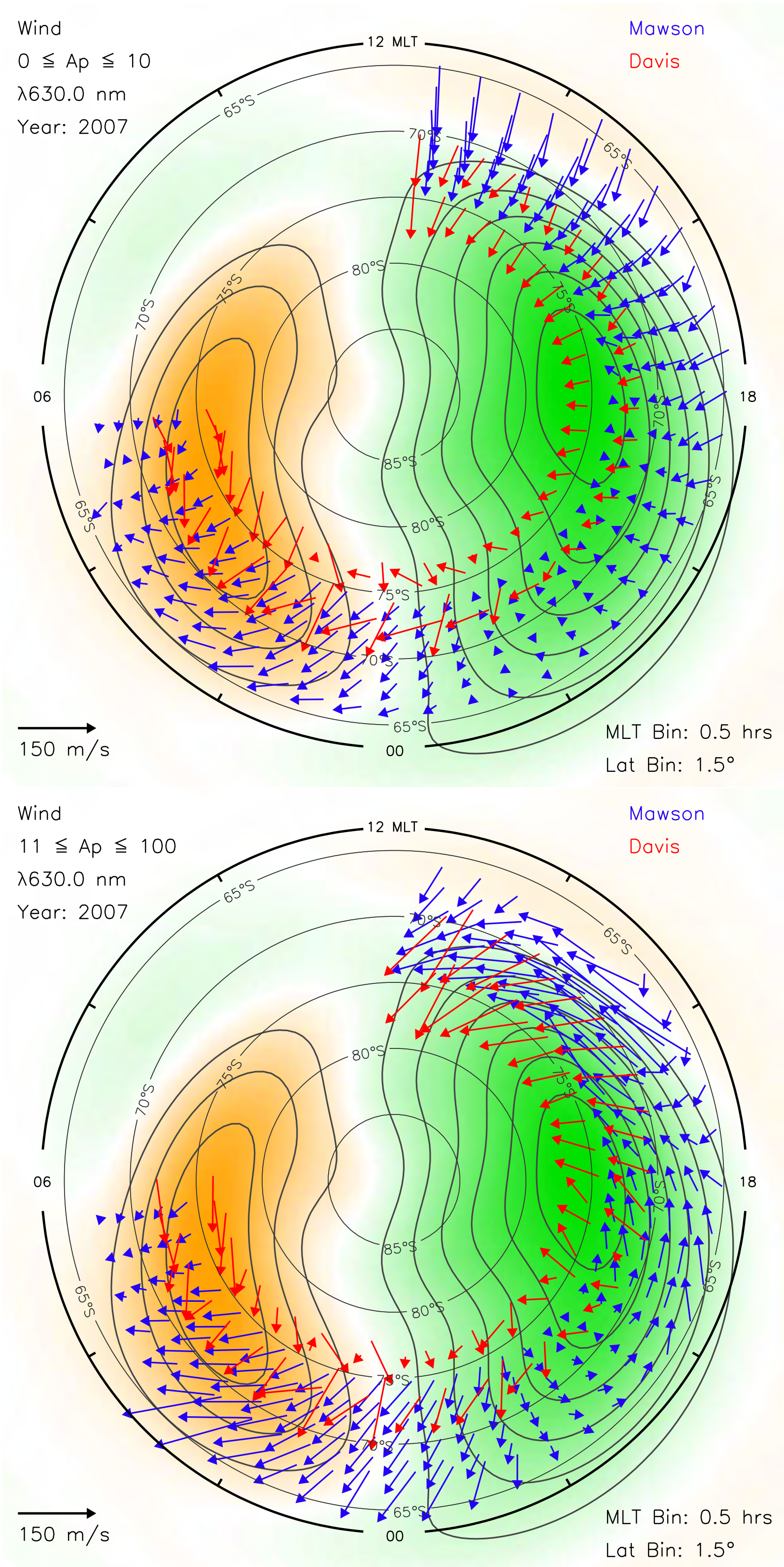


Figure 6.3.11: Average winds observed above Mawson (blue vectors) and Davis (red vectors) during 2007, under quiet (top panel) and active (lower panel) conditions. The model electric potential pattern is indicated by the orange-green shading and black contour lines. Magnetic local times are indicated around the outer circle, and corrected geomagnetic latitude circles are shown. Latitude and MLT bin widths are shown in the lower right corner.

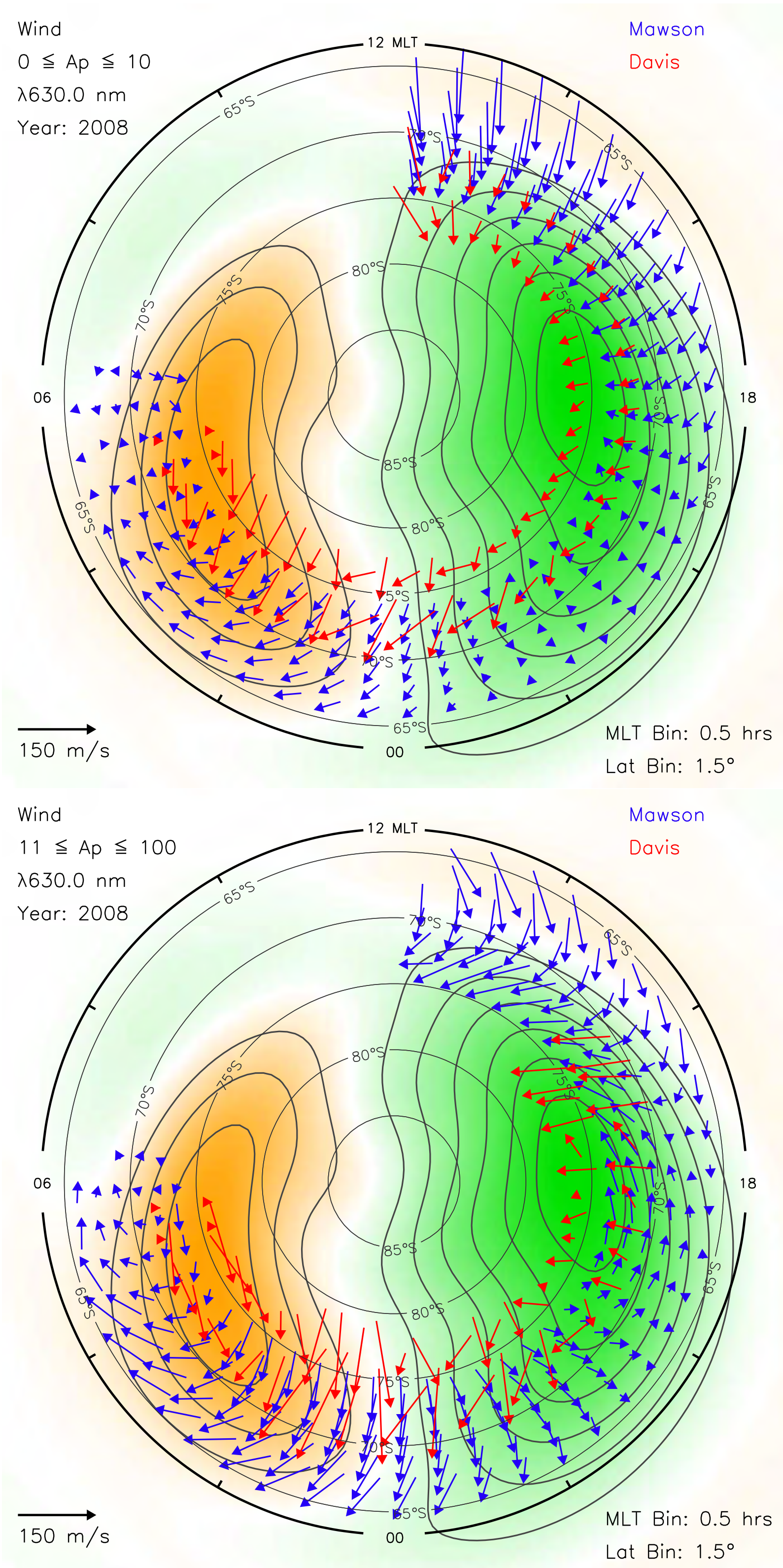


Figure 6.3.12: Average winds during 2008. Plot format is the same as Figure 6.3.11.

Average quiet-time wind speed decreased until the period of almost complete wind abatement near Mawson between 19:00 and 23:00 MLT. Winds observed from Davis however were non-zero in this time period, and were directed magnetically eastward. Beginning around magnetic midnight flow was magnetically equatorward. Winds observed from Davis were on average approximately twice as fast as those observed from Mawson around this time, however the Mawson winds showed an increase in magnitude with decreasing absolute magnetic latitude. Also, winds observed at lower latitudes showed a greater rotation toward eastward flow than did those at higher latitudes, an effect which was particularly evident in the 2008 data. By 06:00 MLT wind speeds observed from Mawson showed another period of abatement. Winds observed from Davis were still around 50 m.s^{-1} near 06:00 MLT during 2007, while in 2008 they showed a similar abatement to the Mawson data.

Active-period winds were very noticeably perturbed from their quiet-time morphology. Close to magnetic noon wind vectors were still directed across the polar cap, however winds in the 13:00 to 20:00 MLT sector showed a strong magnetic westward component, approximately in the direction of the F-region ionospheric plasma flow (compare with Figure 6.2.9). Sunward flow in the dusk sector has been observed many times (e.g. Rees et al., 1983a; Hays et al., 1984; Sica et al., 1986b; Rees et al., 1987; Smith et al., 1988; Aruliah et al., 1991; Conde and Dyson, 1995a; Killeen et al., 1995; Conde and Smith, 1998; Conde et al., 2001; Emmert et al., 2006b), and is associated with momentum forcing from sunward convecting ions.

During 2007, active-time winds observed from both Mawson and Davis were on average faster than during quiet times, and did not show a deceleration with increasing absolute magnetic latitude which was evident in the magnetic afternoon in the quiet-time data. Average active-time winds during 2008 were noticeably different from the 2007 winds in the magnetic afternoon sector. These winds showed an increase in magnitude with increasing absolute magnetic latitude in this sector. The higher latitude winds also showed a greater westward component than those at lower latitudes before 18:00 MLT, with the lowest latitude winds showing an eastward zonal flow in this time sector. This may have been due to a smaller polar cap on average, and a correspondingly higher-latitude sunward ion convection, associated with lower average geomagnetic activity.

The wind abatement, which was observed in the broad MLT region from 19:00 to 23:00 MLT during quiet times, appeared in a narrower time sector during active periods, covering approximately 2 hours of local magnetic time (20:00-22:00 MLT). This region was slightly narrower in 2008 compared to 2007. Periods of very small winds have been observed previously (e.g. Aruliah and Rees, 1995; Conde and Dyson, 1995a; Conde and Smith, 1998; Greet et al., 1999), likely due to a force balance between ion-drag in the sunward direction and solar pressure-gradient in the antisunward direction. The period of low wind speeds observed in the present work appeared somewhat more extended in time than in the works just cited, however comparison with the low solar activity wind maps presented by Emmert et al. (2006b) indicated a similar broad region of very low average wind speed extending from at least 18:00 to 21:00 MLT at the latitude of Mawson. At higher magnetic latitude Emmert et al. observed this feature over a significantly narrower range of local times, consistent with the winds observed above Davis in the present work, where a complete reduction in wind speed was not observed, presumably because it was inside the polar cap at these times and therefore sampled more of the region where pressure-gradient and ion-drag forcing were acting in the same direction.

The wind abatement in the magnetic morning period was also observed during active times, and in 2008 was observed approximately 1 hour later than it was during quiet times. During 2008 winds between 01:30-04:00 MLT (quiet-times) and 01:30-05:00 MLT (active-times) showed a significantly larger westward component at lower absolute magnetic latitudes than was evident during 2007. These winds indicated a flow approximately parallel to the electric potential contours, as would be expected if ion-drag due to the dawn ionospheric convection cell was a dominant contributor to the force balance at these times.

It is interesting that this feature was more apparent during the quieter year, and possibly indicates that the region of polar cap convection was on average smaller (contracted poleward) during 2008 than it was during 2007. This would be consistent with Figure 6.2.9, whereby if the convection pattern were larger (or similarly if Mawson were effectively placed at a higher geomagnetic latitude) then the antisunward flow across the polar cap would be sampled for longer in the magnetic morning, followed by the lower velocities associated with the centre of the dawn vortex. With a smaller convection pattern (Mawson placed at a lower geomagnetic latitude on Figure 6.2.9), the nightside portion of the dawn convection cell would be more readily observed.

The eastward rotation also suggests that ion-drag was effective at modifying the wind flow after magnetic midnight even during periods of low average geomagnetic activity. The Coriolis force would be expected to reinforce the duskside circulation, and inhibit it on the dawnside (Thayer and Killeen, 1993), thus the signature of dawnside convection is often much weaker than on the duskside, where strong sunward flow is frequently observed. No evidence of a sunward rotation in the dawn sector was observed by Emmert et al. (2006b), nor in studies by Conde and Dyson (1995a) (at Mawson) or Hernandez et al. (1990) (at South Pole station), however Smith et al. (1998) did observe the same signature of the dawn convection cell in quiet-time (*ap* 5-15) observations from Mawson. In that data, the eastward rotation was negligible poleward of Mawson, while there was significant eastward rotation equatorward of the station after magnetic midnight, in agreement with the current work. In the present data this eastward rotation suggests that increased ion-neutral coupling after magnetic midnight was sufficiently enhanced to allow ion-drag forcing to dominate the pressure-gradient and Coriolis forces in this time sector, for $\sim 2-4$ hours after magnetic midnight.

6.4 Summary

The results presented in this chapter have demonstrated the average behaviour of F-region neutral winds and temperatures above Mawson, and winds above Davis. In Section 6.2

significant changes in wind flow configuration and variability were observed on a daily basis as the level of geomagnetic activity (for which the *ap* index is a suitable proxy) increased. A trend toward a more sunward flow in the magnetic afternoon (presumably due to increased forcing from convecting ions), and increased levels of divergence (associated with highly variable vertical winds) with increased magnetic activity was observed, along with increased temperatures. An often dramatic increase in vertical wind activity (in both magnitude and short time-scale variability) was also observed.

The trends hinted at in the individual night time-series were confirmed in Section 6.3, where results of a superposed epoch analysis of wind and temperature data were presented. Average winds above Mawson and Davis were compared, and separated into quiet ($ap \leq 10$) and active ($ap > 10$) day curves, which showed that under active conditions zonal flow in the magnetic afternoon was much stronger on average at both stations than it was during quiet conditions, and wind speeds were higher, a feature often observed at auroral latitudes (e.g Rees et al., 1983a; Hays et al., 1984; Sica et al., 1986b; Rees et al., 1987; Smith et al., 1988; Aruliah et al., 1991; Conde and Dyson, 1995a; Killeen et al., 1995; Conde and Smith, 1998; Conde et al., 2001; Emmert et al., 2006b).

A region of antisunward flow around magnetic midnight (identified here with the emergence of the cross-polar jet on the nightside) was also observed earlier and with faster average wind speeds during active conditions, although the dawn-side edge of this flow region was observed at the same time regardless of activity level, and it was postulated that this may have been a result of a larger high-latitude convection pattern (equivalently a larger polar cap region) during increased levels of geomagnetic activity. A similar broadening of the midnight sector antisunward flow with increased activity was observed by Aruliah et al. (1991) in the northern hemisphere. In the discussion on substorms (Chapter 1, Section 1.2.3) it was mentioned that the auroral oval is often observed to expand during the growth phase of substorms, due to energy input from the solar wind. Kamide et al. (1999) concluded that the size of the auroral oval is a good indicator of substorm energy. An average increase in the level of geomagnetic activity would likely be related to increased substorm activity, supporting the conclusion that a larger polar cap convection pattern was at least partially responsible for the observed changes in average wind flow.

Average temperatures were observed to increase with increasing geomagnetic activity, in agreement with previous studies (e.g. Sica et al., 1986b; Killeen et al., 1995). This behaviour was predicted by the NRLMSISE-00 model, however at low levels of activity the model underestimated average temperatures by 50-100 K, and predicted a larger rate of increase of temperature with increasing geomagnetic activity than was actually observed. The links that have been found between increased geomagnetic activity and (in particular) Joule heating (Chun et al., 1999, 2002, for example) make it unsurprising that neutral temperatures showed a clear increase as magnetic activity increased.

For *ap* greater than approximately 35 the observed and modelled temperatures were in close agreement. As discussed in Section 6.2.1, the discrepancy between modelled and measured temperatures at low *ap* may be due to a lack of spatial coverage in the underlying model dataset near Mawson, or, given that the agreement improved with increasing *ap*, a lack of measurements made under conditions like those observed in the present study, namely geomagnetically quiet conditions during an unusually deep and prolonged solar minimum. These factors combined with the proximity of Mawson to the auroral oval, where heating sources can be strong and variable, would make it difficult for a large-scale model like NRLMSISE-00 to accurately predict the intraday temperature variations on which the averages are based.

Also presented in this chapter were average wind, intensity and temperature 'maps', where observations were binned by magnetic latitude and magnetic local time, as well as separated into quiet and active periods. These maps showed, in addition to the variation with magnetic local time, the spatial morphology of the geophysical parameters just mentioned, over the field-of-view of the Mawson spectrometer. In the case of winds, the addition of Davis data provided even greater spatial coverage. The maps showed that in general intensities were well correlated with the location of the statistical auroral oval corresponding to the approximate level of geomagnetic activity. Temperatures also showed a correlation with the location of the statistical oval, with generally lower temperatures located underneath the oval, possibly due to an anti-correlation between intensity and temperature due to the energetic particle precipitation lowering the peak emission height of the airglow.

By combining average winds observed above Davis with the wind maps above Mawson, the detail of the latitudinal variation of the wind flow over a relatively broad range of magnetic latitudes was revealed. During quiet times, the maps showed a decrease of the dayside (12:00-14:00 MLT region) poleward flow magnitude with increasing absolute magnetic latitude. This feature was not evident during active times. The rotation of the vectors toward a westward (sunward) flow in the magnetic afternoon was clearly evident in the active-time data, but more so in 2007, where the rotation could be seen clearly at all sampled latitudes. In the 2008 active-time data, the rotation was not evident at all latitudes, with the lowest latitudes actually showing eastward flow before 18:00 MLT, while the westward flow was largely confined to magnetic latitudes poleward of the station.

The appearance of eastward (sunward) flow in the magnetic morning sector was apparent during quiet times (both years) and active times during 2008. This appeared as an increased eastward component of the vectors with decreasing absolute magnetic latitude, and would be consistent with the appearance of the night-side dawn convection cell. Sunward flow in the dawn sector is usually expected to be much weaker than in the dusk sector, since the Coriolis force acts to oppose the generation of the clockwise circulation on the dawnside. While many previous studies have not observed an average sunward flow in this sector (e.g. Hernandez et al., 1990; Conde and Dyson, 1995a; Emmert et al., 2006b), similar results to those observed in the present work were obtained by Smith et al. (1998), who saw strong eastward rotation after magnetic midnight at magnetic latitudes equatorward of Mawson, also during quiet geomagnetic conditions.

This morning sector sunward flow appeared to be consistent with the geometry of the dawnside ionospheric convection cell, and this may indicate that ion-neutral coupling was sufficiently great in this sector that the convecting ions were able to dominate the force balance in this sector, even during quiet geomagnetic conditions. The ionospheric convection pattern and the neutral flow configuration are also dependent upon the sign of the B_y component of the interplanetary magnetic field (e.g. Förster et al., 2008b), and thus it possible that the wind averages presented in this work were biased toward a particular flow configuration. A more detailed analysis would be required to unambiguously identify the processes responsible for driving the observed sunward flow.

Many of the results presented in this chapter are consistent with an increased level of coupling between ion and neutral species, particularly under active conditions. Ion-neutral coupling has been shown by this and many previous studies to be effective in modifying the large-scale morphology of auroral region thermospheric winds and temperatures. It also raises the possibility of changes on smaller temporal and spatial scales, as the convection electric field can change rapidly in time, and over sufficiently small spatial scales, to provide localised forcing of the neutral thermosphere. Particle precipitation, associated with aurora, can also create highly localised regions of temperature and conductivity enhancements, as can Joule heating. As such, we might expect that on smaller spatial and temporal scales (100's of kilometers and tens of minutes) significant variability could exist in neutral dynamics and temperatures. Such small-scale structure is the focus of the following two chapters.

Chapter 7

Common-Volume Vertical Winds

7.1 Introduction

In this chapter results will be presented of vertical winds derived from bistatic wind observations from the Davis and Mawson spectrometers. Davis station is situated ~ 635 km geographically east of Mawson. These instruments are sufficiently close such that, at F-region altitudes, a sizeable fraction of their fields-of-view overlap in a 'common-volume' region lying between them. By combining each station's estimate of the line-of-sight wind within a common-volume region, two of the three components of the wind field can be unambiguously derived, specifically those components lying in the plane which intercepts both stations and the volume being observed. In the special case that this common-volume region lies along the line (great circle) joining Mawson and Davis stations, then these two components are the vertical wind and the component of the horizontal wind in the direction of the great circle.

A campaign of Davis observations was begun late in 2007 (and continued through 2008) which directed the spectrometer to observe, in addition to the four geographic cardinal directions and the zenith, four common-volume regions lying along the Mawson-Davis great circle. The aim of this campaign was to observe the vertical wind at multiple locations between Mawson and Davis in addition to the vertical wind observations routinely made above each station. From a single station vertical wind estimates are of course confined to the zenith viewing direction only, and as such the winds measured throughout the night will in general display variations arising from both local, rapidly time-varying energy inputs and slowly-varying inputs due to the station being carried (along with Earth's rotation) into regions of forcing which are quasi-stationary with respect to the Sun (Rees et al., 1984b). The 'bistatic' arrangement allows the temporal evolution of the vertical wind field at multiple locations along the Mawson-Davis line to be investigated, providing the opportunity to study the spatial scales involved in auroral region vertical winds.

The motivation for such a study comes from the many observations of strong (100 m.s^{-1} or more) vertical winds at auroral latitudes, winds which can be highly localised spatially as well as temporally (see Section 2.4.2 and references therein). For example, Figure 7.1.1 shows vertical wind measurements from 3 different orbits of the Dynamics Explorer 2 (DE-

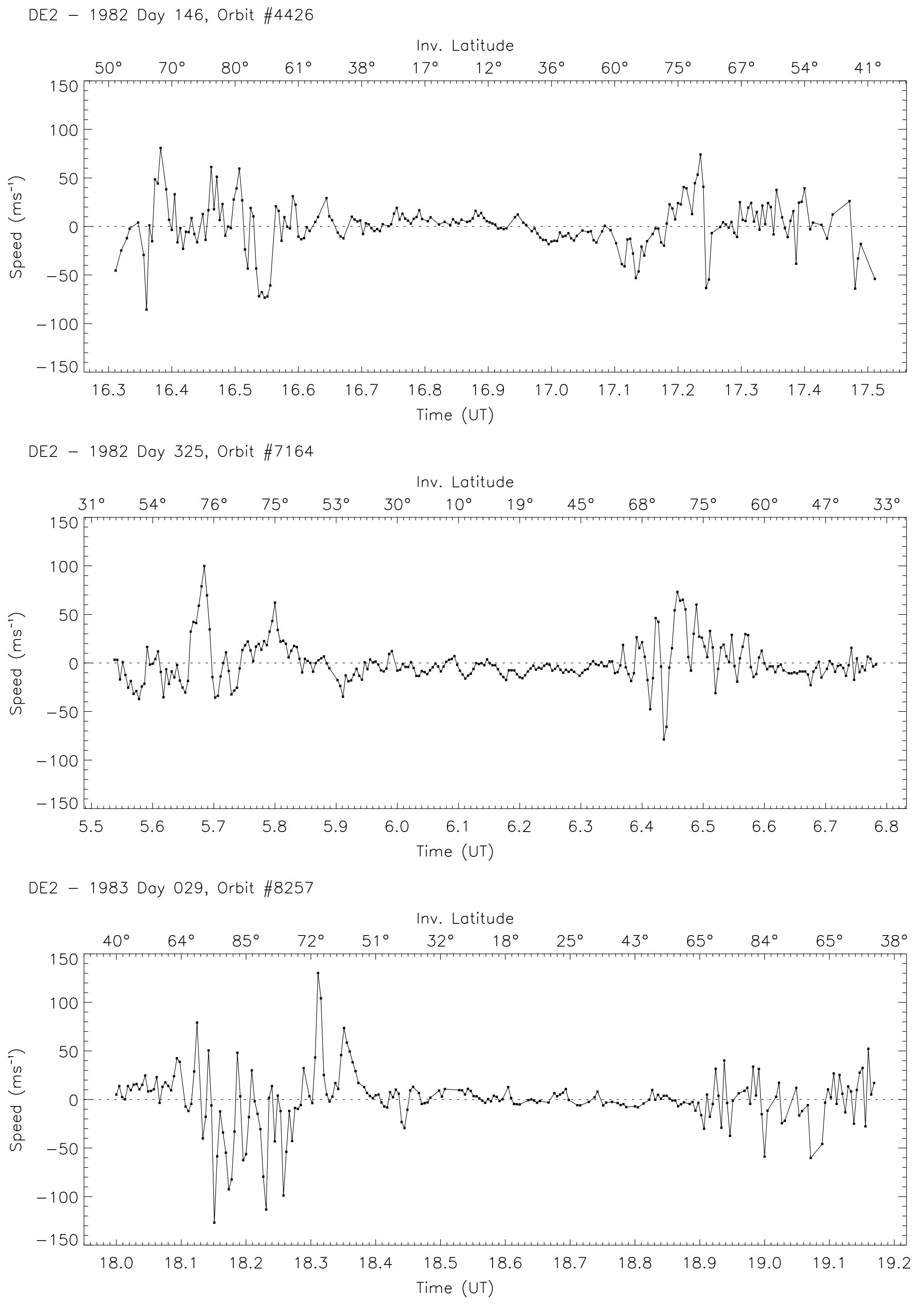


Figure 7.1.1: Vertical wind data from the Wind and Temperature Spectrometer (WATS) instrument on board the Dynamics Explorer 2 satellite. Three individual orbits (on three different dates) are shown. Vertical winds are plotted as functions of universal time and invariant latitude.

2) satellite. These measurements were made by the Wind and Temperature Spectrometer (WATS) instrument on board the satellite (Spencer et al., 1981), which, due to the high orbital velocity of the satellite, essentially provide a 'snapshot' in time of the latitudinal variation of vertical wind activity (over a suitably small range of latitudes). In these solar maximum data the strongest vertical winds are observed predominantly at high-latitudes, with speeds approaching 150 $m.s^{-1}$. Also, the increased variability associated with the high latitude regions is particularly clear. The work of Innis and Conde (2002), which was based on this data, also showed that the greatest vertical wind activity was largely confined to the region bounded by the nominal auroral oval. Previous vertical wind studies from Mawson (Innis et al., 1996, 1999) indicated that strong upwelling events were usually observed when the auroral oval was equatorward of the station. In the first of these studies (Innis et al., 1996), a simultaneous upwelling was observed above both Davis and Mawson, with the largest speeds ($>$ 200 $m.s^{-1}$) observed above Davis, and smaller speeds ($\simeq$ 100 $m.s^{-1}$) above Mawson. This upwelling appeared to move in concert with changes in the location of the poleward edge of the auroral oval.

The locations of Davis and Mawson stations are thus optimally situated to investigate these auroral oval/polar cap region vertical winds, and the bistatic arrangement allows this investigation to be carried out at higher spatial resolution than was previously possible. The all-sky imaging capability of the Mawson instrument also allows the vertical wind dynamics to be compared with gradients and small-scale features in the horizontal wind field, and related to auroral structures in the sky.

7.2 Viewing Geometry and Analysis Technique

The common-volume viewing geometry is shown in Figure 7.2.1. The Mawson zone configuration shown in this figure corresponds to the zone map on the right-hand side of Figure 4.6.1, where these zones have been projected onto the sky at an altitude of 240 km, at the assumed peak of the oxygen red line emission. The spatial extent of these Mawson viewing zones is shown, with the zone number indicated in the corner of each zone. The fields-of-view of the three common-volume Davis look-directions are shown as orange ellipses. Four common-volume viewing regions were included in the usual cycle of Davis observations, however only the three shown in the figure were used in this study, since the fourth region was centred on the Mawson zenith, and therefore provided no new information about the vertical wind directly above Mawson station. The details of the three viewing regions of interest are given in Table 7.1. In this table, Mawson bearings and zenith angles refer to the centre of the given (numbered) zones.

The geometry of an arbitrary common-volume observation is shown in Figure 7.2.2. The positive x-axis has been chosen to point in the direction from Mawson to Davis, while vertical wind is positive upwards. Purple and orange arrows show respectively the Mawson and Davis line-of-sight wind speeds, labelled v_m and v_d. If $\hat{\mathbf{m}}$ and $\hat{\mathbf{d}}$ are used to denote the unit vectors along the lines-of-sight of the Mawson and Davis observations respectively,

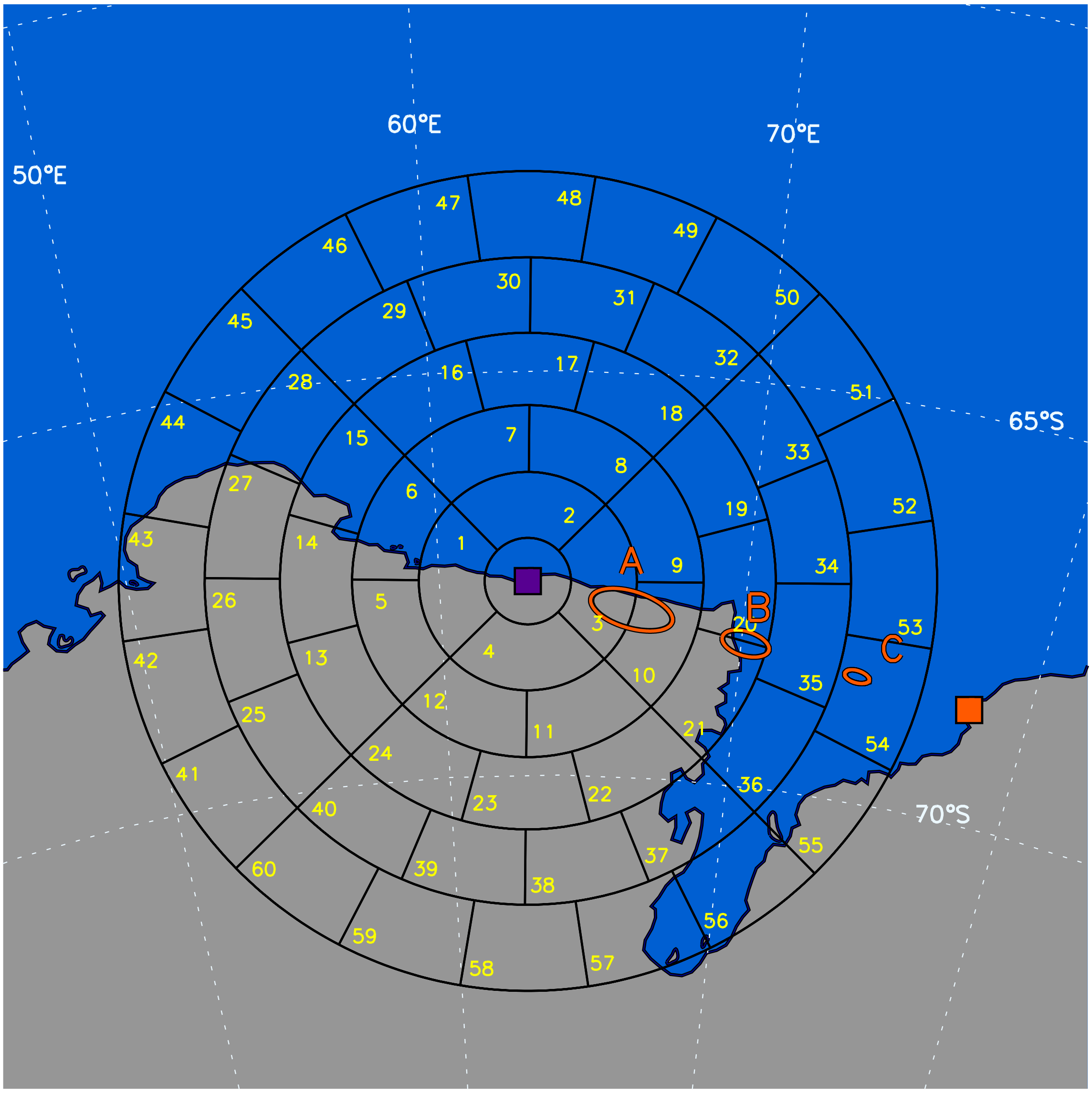


Figure 7.2.1: Map showing the location of Mawson and Davis stations (purple and orange squares respectively) along with the projection of the Mawson viewing zones and the Davis common-volume look-directions (orange ellipses) onto the sky at an assumed height of 240 km. The Mawson zone numbers are indicated in the corner of each zone, while the three common volume regions of interest are labelled A, B and C.

Volume	A	B	C
Davis Zenith Angle	66°	55°	35°
Davis Bearing	N87°W	N87°W	N87°W
Distance from Davis (km)	480	318	160
Mawson Zenith Angle(s)	23.76°, 39.96°	51.84°, 51.84°	66.96°
Mawson Bearing(s)	N91°E, N113.5°E	N91°E, N121°E	N109°E
Mawson Zone(s)	3, 10	20, 21	54
Distance from Mawson (km)	142	286	498

Table 7.1: Common-volume look-directions. Distances from Davis are calculated using the Davis zenith angle, distances from Mawson are calculated using the central zenith angle of the given zone (or the average of two zones where applicable).

and $\mathbf{v}$ represents the true wind vector in the common-volume, then we have:

$$\begin{aligned} v_m &= \mathbf{v} \cdot \hat{\mathbf{m}} \\ &= v_x \hat{m}_x + v_z \hat{m}_z \qquad (7.2.1) \\ v_d &= \mathbf{v} \cdot \hat{\mathbf{d}} \\ &= v_x \hat{d}_x + v_z \hat{d}_z \qquad (7.2.2) \end{aligned}$$

Equations 7.2.1 and 7.2.2 can be written in the form of a matrix equation:

$$\begin{bmatrix} v_m \\ v_d \end{bmatrix} = \begin{bmatrix} \hat{m}_x & \hat{m}_z \\ \hat{d}_x & \hat{d}_z \end{bmatrix} \begin{bmatrix} v_x \\ v_z \end{bmatrix} \qquad (7.2.3)$$

The 2×2 matrix in Equation 7.2.3 is made up of the components of the unit vectors along the lines-of-sight from Mawson and Davis to the common-volume point. For volumes lying between Mawson and Davis, these components are given in terms of the zenith angles of the look-directions from each station, θ_m and θ_d. Equation 7.2.3 can then be solved to give the components of the two-dimensional vector along the $\hat{\mathbf{x}}$ and $\hat{\mathbf{z}}$ directions:

$$\begin{bmatrix} v_x \\ v_z \end{bmatrix} = \begin{bmatrix} \sin(\theta_m) & \cos(\theta_m) \\ -\sin(\theta_d) & \cos(\theta_d) \end{bmatrix}^I \begin{bmatrix} v_m \\ v_d \end{bmatrix} \qquad (7.2.4)$$

These components, along with the full two-dimensional vector, are shown for an arbitrary case as dark green arrows in Figure 7.2.2. One consequence of the common-volume geometry is that a purely vertical wind in the common-volume will result in line-of-sight wind estimates which are correlated between the two stations, while a purely horizontal wind will produce line-of-sight wind estimates at each station which are anti-correlated.

The Davis and Mawson spectrometers, being very different in their design and operation, each produced line-of-sight wind estimates at a different cadence. While exposure times for the Davis FPS were in general shorter than those for the Mawson SDI, the delay introduced at Davis by cycling through the scheduled look-directions generally ensured that the data cadence of the Davis FPS was lower (sometimes considerably) than that of the Mawson instrument. Therefore, in order to fully utilise the available data, obser-

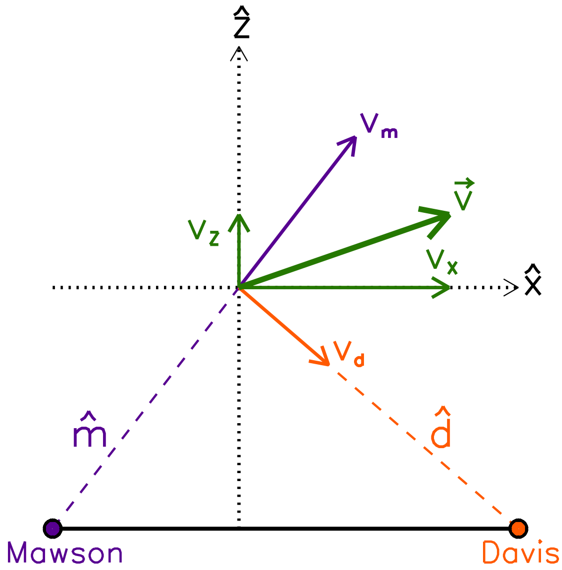


Figure 7.2.2: Diagram showing the Mawson and Davis line-of-sight wind estimates (purple and orange lines) from the given true wind vector (green).

vations from each station were linearly interpolated to a set of 'common times', defined simply by the union of the sets of observation times (universal times) from each station. Thus, for each time at which a common-volume observation was made by either station, two line-of-sight wind estimates were returned, one being an actual measurement from one station, the other being the interpolated estimate from the other station. These estimates (the directly measured and the interpolated) were then combined, using the procedure outlined above, to generate an estimate of the vertical wind and the component of the horizontal wind along the Mawson-Davis great circle.

For periods when the Mawson SDI was acquiring spectra at a higher cadence than the Davis FPS, the most rapid variations of the vertical and horizontal wind estimates derived in the common-volume regions were driven mostly by variations in the line-of-sight wind estimates generated by the Mawson SDI. This is considered acceptable if the high-frequency variation in the line-of-sight winds were due predominantly to variations in the vertical component of the wind, and not the horizontal component. The variability of the line-of-sight wind time-series is frequently observed to decrease with increasing zenith angle, as the example shown in Figure 7.2.3 clearly demonstrates. The data shown in the figure were recorded on April 8, 2008. The point-to-point coherence of the individual wind measurements (after the slowly varying background component was removed by subtracting a 3$^{\text{rd}}$-order polynomial fitted to the original horizontal wind time series) dramatically increased with increasing zenith angle on this night, a behaviour which is frequently observed in the Mawson SDI data. This behaviour can be quantified by examining the variance of line-of-sight wind estimates (within a suitably small time window) as a function of zenith angle. If the high-frequency wind variations were due to the vertical component of the wind we would expect to see the variance decrease with increasing zenith angle, as the contribution from the vertical component decreases with increasing zenith angle.

To measure the variance, line-of-sight wind estimates from each zone (for all clear nights - mean cloud ≤ 3/8 - in 2008) were first filtered by subtracting from them a 3rd-order polynomial fitted to the horizontal wind estimates for that zone. This removed the slowly time-varying background component (similar to applying a high-pass filter). The variance was then calculated for each zone within a 30 minute sliding time window throughout the night to yield a time-series of line-of-sight wind variance from each zone, and the mean variance throughout the night calculated. The width of the time window was chosen such that a sufficient number of data points existed within each window on average, while the window was narrow enough to represent the time-scale of vertical wind variations. The mean variance from all the zones on each clear night were binned by zenith angle, and the average of each zenith angle bin calculated. Averages of the signal-to-noise ratio and χ^2 goodness-of-fit estimate within each zenith angle bin were also calculated, and the results are shown in Figure 7.2.4. As can be seen from the figure the variance clearly

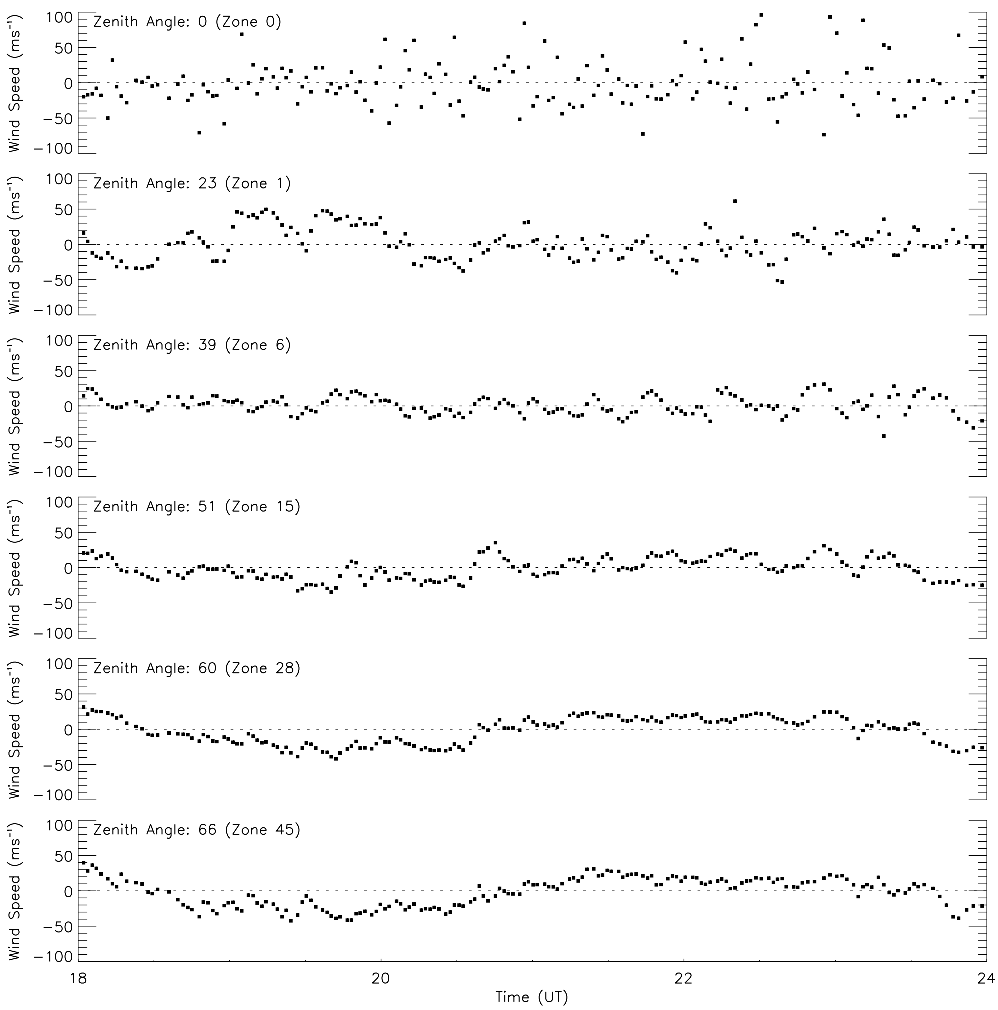


Figure 7.2.3: Horizontal line-of-sight winds measured at different zenith angles on April 8, 2008, after subtraction of a 3rd-order polynomial (fitted to the original horizontal wind time series at each zenith angle) to remove the slowly time-varying background wind.

decreased with increasing zenith angle, as would be expected if the major contribution to the high-frequency variation of wind estimates was due to the vertical component of the wind. For zenith angles $\geq 39°$ the rate-of-change of decrease in variance with zenith angle was greatly reduced, and this was likely due to the fact that the average signal-to-noise ratio also decreased with increasing zenith angle (this is also seen in Figure 7.2.3, as the lower four panels show relatively similar variability, but significantly less than the upper two panels). Lower signal-to-noise increases the uncertainty in the spectral fit, and would therefore be expected to increase the variance. Figure 7.2.4 suggests that for zenith angles $\geq 39°$, the lower signal-to-noise ratio of the spectra in the outer zones was partially compensating for the decrease in variance due to the smaller contribution from the vertical component of the wind.

There are of course other mechanisms which could contribute to the trend of decreasing line-of-sight wind variance with increasing zenith angle. One such mechanism is due to the phase dispersion of the instrument. A zone of a given size which is defined to be closer to the optical axis (closer to the centre of the detector) will span a smaller range

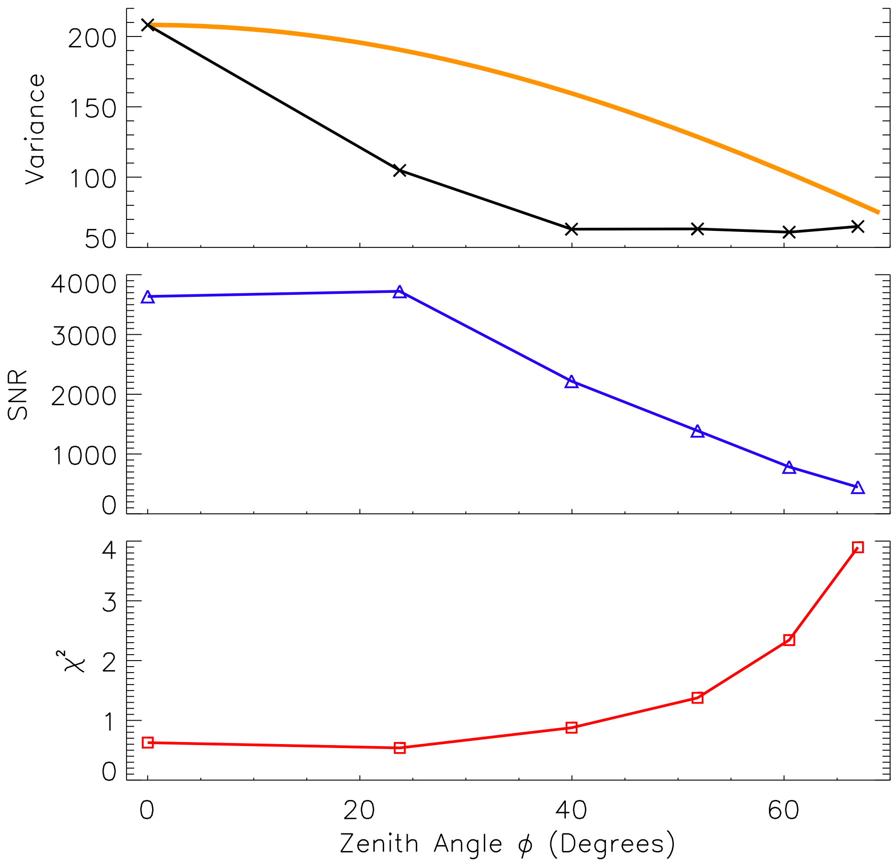


Figure 7.2.4: Top panel: average variance of the (background subtracted) line-of-sight wind speed (black line) as a function of the observation zenith angle. The orange line shows how the vertical contribution to the line-of-sight wind decreases with increasing zenith angle (which decreases in proportion to $\cos(\phi)$). Middle panel: average signal-to-noise ratio as a function of zenith angle. Lower panel: average χ^2 goodness of fit estimate as a function of the zenith angle (lower χ^2 indicates a better fit).

of interference orders than a zone of the same size (pixel dimensions) which is defined closer to the edge of the detector (and therefore mapping to larger zenith angles). This can clearly be seen from the phase map shown in Figure 4.5.2 (top left panel). For the outer zones therefore, at a given step in the etalon scan, the pixels within the zone will contribute to many different spectral channels, effectively 'spreading out' the recorded signal over many spectral channels. For zones closer to the optical axis, the pixels will all contribute to very similar spectral channels. If a time varying brightness were present then this would cause less distortion to the spectra of the outer zones, because of their tendency to 'spread out' the signal over many more spectral channels, reducing the relative contribution to any given spectral channel. However, this effect would appear to be very small, since Figure 7.2.4 shows the χ^2 goodness-of-fit estimate increasing with increasing zenith angle, which would not be the case if the spectral distortion due to time varying brightness were a significant contributor.

There is also the possibility of a statistical bias associated with the likelihood of aurora being present in the field-of-view of any given zone. It is highly likely that aurora will intersect at least one outer zone while the chance of it intersecting an inner zone is relatively lower. Thus an outer ring of zones would be more likely to include spectra with higher signal-to-noise ratio (due to the presence of auroral brightness) than an inner ring, although the relative contribution would be less due to the greater number of zones in the outer rings. The decrease in signal-to-noise with increasing zenith angle does not support this mechanism as a significant contributor. A third possibility is that the perturbations which contribute to the variance are caused by gravity waves (see Section 2.4.4). Wind perturbations due to waves would have both vertical and horizontal components. Inner zones would be sensitive to the vertical component, however the outer zones, being more sensitive to the horizontal component, would not in general be viewing along a line-of-sight which is parallel to the direction of the horizontal wind perturbation, and would therefore be expected to observe less variance in general than the inner zones. The extent to which this last mechanism contributed to the variance curve in Figure 7.2.4 is difficult to assess, however since the majority of the Mawson SDI data show the behaviour exhibited in Figure 7.2.3, the effect of horizontal gravity wave perturbations on the line-of-sight winds, in common-volume regions where the Davis spectrometer is most sensitive to them, are expected to be small.

Since the line-of-sight wind estimates generated by the Mawson SDI only contribute some fraction of the derived vertical wind in any common-volume region (dependent upon the zenith angle of the common-volume region from Mawson), the magnitude of the high-frequency variations in the derived vertical winds (which are driven by variations in the Mawson line-of-sight winds estimates) will in general be underestimated by this technique.

7.3 Results

The technique which has been outlined above requires cloud-free conditions above both Mawson and Davis stations and also within the common-volume regions A, B and C.

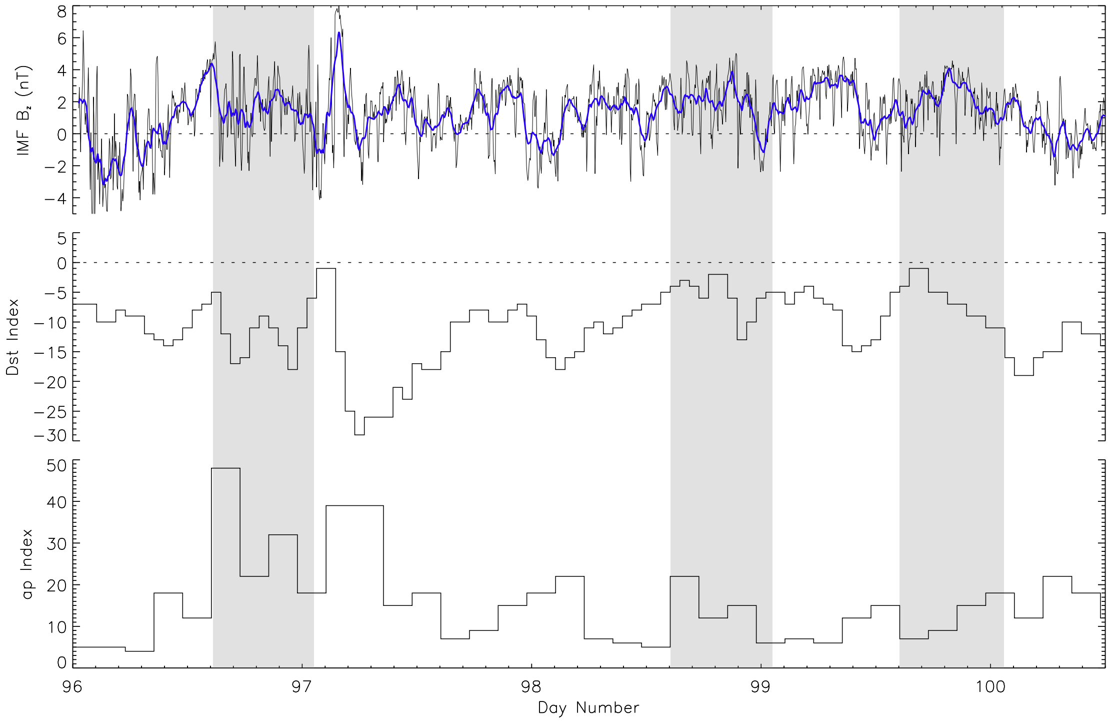


Figure 7.3.1: IMF and geomagnetic conditions for the 6-day period starting April 5, 2008. From the top panel down are plotted the B_z component of the IMF, the Dst index and the *ap* index respectively. The abscissa indicates 00:00 UT of the given day number. Approximate observation periods for each day are shaded grey.

Cloud observations above each station are routinely made by the Australian Bureau of Meteorology, allowing an estimate to be made of the total cloud cover at each site. However, no directional information was provided by these cloud estimates. Thus if a cloud level of only $1/8$ was reported above each station, this cloud may still have been located within a common-volume region, and thus have biased the results. In practice therefore good days were chosen based on both the cloud estimates made by the Australian Bureau of Meteorology and by visual inspection of the resulting vertical wind time series in the common-volume regions. Days which were reported as being heavily overcast above both stations were usually easily identified by common-volume vertical winds which showed appreciable DC offsets or by line-of-sight wind estimates from each station which were highly correlated between common-volume regions and with the station vertical wind.

Nights for which the skies were relatively cloud-free above both stations occurred very infrequently, thus limiting the available dataset. There was however a period of approximately 6 days during 2008 when skies were generally clear and when the derived common-volume vertical wind time-series appeared to have produced consistent results. Three nights from this period will be presented. For reference in what follows, Figure 7.3.1 displays the IMF and geomagnetic conditions which prevailed during the nights for which results are presented, as measured by the B_z component of the IMF (top panel, 4-minute average in black and 1-hour average in blue), and the Dst (middle panel) and *ap* indices (lower panel) respectively. It is worth noting that for the nights reported here, the B_z

component of the IMF was usually (weakly) northward.

Comparison between the emission intensity estimates from each instrument in the common-volume regions revealed a consistent time offset between the two instruments, which was later attributed to a computer clock synchronization problem at Davis. Since each instrument observes along a different line-of-sight, differences in recorded intensity (even when viewing the same nominal volume) can and will occur between the two instruments due to these different lines-of-sight. Such differences are unlikely to be consistent across many nights, however, as was observed. Since a consistent time offset was observed, this could only be due to a clock error. Therefore, in order to maximize the correlation between intensities measured in the common-volume regions, the data from the Davis FPS have been shifted forward in time by 12 minutes.

7.3.1 April 8, 2008 - Day 99

The first night that will be presented, April 8, 2008, was a relatively quiet night (as was most of the 6-day period mentioned above). This night is presented to demonstrate the results of the technique on a night where no major vertical wind activity was observed (however three periods of moderate disturbance are discussed). Figure 7.3.2 shows the line-of-sight wind estimates from Mawson (blue lines) and Davis (red lines) in each of the common-volume regions A, B and C (as labelled in Figure 7.2.1) and also directly above each station, as a function of universal time. The plots are ordered from top-to-bottom by increasing distance from Davis. Of course the line-of-sight wind recorded in the zenith is simply the vertical wind above the station. Error bars indicate the 1σ uncertainties in the fitted spectral peak positions.

In Figure 7.3.3 are shown the vertical wind in each of the common-volume regions derived using the line-of-sight wind estimates from each station, as shown in Figure 7.3.2, along with the vertical winds recorded above each station. The ordering of the plots is the same as that in Figure 7.3.2. For the common-volume regions the error bars indicate the uncertainties in the derived vertical wind estimates, calculated from the 1σ uncertainties in the underlying line-of-sight wind data. Figure 7.3.4 shows the observed (normalised) intensities for each of the common-volume regions and above each station. Intensities estimated from Mawson are shown in blue, those from Davis in red. For this plot, the intensities recorded in each region have been normalised to the maximum intensity observed in that region over the whole night, where this normalisation was done separately for each station. This allowed the observed intensities to be more easily compared between the two stations. On this day the correlation between the intensities recorded by each station was quite good, and gives some measure of confidence that the common-volume wind observations are valid. The recorded intensities represent column integrals of emission brightness along the given line-of-sight, and thus differences between the observed intensities can (and will) arise due to the fact that each station was viewing a given common-volume region along a different line-of-sight. Differences in magnitude between each station are also expected, since neither instrument has been calibrated to give absolute intensities (in

Rayleigh, for example). As mentioned previously, the Davis FPS data have been shifted forward in time by 12 minutes to maximise the correlation between the common-volume intensities measured from each station.

Vertical winds were quiescent for much of this night. Some large, short-lived vertical winds were observed above Mawson, with peak magnitudes of approximately 100 $m.s^{-1}$, while in region A, the closest common-volume region to Mawson, the derived vertical winds were very small. A moderate disturbance was observed in region C between approximately 17:30 and 19:30 UT, which manifested as an initial downwelling of $\simeq 50$ $m.s^{-1}$ followed by an upwelling of comparable magnitude and then another downwelling, again reaching speeds of approximately 50 $m.s^{-1}$. Each up/downwelling lasted approximately 45 minutes. There was a concurrent, albeit less prominent, disturbance observed in region B, although this disturbance did not show such a clear down/up/down structure: an initial downwelling very similar to that seen in region C (peak downward wind speed 50 $m.s^{-1}$) followed by a reduction in downward velocity (reaching ~ 0 $m.s^{-1}$ at approximately 18:00 UT) and another downwelling of slightly smaller magnitude which had abated by 19:15 UT, as had the second downwelling in region C. A second upwelling was observed in region C, which began at approximately 20:40 UT and lasted for ~ 45 minutes. The upwelling reached a peak magnitude of 67 $m.s^{-1}$. At around the same time a similar upwelling was observed above Davis, with a peak magnitude of 80 $m.s^{-1}$. While the beginnings of this disturbance were observed in region B and also to some extent in region A, there was no clear signature of it above Mawson.

In Figure 7.3.5, the vertical winds have been plotted above a keogram derived from the Mawson all-sky images recorded on the night. Each vertical wind time-series has been offset along the y-axis, with the local 0 $m.s^{-1}$ baseline shown for each time-series as a dotted line. The wind scale (y-axis) must be offset accordingly for each time-series. The keogram is comprised of vertical cross-sections through each all-sky image (taken along the geomagnetic north-south meridian) plotted as a function of universal time. Each all-sky image was scaled and smoothed with a 2-point boxcar average before the vertical cross-section through the middle of the image (pixel 'x' coordinate 256) was extracted. The ordinate of the keogram therefore corresponds to projected latitude (with geomagnetic north at the top), and thus the dynamic position of the aurora is clearly tracked by the regions of high intensity. The common-volume vertical winds are plotted using the same time axis to aid comparison, while the vertical wind error-bars have been omitted for clarity. Grey shaded regions are used to indicate the periods of interest.

Each of the vertical wind disturbances mentioned above were seen to coincide with periods of increased auroral brightness. The initial downwelling which was observed in regions B and C around 17:30 UT coincided with the appearance of an auroral arc at the latitude of region B (lower panel of Figure 7.3.5). As the auroral arc moved equatorward, the downwelling in both regions B and C decreased in magnitude, with the most rapid abatement observed in region C. At approximately 18:00 UT the vertical wind in region C became positive, and increased in magnitude over the next ~ 20 minutes until the auroral arc stopped moving equatorward and maintained its position approximately

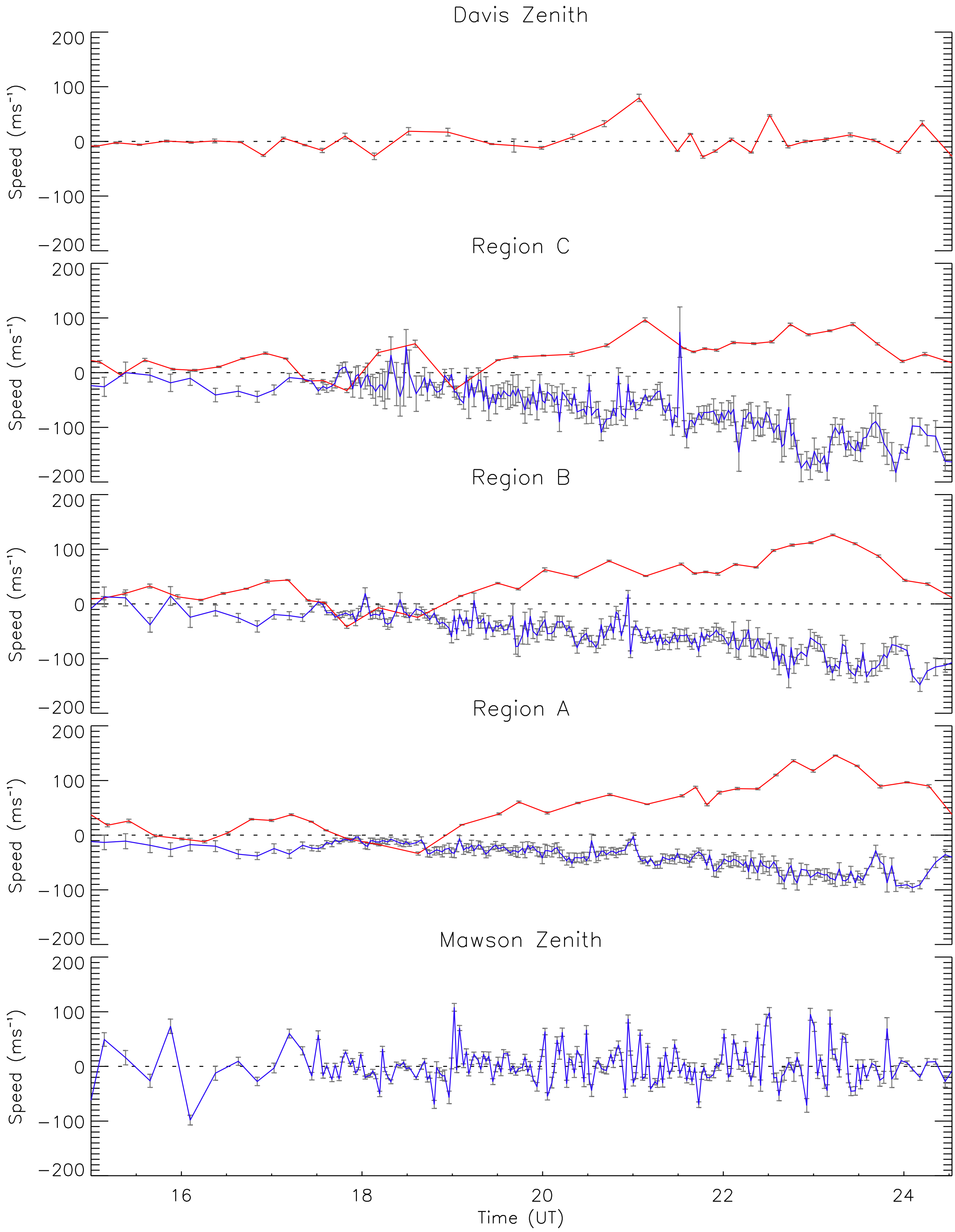


Figure 7.3.2: Line-of-sight wind estimates from Mawson (blue) and Davis (red) in each of the common-volume regions and above each station, for April 8, 2008 (day 99).

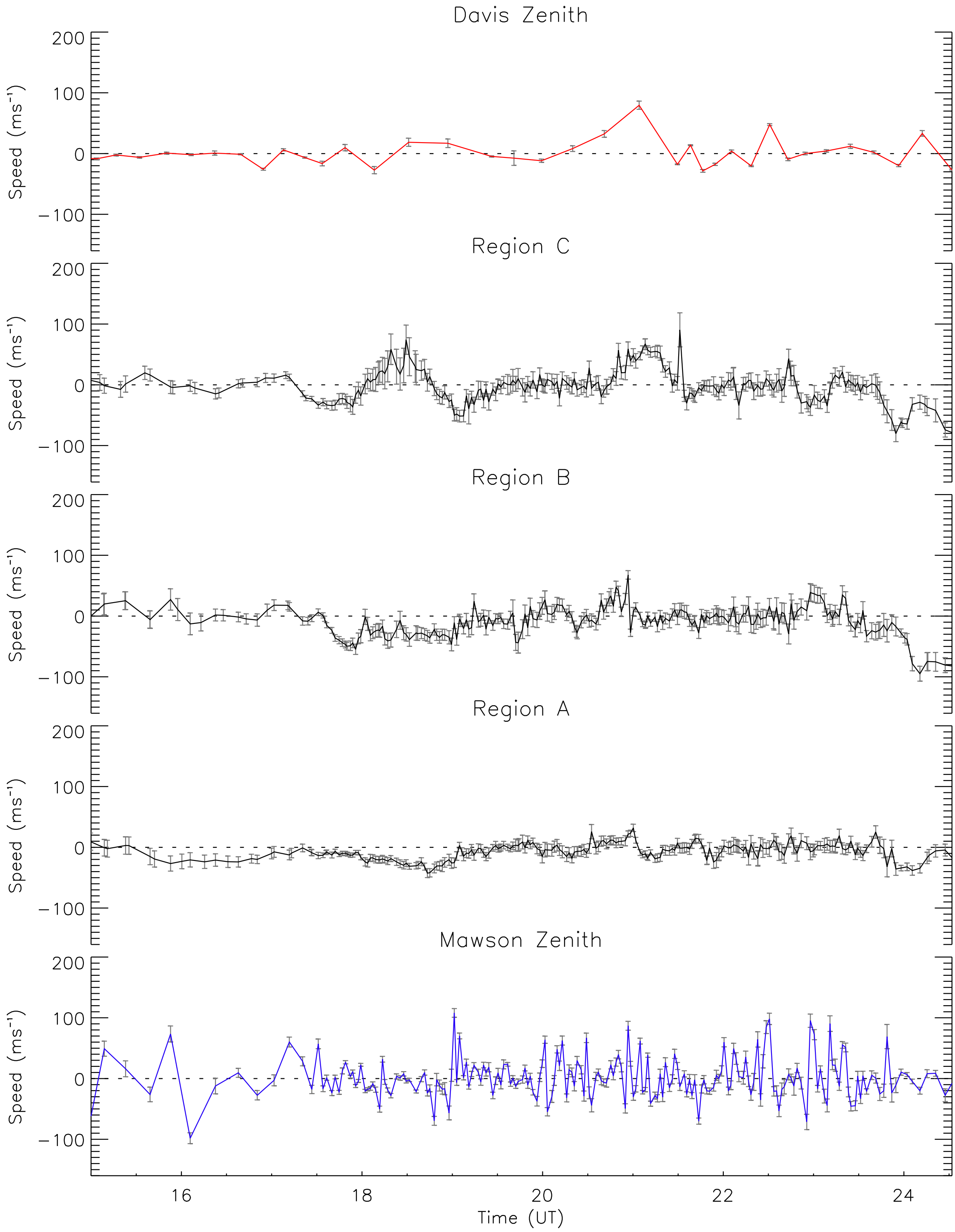


Figure 7.3.3: Derived vertical winds in each of the common-volume regions, and vertical winds recorded above each station, for April 8, 2008 (day 99).

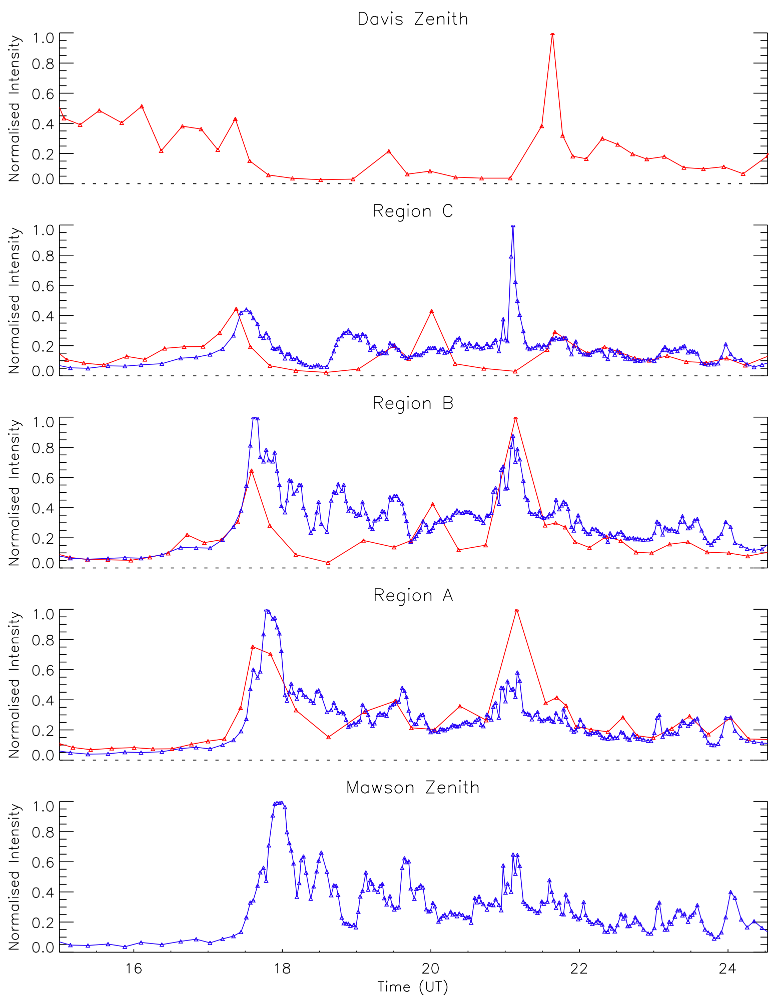


Figure 7.3.4: Normalised intensities observed in each of the common-volume regions, and above each station, for April 8, 2008 (day 99).

2° geomagnetically north of Mawson station. By the time the aurora had halted its equatorward motion, the downwelling in region B had begun to increase in speed once more, and was tracked by vertical winds of similar magnitude and direction observed in region A, although with less variability. The keogram indicates that some auroral brightness was still present at the latitude of region B during this period of up/downwelling, while at the latitude of region C the auroral intensity was very low. Two short-lived vertical wind spikes were observed in region C before the upwelling in that region subsided. This was followed by another downwelling in region C and the continued abatement of the downwelling in region B, which was coincident with a fading of the auroral arc and a (relative) increase in the region C auroral brightness. By approximately 19:30 UT the downwelling in all three common-volume regions had subsided, as had the auroral activity.

Following this disturbance was a period of approximately 1 hour during which time auroral brightness levels remained low and vertical wind magnitudes were generally less than 30 m.s^{-1} (except above Mawson, where there was significant variability throughout the night). At approximately 20:40 UT another upwelling was recorded above Davis

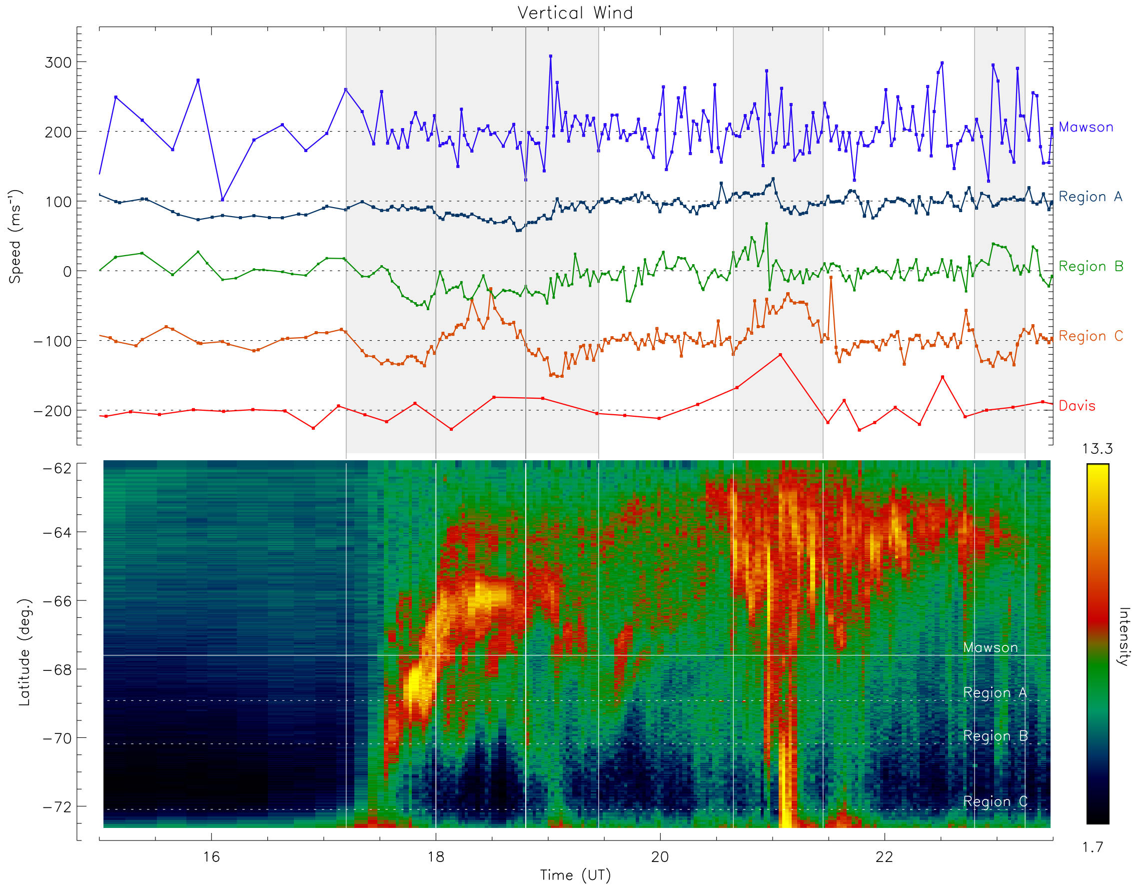


Figure 7.3.5: Top panel: Vertical winds observed above each station and derived in regions A (blue), B (green) and C (red), for day number 99 (error-bars omitted). Lower panel: Keogram derived from the Mawson all-sky images (explained in the text). The approximate latitudes of the common-volume regions and of Mawson station are indicated by horizontal lines. The intensity scale is indicated by the color bar on the right (arbitrary units).

(though it may have started earlier than this time) and in the common-volume region closest to Davis, region C. The beginning of this disturbance was also observed in region B. At the start of this upwelling the auroral brightness had increased substantially at latitudes magnetically equatorward of Mawson, while during the peak of the upwelling the aurora had expanded to cover most of the field-of-view. By the time the upwelling had subsided the aurora was once again located equatorward of Mawson, and was beginning to decrease both in brightness and in latitudinal extent.

One final event worth noting was a moderate (40 m.s^{-1}) upwelling which was observed in region B between 22:45 and 23:15 UT. This upwelling coincided with a 30 m.s^{-1} downwelling in region C. No disturbance was observed above Davis or in region A, while above Mawson some strong 50-100 m.s^{-1}, short-lived vertical winds were observed, although such winds were observed above Mawson relatively frequently throughout this night. During this last disturbance the auroral arc was dimming, and by the time the up/downwelling had abated the discrete arc had disappeared, and only diffuse auroral brightness remained in the field-of-view. In Section 7.3.3 a similar (stronger) anti-correlated vertical wind event will be presented, which also coincided with the disappearance of an auroral arc.

7.3.2 April 5, 2008 - Day 96

This night, April 5, 2008, comes from the beginning of the 6-day period mentioned previously, and was the most geomagnetically active night from this period. From Figure 7.3.1 the *ap* index was seen to reach a maximum of 48 at the start of the night, while the IMF was directed predominantly southward during the daylight hours before observations began. The line-of-sight winds, vertical winds and intensities are shown in Figures 7.3.6 - 7.3.8 (the format of these figures is the same as in Figures 7.3.2 - 7.3.4). Common-volume vertical winds showed significantly more variability on this night compared for example with the night of April 8, although above Mawson the variability was perhaps comparable between the two nights.

Between 16:00 and 16:15 UT a disturbance was observed in all vertical wind time-series to some extent. Above Mawson there was a strong upwelling of approximately 80 m.s^{-1}, coincident with a small downward wind spike of 30 m.s^{-1} in region A. A small downwelling which had begun in region B around 10 minutes prior to this event was interrupted by an abatement of the downwelling during this disturbance, with the vertical wind decreasing almost to zero before increasing again to a 40 m.s^{-1} downwelling after the disturbance had passed. In region C a 50 m.s^{-1} downward wind was observed during the disturbance, and above Davis a minor (30 m.s^{-1}) upward wind was observed. Prior to this disturbance (see Figure 7.3.9) an auroral arc which had appeared on the poleward edge of the field-of-view had moved rapidly equatorward. The disturbance was approximately coincident with a breakup of the discrete auroral arc as it reached its greatest equatorward extent.

Following the auroral breakup another discrete arc formed closer to, but still equatorward of, Mawson. This new arc then moved 1-2° magnetically northward, at which time a strong 100 m.s^{-1} downwelling was observed above Mawson. Downwellings were also observed in each of the common-volume regions, strongest in regions A and B, closest to Mawson, but also visible in region C. No significant event was observed above Davis at this time, however a 40 m.s^{-1} upwelling was observed to peak there near 17:30 UT (accompanying a much stronger upwelling above Mawson), during a brightening of the auroral arc associated with an increase in its latitudinal extent. By the time this arc had disappeared around 18:15 UT the vertical wind activity in all regions had decreased significantly. However, during the period of relatively low auroral intensity between 18:10 and 20:00 UT there appeared periodic variations in regions A and B which were suggestive of waves.

Beginning at 20:00 UT the previously diffuse aurora which had been covering the field-of-view contracted equatorward, forming a narrow region of relatively higher intensity in a band centred 4° equatorward of Mawson, and as a result the intensity observed overhead and at latitudes magnetically southward (poleward) of Mawson dropped to very low levels. As the aurora became concentrated in the northward band an upwelling was observed in all three common-volume regions and above both stations. The upwelling had a different structure in each of these regions. Above Davis, a steady increase in upward vertical wind was observed, beginning at 20:00 UT and reaching a maximum speed of 100 m.s^{-1}. The

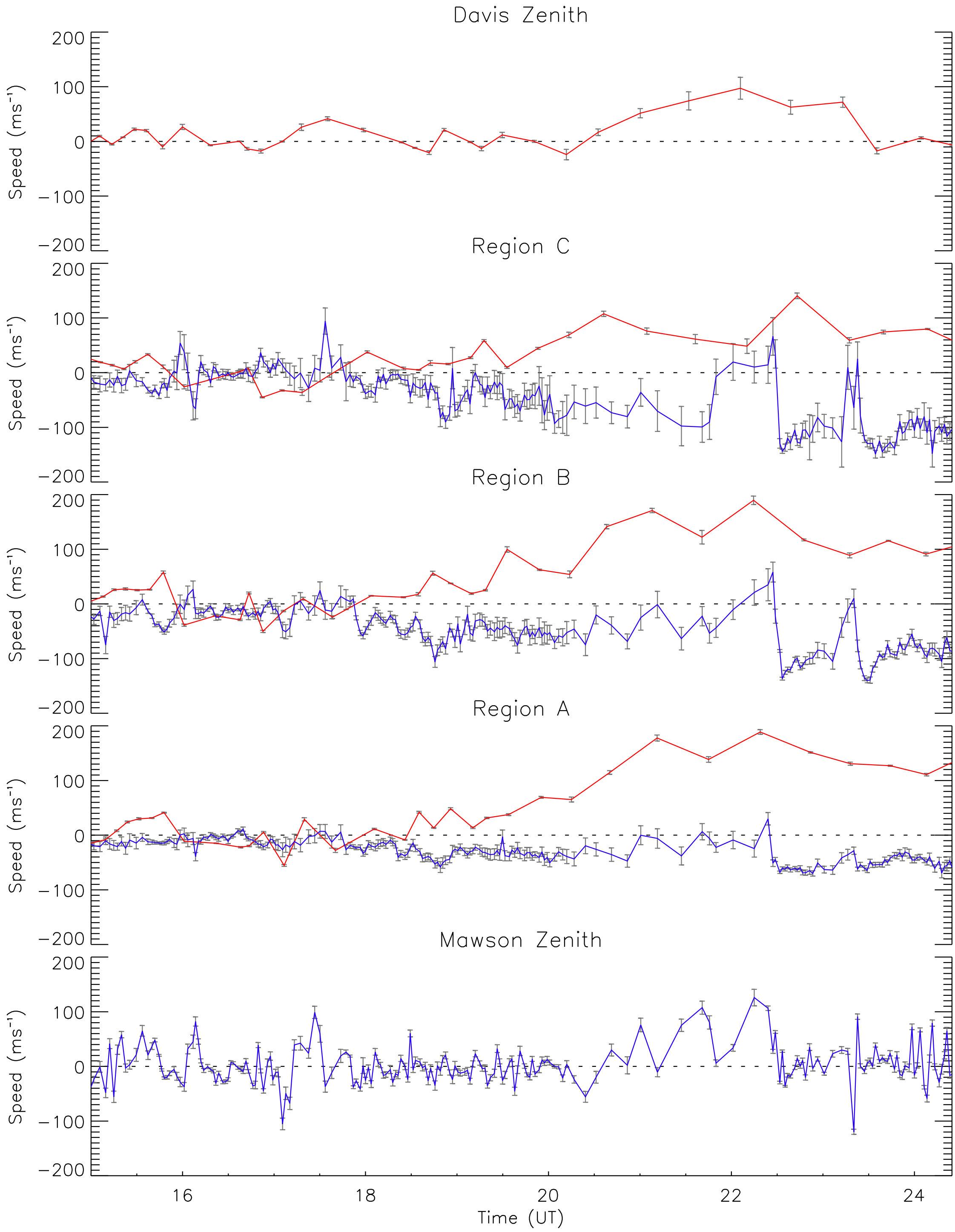


Figure 7.3.6: Line-of-sight wind estimates from Mawson (blue) and Davis (red) in each of the common-volume regions and above each station, for April 5, 2008 (day 96).

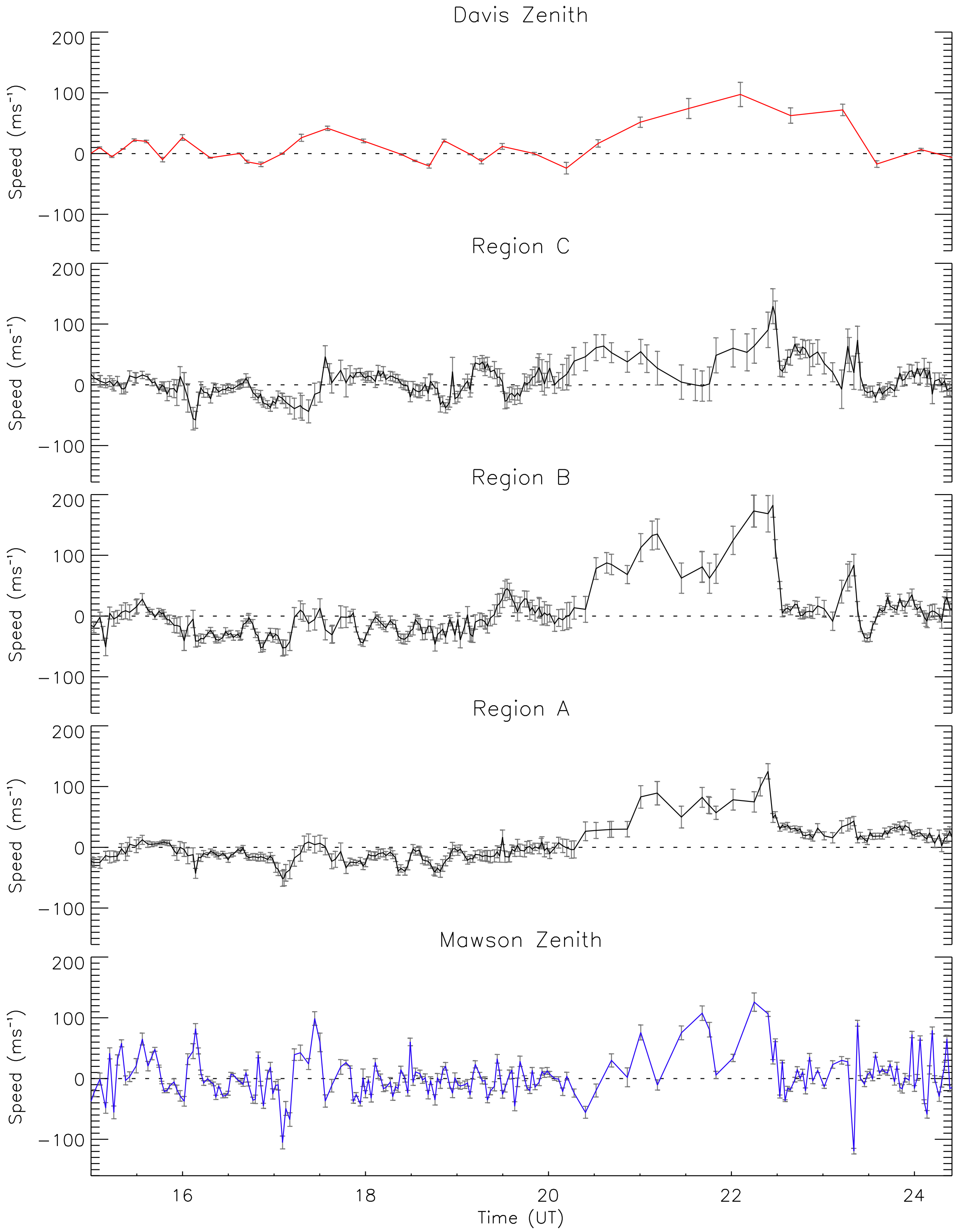


Figure 7.3.7: Derived vertical winds in each of the common-volume regions, and vertical winds recorded above each station, for April 5, 2008 (day 96).

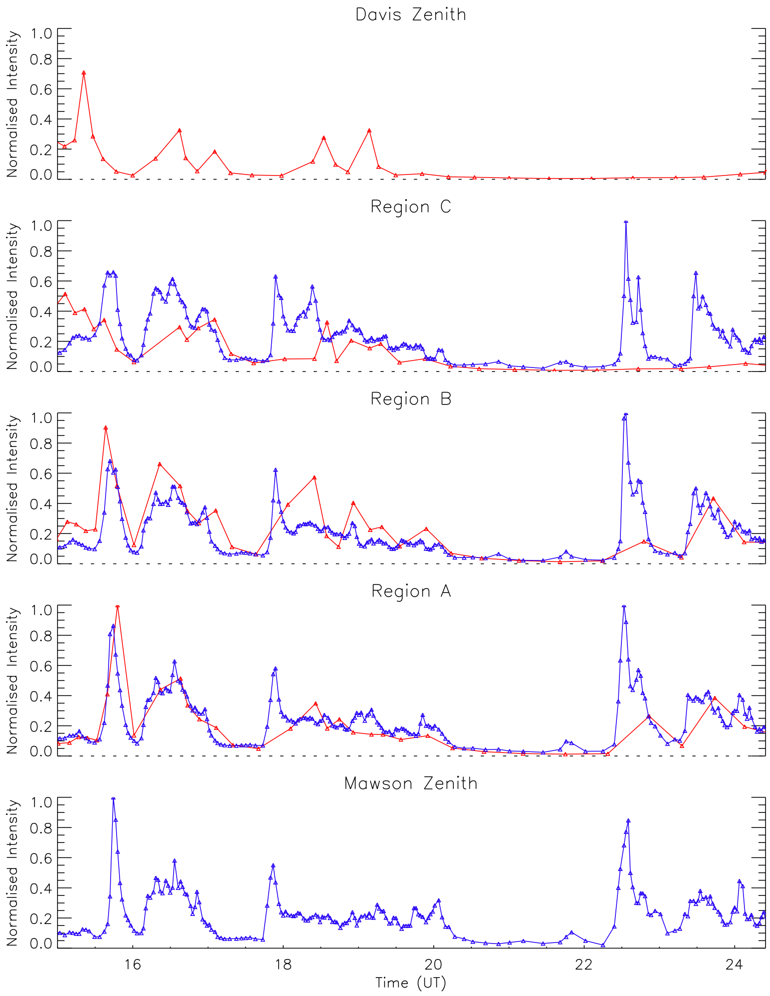


Figure 7.3.8: Normalised intensities observed in each of the common-volume regions, and above each station, for April 5, 2008 (day 96).

upwelling above Davis had abated by 23:30 UT.

In region C, the region closest to Davis, an initial upwelling of 40 m.s^{-1} was observed at the same time as the upwelling above Davis was beginning. By 21:40 UT this upwelling had abated, and was followed by a much stronger upwelling which reached a peak speed of 130 m.s^{-1} at 22:30 UT before abruptly dropping back to 30 m.s^{-1} within $<$ 6 minutes, and subsequently showing another upwelling followed by a more gradual abatement. Finally, between 23:10 and 23:30 UT, as the upwelling above Davis was ending, a 60-70 m.s^{-1} upwelling was observed in region C. The trend in region B was very similar to that in region C. An initial upwelling beginning at 20:30 UT reached a peak speed of 130 m.s^{-1} which then decreased to around 60 m.s^{-1} at 21:30 UT before strengthening again and reaching a speed of approximately 180 m.s^{-1}. As was observed in region C this upwelling stopped abruptly at 22:30 UT, however the vertical wind speed in region B dropped back to zero almost immediately, showing no sign of the upwelling and gradual abatement which was observed after the sudden drop in speed in region C. Following this extended upwelling another short-duration upwelling of 80 m.s^{-1} was observed, concurrent with the similar

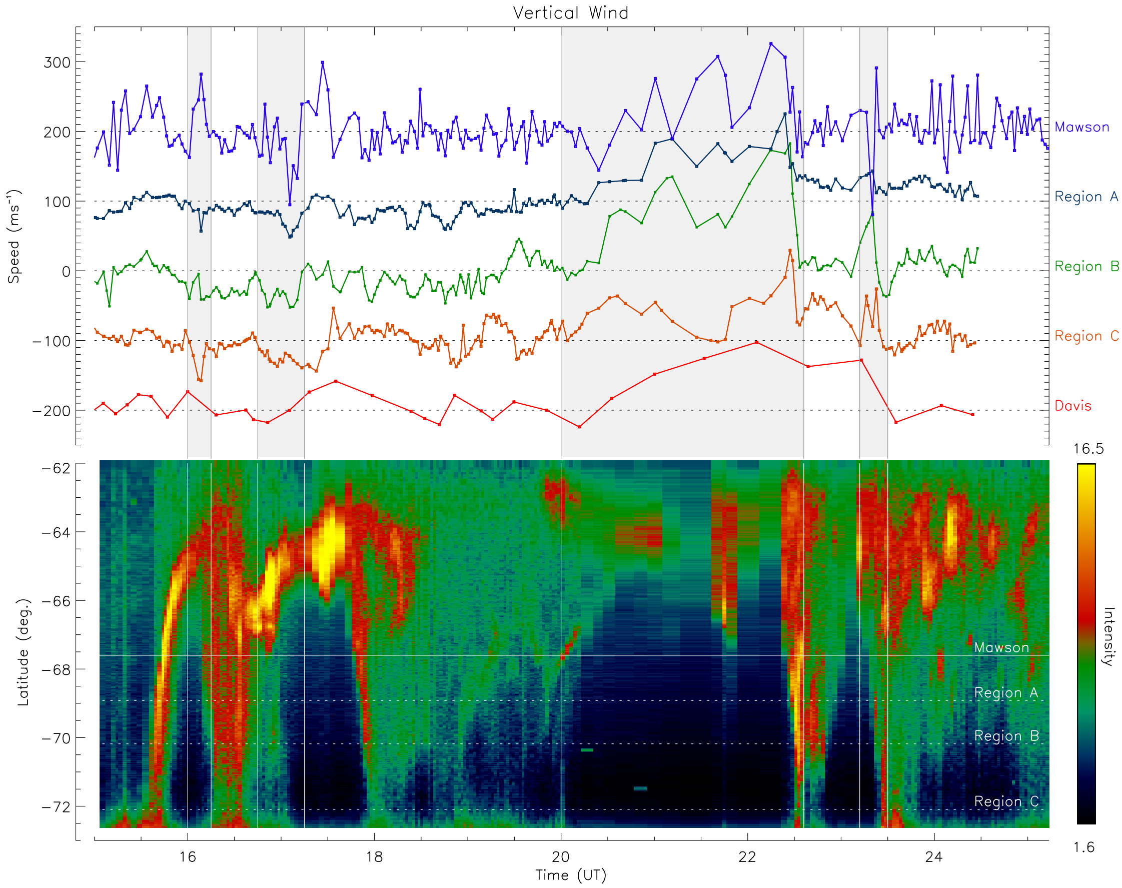


Figure 7.3.9: Top panel: Vertical winds observed above each station and derived in regions A (blue), B (green) and C (red), for day number 96 (error-bars omitted). Lower panel: Keogram derived from the Mawson all-sky images (explained in the text). The approximate latitudes of the common-volume regions and of Mawson station are indicated by horizontal lines. The intensity scale is indicated by the color bar on the right (arbitrary units).

upwelling in region C. In region A the trend was again similar, however the upwelling was more moderate, reaching an initial speed of 90 m.s^{-1} before it decreased to 50 m.s^{-1}, followed by an increase to 120 m.s^{-1}. The upwelling ended abruptly at around 22:30 UT, however the trend was similar to that observed in region C, with a sharp decrease to $\sim$ 50 m.s^{-1} followed by a more gradual subsequent decrease. As with regions B and C a smaller upwelling was then observed between 23:10 and 23:30 UT. This upwelling was weaker than those observed in the other common-volume regions, reaching a maximum speed of only 40 m.s^{-1}.

The vertical wind above Mawson during this period showed quite different behaviour. As the upwelling was beginning above Davis and in each of the common-volume regions Mawson observed a 50 m.s^{-1} downwelling, which was followed by a series of progressively stronger upwellings, the largest reaching 125 m.s^{-1}. As with the common-volume regions the last upwelling ended suddenly at 22:30 UT in a manner similar to region B. Between 23:10 and 23:30 UT, as regions A, B and C observed upwellings, Mawson observed a strong (120 m.s^{-1}) downwelling which lasted $<$ 5 minutes followed by an equally rapid upwelling of 90 m.s^{-1}.

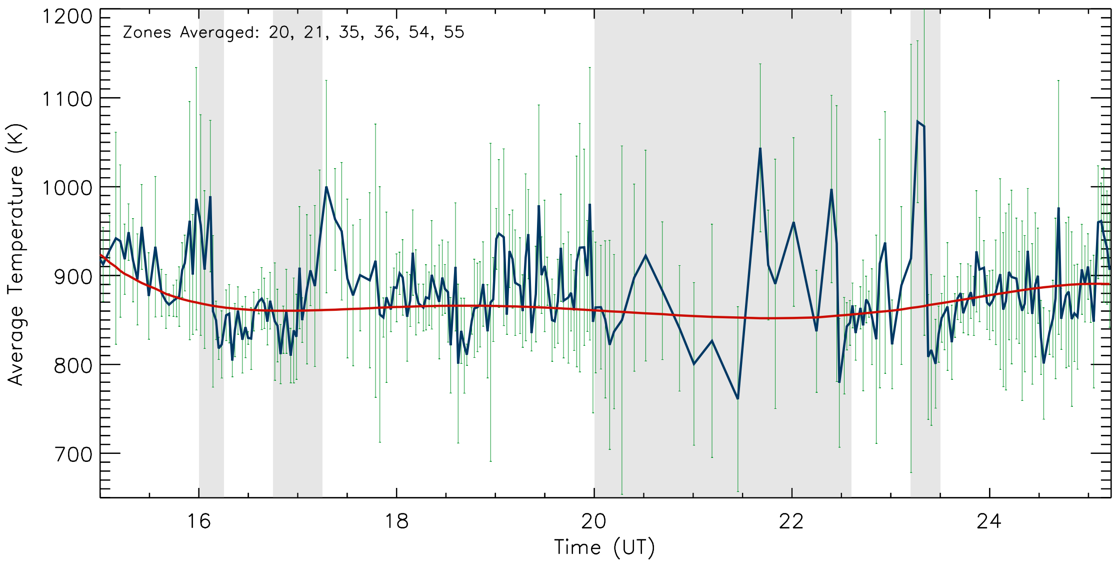


Figure 7.3.10: Average temperature (blue line) in zones 20, 21, 35, 36, 54 and 55 from Mawson, on April 5, 2008 (day 96). Error-bars are the 1σ deviations of the temperature estimates from these zones. The red line is a 5^{th}-order polynomial fit to the mean temperature across all zones, and represents the background temperature variation throughout the night. The periods shaded in grey correspond to the same periods which were highlighted in Figure 7.3.9 (and in text).

From the keogram in Figure 7.3.9 there were some interesting correspondences between changes in the auroral structure and the observed vertical winds. The period of initial upwelling above Davis and in the common-volume regions between 20:00 and 20:30 UT (which was a downwelling above Mawson) coincided with the contraction of the auroral arc to a more latitudinally confined area and a dimming of the auroral brightness above Davis and the common-volume regions. The region above Mawson, being closest to the arc, was the last to observe the dimming, and this may be related to the observation of a downwelling above Mawson at this time in contrast to the upwellings observed in all other regions. The decrease in upward wind speed which followed the initial upwelling in each of the common-volume zones coincided with a noticeable reduction in the brightness of the auroral arc, such that between two exposures the arc almost completely disappeared. The subsequent strengthening of the upwelling in each of the common-volume zones was accompanied by an equally sudden brightening and subsequent dimming of the aurora in the northern half of the field-of-view. The very fast decrease in upward wind speed which was observed above Mawson and also in the common-volume regions occurred as the aurora to the (geomagnetic) north increased rapidly in brightness and expanded to fill the entire field-of-view, before dimming once again as the upwellings (in all regions except above Davis) subsided. The final disturbance which appeared in the common-volume regions as an upwelling lasting $\sim$ 20 minutes was again observed concurrently with a sudden brightening of the aurora in the sky equatorward of Mawson.

From Figure 7.3.7 it is evident that the uncertainties in the derived vertical winds were large during the relatively long period of upwelling between 20:00 and 22:30 UT. This was due to the low auroral intensity, which had the effect of decreasing the signal-to-noise ratio of the observed spectra, and hence increased the statistical uncertainty (in this case the 1σ deviations) of the spectral fits. However this does not mean that the vertical winds observed during this period are to be mistrusted. Above Mawson, for example, during this period the large (> 70 m.s^{-1}) vertical wind speeds were at least 6 times greater than the statistical uncertainty in the spectral fits. In the common-volume regions where the uncertainties were in general greater, these large winds were at least greater than 3σ. Also, while the uncertainties were large, the consistency between adjacent data points was good. Innis et al. (1996) presented similar arguments for the winds they observed above Mawson during three 'zenith direction only' observing campaigns in 1993, using an earlier-generation Fabry-Perot spectrometer. In that study Innis et al. (1996) observed large vertical winds during periods of very low intensity, and identified those low intensity periods with Mawson being underneath the polar cap region, as they were often observed concurrently with higher temperatures.

In the present study, the absolute uncertainties in the temperature data during the low intensity periods on the night of April 5, 2008, were often very large relative to the variations in the temperature estimates. However, the correlation between the temperature estimates from neighbouring zones was generally very good, and since these were independent, concurrent estimates of the temperature within a relatively small region of the sky, it is reasonable to average the temperature estimates from nearby zones and use the standard deviation of the temperatures from the contributing zones to estimate the uncertainty. Figure 7.3.10 shows the results of averaging the temperature estimates in this manner. In this case, temperature estimates from zones 20, 21, 35, 36, 54 and 55 (in the vicinity of regions B and C, see Figure 7.3.11) were averaged, and the standard deviation of those temperature estimates in each exposure used to measure the uncertainty. The background temperature variation is indicated by the red line, which is a 5^{th}-order polynomial fit to the mean temperature across all zones.

As can be seen from this figure, there were clear temperature increases associated with the long period of upwelling between 20:00 and 22:30 UT. However, these increases (which were between 100 and 200 K) were only observed during the second half of the upwelling period. During the first half of the period a decrease in average temperature was observed. The temperature increases were in general larger than the standard deviation of the underlying temperature estimates over the 6 zones. The end of the upwelling in the common-volume regions coincided with temperatures returning to their background values. Not long after this, a sharp temperature spike was observed at the same time as the brief upwelling was observed in the common-volume regions (predominantly in regions B and C) between 23:10 and 23:30 UT. The early disturbance observed in all regions between 16:00-16:10 UT coincided with a temperature increase of around 100 K above the background level, although the 1σ temperature uncertainties were large. Similarly, the upwelling which began above Davis at 17:00 UT (and coincided with downwellings in all other regions) occurred as the temperature in the common-volume region was rising to 140 K above the background temperature.

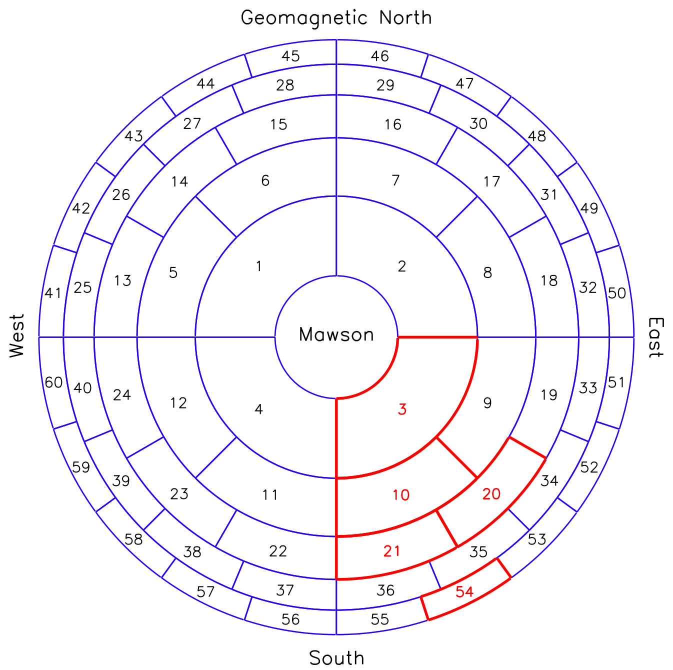


Figure 7.3.11: The zone map used during the bistatic campaign, with zones contributing to the common-volume regions outlined in red.

7.3.3 April 7, 2008 - Day 98

April 7, 2008 was also a relatively quiet day in terms of geomagnetic activity, with *ap* reaching a maximum of 22 during the night, while the IMF was predominantly northward. The aurora was however quite active on this night, as were the derived common-volume vertical winds. The line-of-sight wind estimates, derived vertical winds, and normalised intensities are shown in Figures 7.3.12 - 7.3.14. Multiple vertical wind disturbances were observed on this day. The first began at approximately 16:50 UT, when a strong (150 m.s^{-1}) downwelling was observed in region C, concurrent with a 30 m.s^{-1} downwelling above Davis, and 50 m.s^{-1} upwellings above Mawson and in region A. Region B observed no disturbance. Prior to this event, the all-sky images showed the aurora moving equatorward overhead, as can be seen from the northward track of the high intensity region in Figure 7.3.15 (lower panel). At the time the upwelling/downwelling began, the auroral arc had reached its furthest northward position, while diffuse aurora filled most of the field-of-view. By 17:00 UT, at the peak of the disturbance, the equatorward arc suddenly disappeared and was replaced by an arc on the poleward edge of the field-of-view.

Approximately 15 minutes after the strong downwelling in region C, the direction of the vertical winds in each of the common-volume regions reversed: the region C vertical wind became positive, while the region A and Mawson vertical wind became negative (region B also registered a downwelling at this time), each region reaching maximum speeds in the range 50-70 m.s^{-1}. A positive vertical wind was also observed above Davis concurrently with the upwelling in region C. The wind direction reversal occurred at approximately the same time for each zone (within $\sim$10 minutes). There was then a gradual return to lower

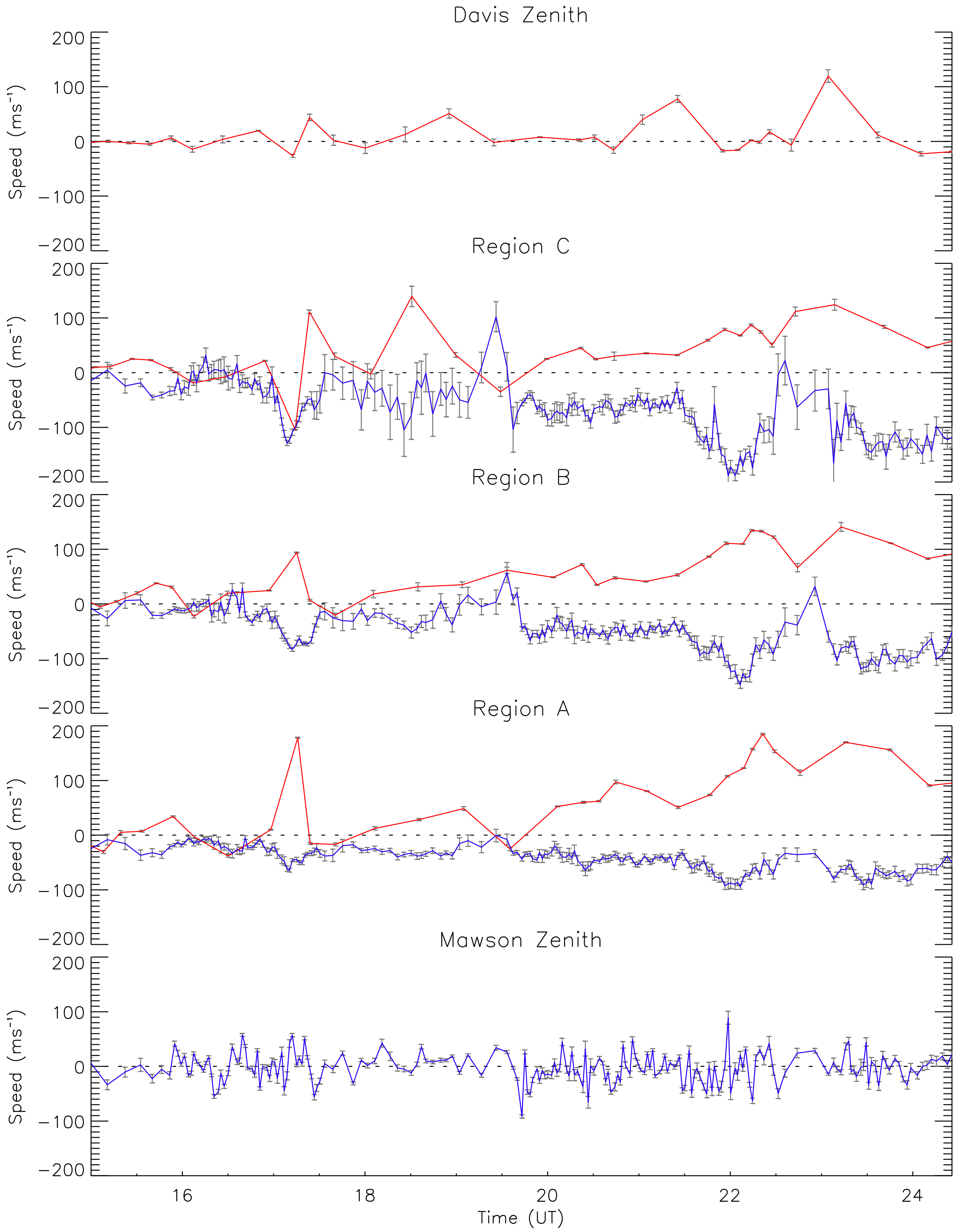


Figure 7.3.12: Line-of-sight wind estimates from Mawson (blue) and Davis (red) in each of the common-volume regions and above each station, for April 7, 2008 (day 98).

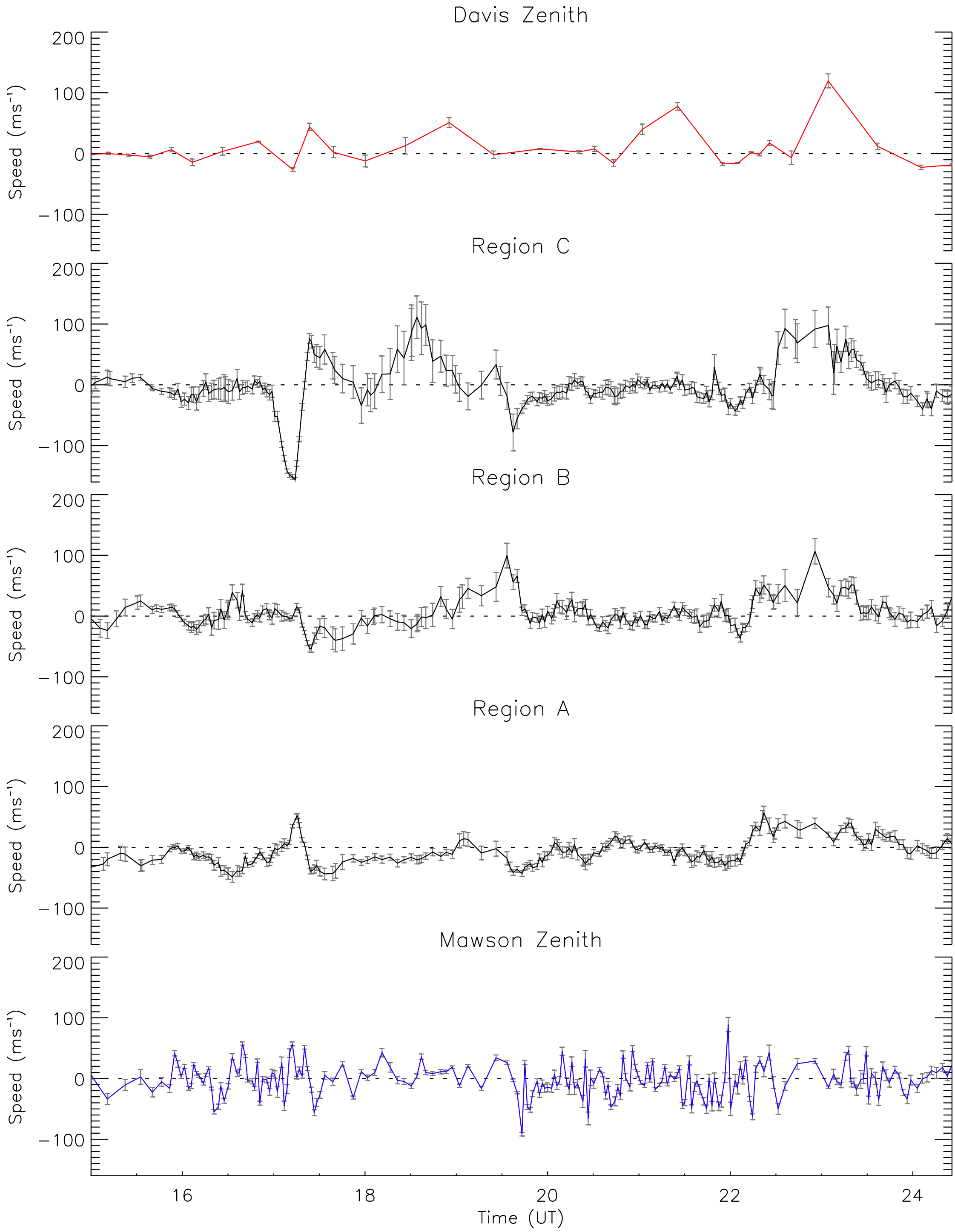


Figure 7.3.13: Derived vertical in each of the common-volume regions, and vertical winds recorded above each station, for April 7, 2008 (day 98).

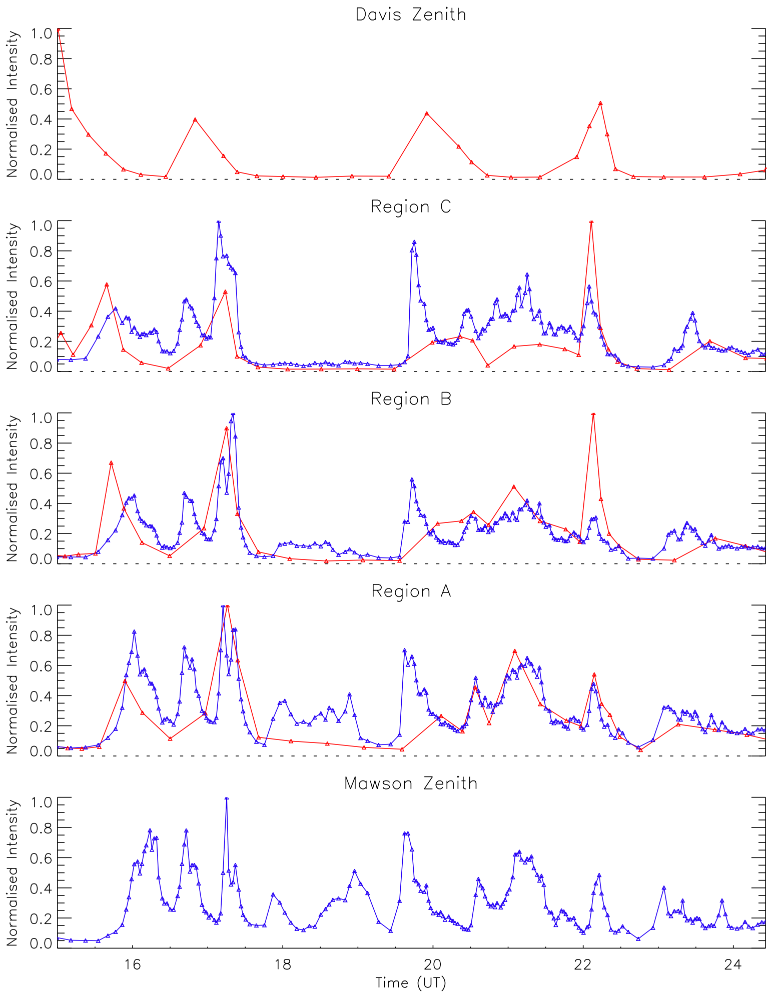


Figure 7.3.14: Normalised intensities observed in each of the common-volume regions, and above each station, for April 7, 2008 (day 98).

absolute wind speed in each region over 40-50 minutes, the vertical winds having abated by 18:00 UT. During this period of wind direction reversal and subsequent abatement (17:20-18:00 UT) an auroral arc that was poleward of Mawson moved rapidly equatorward across the field-of-view. By the time the strong vertical winds had abated, this bright arc had been replaced by a much dimmer arc equatorward of Mawson. The motion of the aurora into and out of the common-volume regions during this period between 16:00 and 18:00 UT was also clearly tracked by the intensities plotted in Figure 7.3.14.

The second period of interest occurred between 18:10 and 19:45 UT, and was marked by three localised upwellings, separated both spatially and temporally. The first began around 18:10 UT in region C. A maximum upward wind speed of 110 m.s^{-1} was reached, with the whole upwelling period lasting approximately 50 minutes. A moderate upwelling of 50 m.s^{-1} peak speed was observed above Davis, beginning at the same time as the event in region C, but lasting approximately 20 minutes longer and peaking some 15 minutes after the peak speed was reached in region C. The upwelling observed in region C appeared to be followed by a second smaller upwelling which peaked at 19:30 UT with a speed of

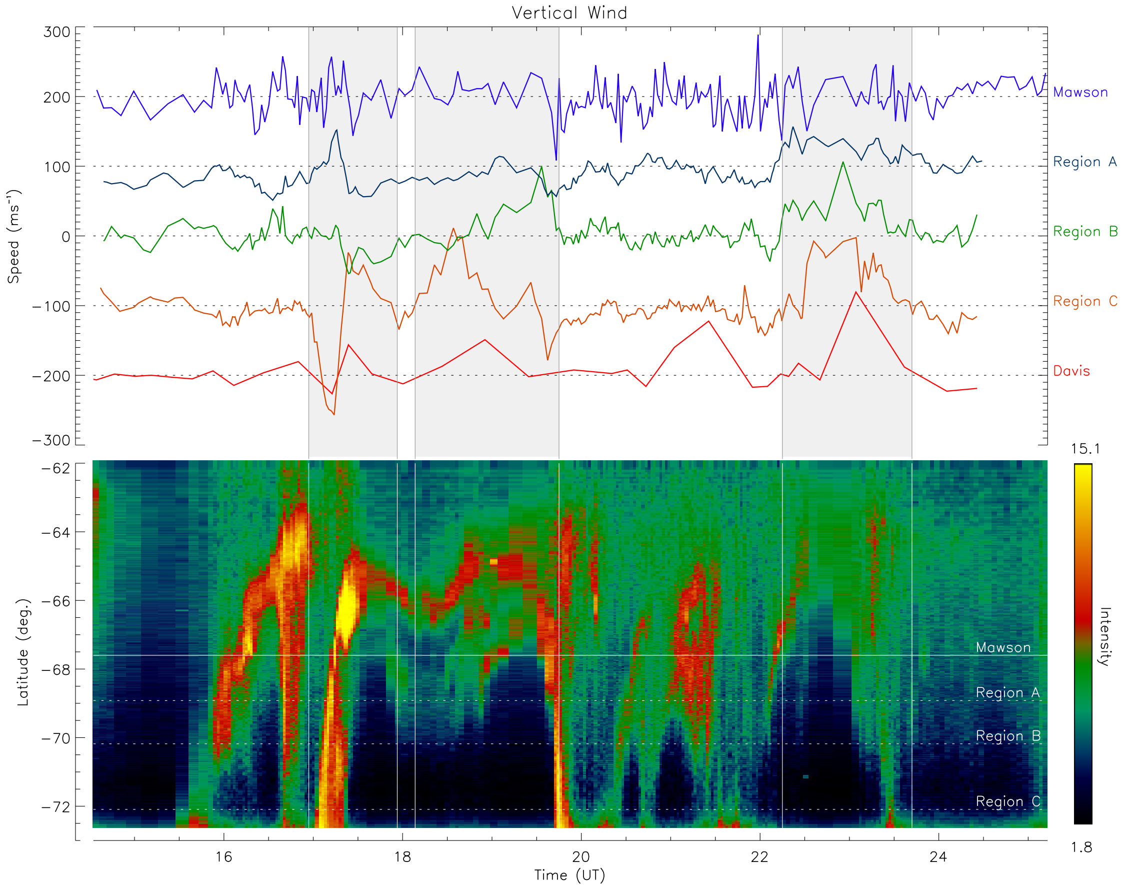


Figure 7.3.15: Top panel: Vertical winds observed above each station and derived in regions A (blue), B (green) and C (red), for day number 98 (error-bars omitted). Lower panel: Keogram derived from the Mawson all-sky images (explained in the text). The approximate latitudes of the common-volume regions and of Mawson station are indicated by horizontal lines. The intensity scale is indicated by the color bar on the right (arbitrary units).

30 m.s^{-1}, and lasted 10-15 minutes, itself followed by an 80 m.s^{-1} downwelling.

Following the first upwelling in region C, an upwelling was observed in region B beginning at approximately 19:00 UT and reaching a peak upward wind speed of 100 m.s^{-1}. This upwelling also lasted approximately 50 minutes. No significant upwelling was observed in region A during the period 18:00-19:45 UT, however a 40 m.s^{-1} downwelling was observed at approximately 19:40 UT, coincident with the 80 m.s^{-1} downwelling in region C and a similar strong downwelling above Mawson. Figure 7.3.14 showed that during this period between 18:10-19:45 UT the intensity recorded in each region was quite low, especially above Davis and in regions B and C, due to the aurora being positioned equatorward of Mawson. The keogram (Figure 7.3.15) showed the aurora moving equatorward at the peak of the region C upwelling, then reversing and moving poleward again as the region B upwelling peaked.

Following the events described above there was a period of approximately 1½ hours where vertical winds in the common-volume regions remained small. An 80 m.s^{-1} upwelling was observed above Davis at approximately 21:30 UT, however this event was confined to a region extending no more than $\simeq$ 160 km in the direction of Mawson, as

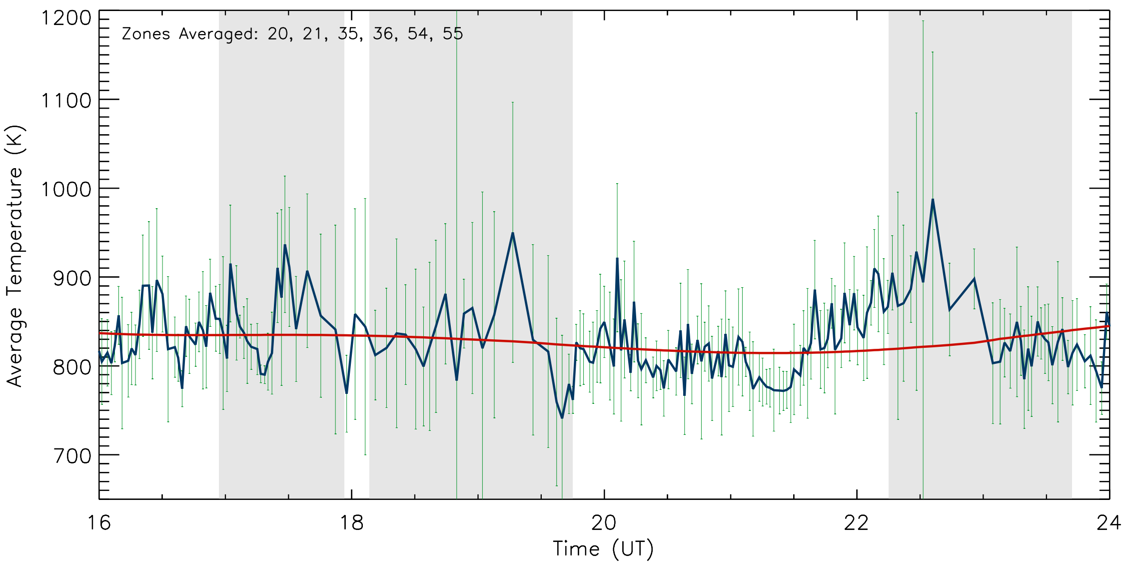


Figure 7.3.16: Average temperature (blue line) in zones 20, 21, 35, 36, 54 and 55 from Mawson, on April 7, 2008 (day 98). Error-bars are the 1σ deviations of the temperature estimates from these zones. The red line is a 5^{th}-order polynomial fit to the mean temperature across all zones, and represents the background temperature variation throughout the night. The periods shaded in grey correspond to the same periods which were highlighted in Figure 7.3.15 (and in text).

no similar upwelling was observed in region C. The third period of interest began at approximately 22:15 UT, when another strong upwelling began in regions B and C, with a smaller magnitude upwelling observed in region A. An upwelling above Davis began approximately 15 minutes later. Upward winds reached speeds of 120 m.s^{-1} above Davis, and $\simeq$ 100 m.s^{-1} in regions C and B. The maximum upward wind recorded in region A was 50 m.s^{-1}, while no sustained upwelling was observed above Mawson.

These observations suggest a large-scale upwelling event, which decreased in magnitude away from Davis. The upwelling lasted for approximately one hour, during which time the aurora was equatorward of the common-volume regions. This upwelling was very similar to that observed on April 5, 2008, as it was a relatively long-duration, large-amplitude upwelling associated with low intensities and dim, equatorward aurora. However in this instance the vertical wind above Mawson showed no significant upwelling, which may be related to the aurora being closer to Mawson on this night compared with the night of April 5, 2008. In Figure 7.3.16 are plotted the average temperature from the zones in the vicinity of the common-volume regions as was described in Section 7.3.2 (and Figure 7.3.10). Again, the red line in this figure represents the background temperature variation throughout the night (across all zones). There was a clear increase in temperature starting approximately 21:30 UT. The temperature increase peaked as the upwelling was beginning, at which time temperatures were approximately 160 K above the background level.

7.4 Discussion

The vertical winds which have been presented in this chapter support many of the findings of previous investigators. For example, it is clear that upwellings of duration $\gtrsim$ 15 minutes were most commonly observed poleward of the poleward edge of the auroral oval, in agreement with previous studies of vertical winds above Mawson (Innis et al., 1996, 1999), and elsewhere (e.g. Price et al., 1995; Ishii et al., 2001). Greet et al. (2002) presented results of hourly mean vertical winds above Davis and Mawson over 4 years, and related these data to the nominal position of the auroral oval. When a station was on the poleward edge of this nominal auroral oval, these authors observed average downward winds in the early magnetic evening and upward winds near magnetic midnight. In addition, vertical winds observed on the poleward edge of the auroral oval were of larger amplitude on average than those observed underneath the oval. From the data presented here, large upwellings (lasting for 1-2 hours) were observed on days 96 and 98 between 20:00 and 23:30 UT. Magnetic midnight at Davis is at approximately 22:00 UT, and at Mawson 22:40 UT. Thus these upwellings were both observed in the magnetic midnight sector.

For each of the three data sets presented in this chapter, the presence of bright aurora somewhere within the Mawson field-of-view heralded often significant vertical wind activity. In Chapter 6 (Section 6.2) it was noted that large magnitude, highly variable vertical wind activity overhead of Mawson was often coincident with periods of increased intensity, due to the presence of bright aurora within the field-of-view. That observation has also been shown to hold true for vertical winds in the common-volume regions between Mawson and Davis. The only events which did not appear to be associated with bright aurora were the large-scale, long-duration upwellings which were observed between 20:00 and 22:30 UT on day 96 and between 22:15 and 23:30 UT on day 98. Other than these two periods, significant vertical wind activity in the common-volume regions was observed during periods of increased auroral brightness and/or latitudinal motion of the aurora on the days presented here.

In order to establish if there is a real link between the presence of aurora and an unusually active vertical field, consider Figure 7.4.1. In the top panels of this figure have been plotted the median of the absolute vertical wind speed observed under different levels of signal intensity (red line). Error-bars represent the standard error within each intensity bin, while the grey bars show the median *ap* within each bin. The middle and lower panels show respectively the median signal-to-noise ratio and the number of data points within each intensity bin. The data for two years of Mawson data, 2007 and 2008, are shown in the left column. Five years of Davis data (2004-2008, inclusive) are shown in the right column. Intensities are the uncalibrated spectral areas returned by the spectral fitting algorithm, divided by the exposure time. For Mawson an intensity bin width of 80 was used, for Davis the width was 0.3.

The assumption that is made here is that elevated intensities correspond to periods where the aurora was within the field-of-view. The correlation between increased intensity and the level of geomagnetic activity (as indicated by the *ap* index) is generally clear, and

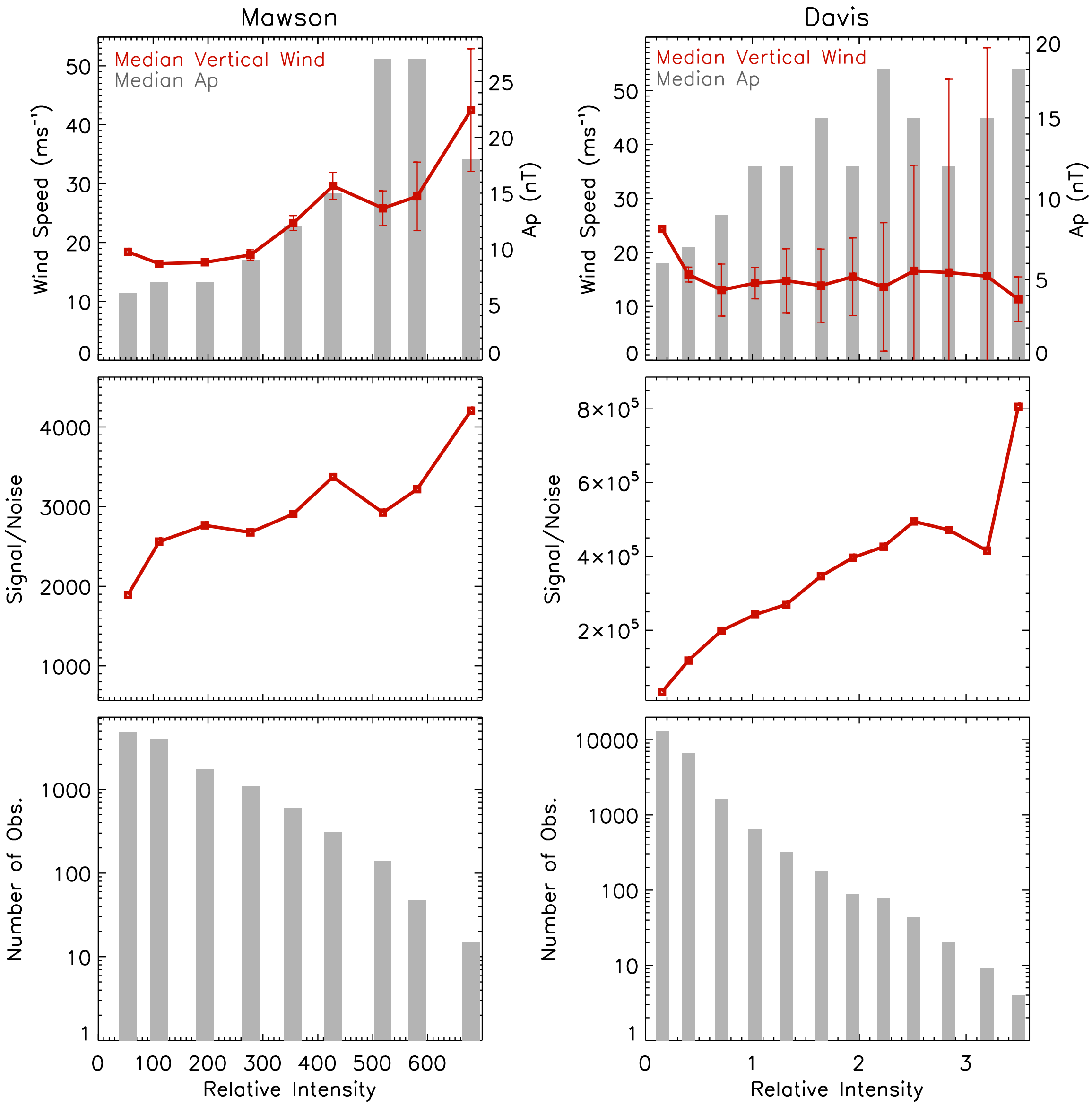


Figure 7.4.1: Median vertical wind magnitudes sorted by intensity level for two years of Mawson data (2007-2008, left column) and five years of Davis data (2004-2008, right column). Top panel shows (red line) the median of the absolute vertical wind speed within intensity bins of width 80 (Mawson) and 0.3 (Davis), along with the median *ap* level within each intensity bin (grey bars). Error-bars are the standard errors (mean absolute deviation about the median divided by the square root of the number of data points). The middle panel shows the median signal-to-noise ratio within each intensity bin, while the lower panel shows the number of observations in each bin.

supports this assumption. The signal-to-noise ratio was higher under levels of increased intensity, as expected, while the number of observations contributing to each intensity bin decreased approximately exponentially with increasing intensity, which is also as expected from the geomagnetically very quiet data set (see Section 6.3).

At Mawson, there was a clear trend toward increased vertical wind speeds as the auroral brightness increased. The median vertical wind at an intensity of 680 was approximately 40 m.s^{-1}, or twice the speed at an intensity of 54. The vertical wind variability was also higher during periods of increased intensity, as indicated by the larger error-bars. At Davis, no trend toward increasing vertical wind speeds was observed. The lack of calibrated intensities makes it difficult to compare the data from the two stations, as it is possible the Davis data does not cover the same range of intensities (in an absolute sense) as does the Mawson data set. In fact, this situation is likely, since for much of the time Davis lies inside the poleward edge of the auroral oval, whereas Mawson usually lies equatorward of or underneath the auroral oval. Particularly as geomagnetic activity increases, and the auroral oval (on average) expands equatorward, Davis is unlikely to observe the aurora and the high emission intensities associated with it. It is therefore possible, that were a larger range of intensities available from Davis, a similar trend to that observed above Mawson would emerge. However, given that Davis spends such a large fraction of the time inside the auroral oval (that is, in the region of the polar cap, which is bounded by the poleward edge of the auroral oval), it is very likely that the different trends revealed in Figure 7.4.1 do in fact reflect real differences in the vertical wind behaviour at the two stations, particularly in light of the vertical wind activity maps generated by Innis and Conde (2002) (discussed below).

The low intensity median winds observed above Mawson were of comparable magnitude to those observed above Davis, however the variability in the vertical winds was much greater above Davis. The greater variability of the vertical winds at the polar cap station is consistent with the study of Innis and Conde (2002), who inferred the spatial distribution of vertical wind variability from DE-2 vertical wind data. The southern and northern hemisphere maps they produced from all the derived variabilities (not sorted by height, solar zenith angle or AE index) are shown in Figure 7.4.2. For these data the region of increased vertical wind activity was quite clearly confined to the region bounded by the (statistical) auroral oval, or in other words the polar cap region. This suggests that a possible explanation for the larger vertical wind variability observed at Davis is due to that station's location within the polar cap.

It is also evident from the data presented here that vertical winds can be correlated on a range of horizontal scales. Kosch et al. (2000), in a study of vertical winds from Norway, observed essentially no correlation between vertical winds observed from two stations which were separated longitudinally by only 45 km. This lack of correlation applied to both E and F-region vertical winds. Ishii et al. (2004) on the other hand observed very good correlation between E-region vertical winds over an approximately 300 km baseline, however poorer correlation was observed in the F-region vertical winds. The low correlation cases indicated that the aurora was in a different location with respect to the two observatories (which were at similar geomagnetic latitudes). Ishii et al. (2004) interpreted their results as suggesting that along an auroral arc the vertical wind structures were uniform, over a distance of at least 300 km.

In the present study, the smallest horizontal separation between vertical wind estimates was on the order of 160 km, while the largest separation was that between Davis and Mawson, approximately 635 km. In general the correlation between vertical winds above Davis and Mawson was very low, although it is difficult to assess this given the often large difference in time-resolution between the instruments at each station. The different geomagnetic locations of the two stations would also be expected to reduce the correlation: Mawson was usually equatorward of or underneath the auroral oval, while Davis was usually situated inside the region bounded by the auroral oval for most of the time. Poor correlation over shorter baselines was also evident at times during the nights presented in this chapter. For example, there was often little or no correlation between winds observed directly above Mawson and those derived in region A, the closest common-volume region to Mawson. Also consider day 98, when a strong 80 m.s^{-1} upwelling was observed above Davis between 20:30 and 21:40 UT, but which was not observed in region C, the closest common-volume region to Davis.

However the data also showed periods of good correlation (and anti-correlation) between vertical winds estimated in different regions. The long-duration upwellings observed near magnetic midnight on days 96 and 98 are examples of good correlation over baselines of $\sim$ 480-635 km. These upwellings were observed during periods of low intensity and elevated temperatures, when the aurora was dim and located equatorward of Mawson. On day 96 a vertical wind spike was observed simultaneously in the three common-volume regions (between 23:10 and 23:30 UT), while a sharp downward/upward wind event was observed above Mawson.

There were also periods when vertical wind events were observed only in adjacent zones, and were thus correlated over baselines of approximately 160 km. On day 99 an oscillation was observed in regions B and C between 17:30 and 19:30 UT, and later an upwelling was observed above Davis and in region C (peaking around 20:40 UT). This day also showed an anti-correlation between adjacent regions B and C at the end of the night, where region B observed an upwelling at the same time that region C observed a comparable downwelling. A second, stronger anti-correlation was observed on day 98, between 17:00 and 18:00 UT, when downwelling was observed above Davis and (very strongly) in region C concurrently with more moderate upwellings above Mawson and in region A. The anti-correlation continued as the direction of the vertical winds in each region reversed at approximately the same time ($\sim$17:20 UT), and gradually abated.

From this data we can conclude that vertical wind events are likely to be present on a range of scales, obviously depending on which mechanisms are driving the observed winds. Figure 7.1.1 provides some qualitative support for this conclusion. There are examples within these three plots of vertical wind structures which are present from $\sim 6^{\circ}$ of latitude (orbits 4426 and 7164) down to at least 1° (orbit 8257) of latitude, and probably less. The vertical winds reported in this chapter have so far been related to the auroral morphology observed with the Mawson instrument without any mention of the response of the horizontal wind field. A closer look at the relationship between vertical winds and horizontal motions, and in particular the divergence of the horizontal wind field, will be left for the following chapter, where it will be shown that, at least in the case of day 98 presented above, small-scale gradients in the horizontal wind field were indeed linked to the observed vertical motions.

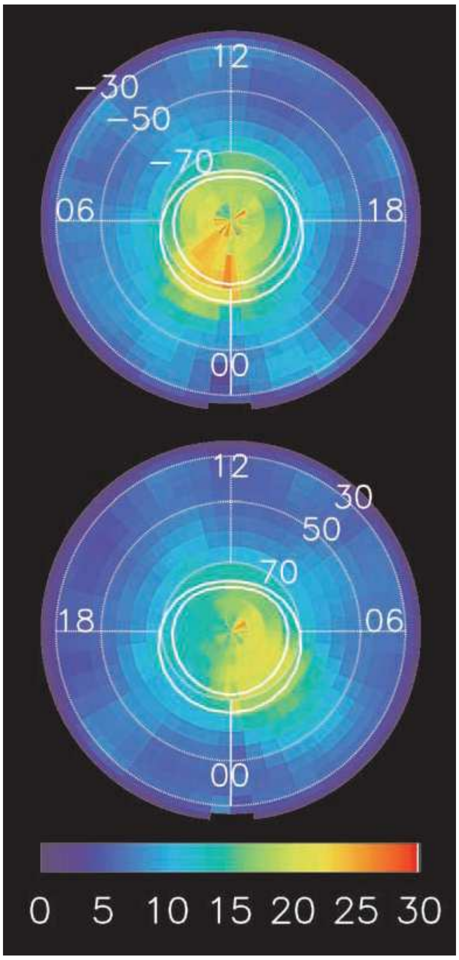


Figure 7.4.2: Distribution of vertical wind activity (defined as the standard deviation of the vertical wind measured within a sliding window equivalent to an along-track (orbit) distance of $\sim$ 900 km). Plots are shown for the southern (top) and northern (bottom) hemispheres. Coordinates are geomagnetic latitude and magnetic local time. From Innis and Conde (2002)

Chapter 8

Divergence and Local-Scale Structure

8.1 Introduction

In Section 2.4.2 it was pointed out that vertical winds must be linked to divergence/convergence of the horizontal wind field in order to conserve mass. Burnside et al. (1981) related the vertical wind (w) to the total horizontal divergence of the wind field using the scale height, H:

$$w = H\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right) \tag{8.1.1}$$

There are four important assumptions which must be satisfied for this relationship to hold, namely: that the time rate of change of pressure at the level of the airglow emission is negligibly small, the horizontal velocity components are constant with height, the atmosphere is isothermal above the level of the airglow, and horizontal gradients in atmospheric density can be ignored. In a recent modelling study, Cooper et al. (2009) demonstrated that the right-hand side of Equation 8.1.1 should in the real atmosphere be a reliable proxy for the vertical wind above the level of maximum (per unit mass) energy deposition, and during periods when the forcing does not change suddenly with time. At auroral latitudes the altitude of maximum energy deposition due to Joule and particle heating is in the E-region. As such, the findings of Cooper et al. (2009) would suggest that the Burnside relation should be applicable to observations of the 630.0 nm airglow, which has a peak emission altitude of around 240 km, at least during periods when the energy deposition is not changing rapidly with time.

Were this indeed the case, then the relation given in Equation 8.1.1 would be a particularly useful way of estimating vertical wind. In addition, comparisons between vertical wind predictions made using the Burnside relation and actual vertical wind measurements would give some insight into the dominant driving forces behind the vertical motions: in the case that the Burnside relation was a reliable predictor of vertical wind, then the likely processes driving those winds would be related to heating at a lower altitude. This chapter will present results of an investigation into the relationship between measured divergence and vertical winds above Davis and Mawson, on both large ($\sim$ 1000 km) and small ($\sim$ 100 km) scales.

8.2 Recurrent Upwelling above Davis

In Section 6.3.1 (and in particular Figure 6.3.3) data were presented which showed the average behaviour of vertical and horizontal winds above Davis for 'quiet' and 'active' geomagnetic conditions, using five years of data from the Davis FPS. An important feature of that data was an average upwelling of approximately 20 m.s^{-1} observed above Davis between 21:00 and 03:00 MLT. It was shown that the effect of increased geomagnetic activity was to confine the upwelling to a narrower time period, between approximately 00:00-04:00 MLT. For this upwelling to appear in a climatological average indicates that it must be a recurrent feature of the wind field above Davis, and indeed it has been observed previously. Greet et al. (2002) examined four years of vertical wind data (1997-2000) from the previous generation of Fabry-Perot spectrometers at Mawson and Davis. These authors observed an upwelling above Davis between approximately 22:00 and 03:00 MLT. The magnitude of the upwelling was in the range 10-20 m.s^{-1}, consistent with the observations reported here.

In order to estimate the horizontal divergence associated with the average upwelling observed in the present study, it was necessary to interpolate, for each day, the off-zenith winds observed in the four (geographic) cardinal directions to a set of common times, which for easy comparison were chosen to coincide with the times at which zenith observations were made for that day. The set of interpolated line-of-sight (horizontal) winds in each cardinal direction, along with the magnetic local time, *ap* index, cloud level and vertical wind were extracted from each night in the five-year data set (2004-2008). The divergence in each of the geographic zonal and meridional directions was calculated by finding the difference between the wind estimates from opposite cardinal directions, and dividing by the distance between those measurements (using the zenith angle and an assumed altitude of 240 km). Thus, for example, the zonal divergence was calculated by subtracting the west-looking estimate of zonal wind (positive eastward) from the east-looking estimate, and dividing by the distance between the two wind estimate locations. Similarly for the meridional divergence, using the north and south-looking estimates of the meridional wind. The total horizontal divergence was then calculated by simply summing the contributions from the two orthogonal directions.

From the final set of vertical and horizontal winds, divergence, *ap* index and cloud level (from all five years), data recorded during quiet and active conditions were separately extracted using the same criteria as in Section 6.3.1, namely $ap \leq 10$ (corresponding to quiet conditions) and $ap > 10$ (active conditions). In both cases observations for which cloud levels were $> 3/8$ were rejected as being too overcast. For each of the two ranges of activity, the vertical wind, horizontal wind and horizontal divergence were ordered by magnetic local time, and a median filter applied. The median filter replaced each value of a

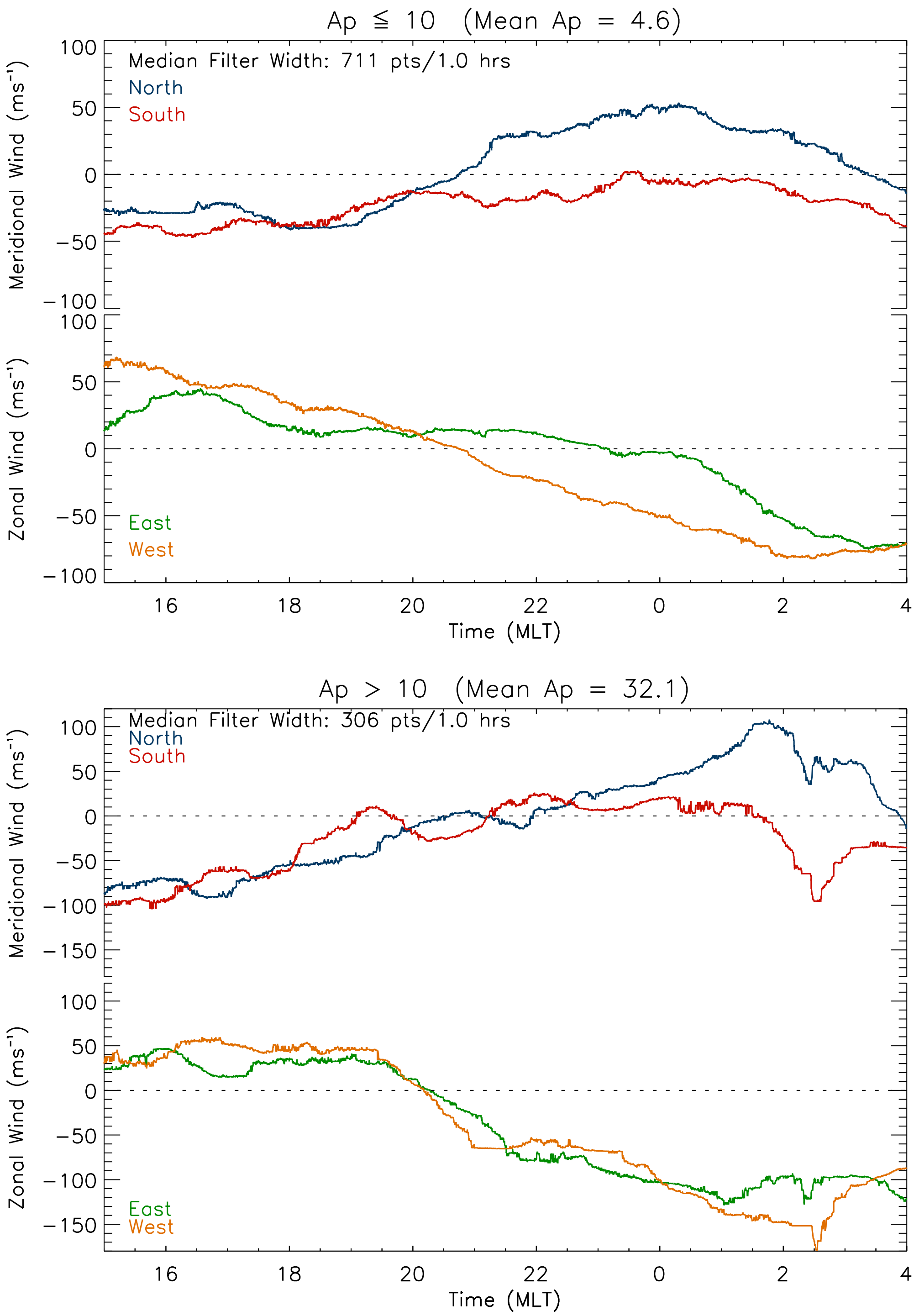


Figure 8.2.1: Average horizontal winds during quiet (top panel) and active (lower panel) conditions, from five years of Davis data (2004-2008). North and south-looking observations of the (geographic) meridional wind are shown in blue and red respectively, east and west-looking observations of the (geographic) zonal wind are shown by green and orange lines respectively. The width of the median filter (in units of data points and time) is indicated near the top of each plot; features of width less than the filter width are not significant. Note the different wind scale between quiet and active conditions.

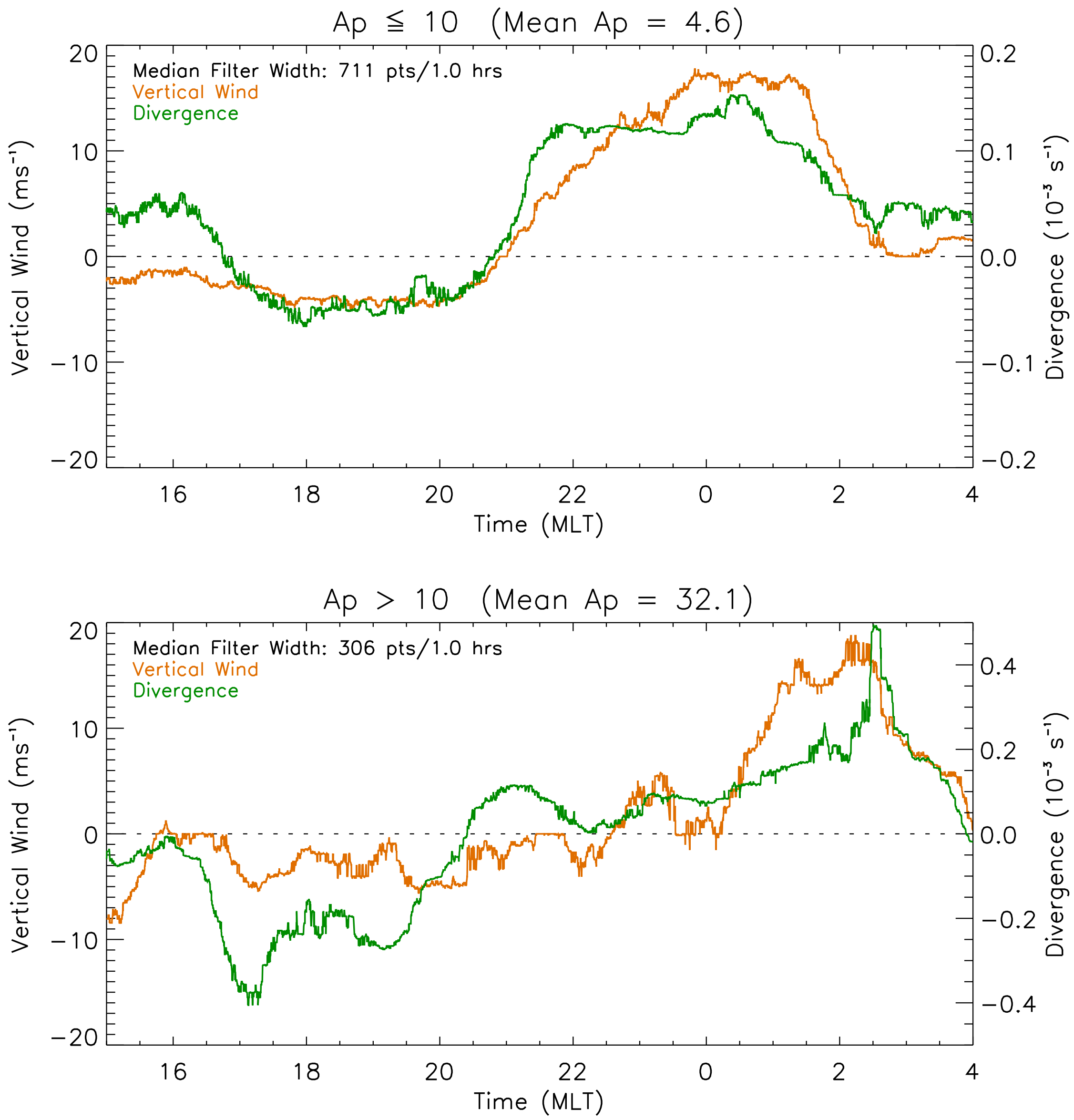


Figure 8.2.2: Average vertical wind and horizontal divergence during quiet (top panel) and active (lower panel) conditions, corresponding to the horizontal winds shown in Figure 8.2.1, from five years of Davis data (2004-2008). Vertical wind is plotted in orange, with the scale given by the left-hand-side y-axis, the divergence is plotted in green (in units of 10^{-3} s^{-1}), with the scale indicated by the right-hand-side y-axis. The width of the median filter (in units of data points and time) is indicated near the top of each plot; features of width less than the filter width are not significant. Note the different divergence scales between quiet and active conditions.

set of data (for example, vertical wind) with the median of the values within a given window centred on the point being replaced. The width of the window was chosen to correspond to approximately 1 hour of magnetic local time. Since the two ranges of activity contained a different number of data points, the width of the median filter's window corresponded to a different number of data points in order to maintain an approximately 1 hour-wide time window.

The result was thus an average of the vertical and horizontal winds and horizontal divergence under quiet and active conditions, using five years of data from Davis. Note that this approach is only slightly different to that outlined in Section 6.3.1, in that a median filter was used instead of the previous approach based on averaging over time bins of fixed width. The results are shown in Figures 8.2.1 and 8.2.2. The first of these figures shows the average horizontal wind observed in each cardinal direction, for quiet (top panel) and active conditions (lower panel). The width of the median filter in each case is indicated, in terms of both the number of data points and the approximate time window to which it corresponded. The average of the north and south-looking estimates of the meridional wind are shown in blue and red respectively, while the average of the east and west-looking estimates of the zonal wind are shown respectively in green and orange.

The vertical wind and divergence associated with these horizontal winds are plotted in Figure 8.2.2, quiet conditions in the top panel and active conditions in the lower panel. Vertical wind is plotted in orange, and the scale is indicated by the left-hand-side y-axis. The horizontal divergence (in units of 10^{-3} s^{-1}) is plotted in green, with the scale indicated by the right-hand-side y-axis. Note that while the vertical wind scale is identical for both activity levels, the active-time divergence scale covers 2.5 times the range of the quiet-time divergence scale. Also, since cardinal point observations from different years were in general made at different zenith angles, there is not always a simple correspondence between the difference in horizontal winds observed in opposite cardinal directions and the horizontal divergence, due to different separations between cardinal point look-directions[1].

It is clear from Figure 8.2.1 (top panel) that the period of upwelling between approximately 21:00 and 03:00 MLT corresponded to a period of relatively large gradients in the average horizontal wind field. For example, the average meridional flow observed to the south had a very small magnitude and was directed southward, while to the north the average flow was much stronger (around 50 m.s^{-1}) and directed northward. Similarly the average zonal wind flow during this time was westward, however the average wind observed to the west of the station was much stronger than that observed to the east. Thus the average divergence (Figure 8.2.2, upper panel) during this period was positive and relatively large (compared with the rest of the 'average' night), reaching a maximum of 1.5×10^{-4} s^{-1}. The variation in the average divergence during quiet conditions clearly correlated quite well with the average vertical wind, which reached speeds of 15-20 m.s^{-1}. There were differences however, as for example when the upwelling was increasing between 21:00

[1]Note that while the zenith angles were not always the same year-to-year (they differ by at most 25° between years) the divergence calculation takes horizontal separation into account. Thus this is only an issue when comparing horizontal wind and divergence by eye, and does not affect the vertical wind/divergence comparison.

and 22:00 MLT, the average divergence 'plateaued' at around $1.2-1.5\times10^{-4}$ s^{-1}. Also at the earliest and latest times the divergence was positive and relatively large ($\simeq 0.4\times10^{-4}$ s^{-1}) considering the average vertical wind was of very small magnitude (< 3 m.s^{-1}).

Average horizontal winds observed during active conditions showed more structure. Between approximately 23:00 and 03:00 MLT the winds observed in opposite cardinal directions showed a larger gradient, particularly in the geographic meridional direction. The average wind flow peaked at $\sim$ 100 m.s^{-1} northward to the geographic north of the station during this period, and very weakly northward becoming southward when observing to the south. Gradients in the average geographic zonal wind were smaller, and reversed from convergent flow between 22:00 MLT and magnetic midnight to divergent flow after magnetic midnight. As was observed during quiet conditions a significant average upwelling was associated with this period of average divergence (Figure 8.2.2, lower panel). Similar peak vertical wind speeds were reached during the active-time upwelling ($\sim 15-20$ m.s^{-1}), and similar or slightly higher values of divergence were observed during the upwelling ($1-2\times10^{-4}$ s^{-1}), with a narrow divergence peak of approximately 5×10^{-4} s^{-1} around 02:30 MLT.

The correspondence between the average vertical wind and horizontal divergence in the active data was not as clear as it was under quiet conditions. Divergence was in general larger under active conditions (the scale for the active-period divergence plot spanning twice the range of the quiet-period divergence plot), and showed more variability, although there could be a statistical bias due to the smaller number of active-time data, which reduced the number of data points within the median filter window. There was however a similar general trend observed in the average divergence (and vertical wind) between the two levels of activity, with a weak downwelling in the late afternoon associated with convergent horizontal winds, followed by a stronger upwelling later in the night, associated with divergent horizontal flow. Under active conditions however the convergent horizontal flow in the early afternoon (before 20:00 MLT) was much stronger relative to the observed downward vertical winds than it was under quiet conditions. Also the divergence associated with the upwelling after magnetic midnight was not as large as might be expected given the magnitude of the convergent flow compared to the relatively weak downwelling earlier in the night.

In Figure 8.2.3 the data from Figure 8.2.2 are shown in a scatter plot, with the (median filtered) vertical winds plotted as a function of the (median filtered) divergence for both quiet (upper panel) and active (lower panel) conditions. The individual data points have been given a colour based on the magnetic local time at which the (average) vertical wind/divergence pair were observed (the scale is indicated by the colour bar on the right). The straight line indicates the behaviour predicted by the Burnside relation (Equation 8.1.1), where the scale height used in this relation was calculated using the NRLMSISE-00 model for 00:00 UT and an *ap* level equal to the mean *ap* for each activity level (which is indicated at the top of each plot). For both activity levels, the data points were quite clearly clustered into the top-right and bottom-left quadrants, indicating that upward (positive) vertical winds were predominantly associated with positive divergence, while

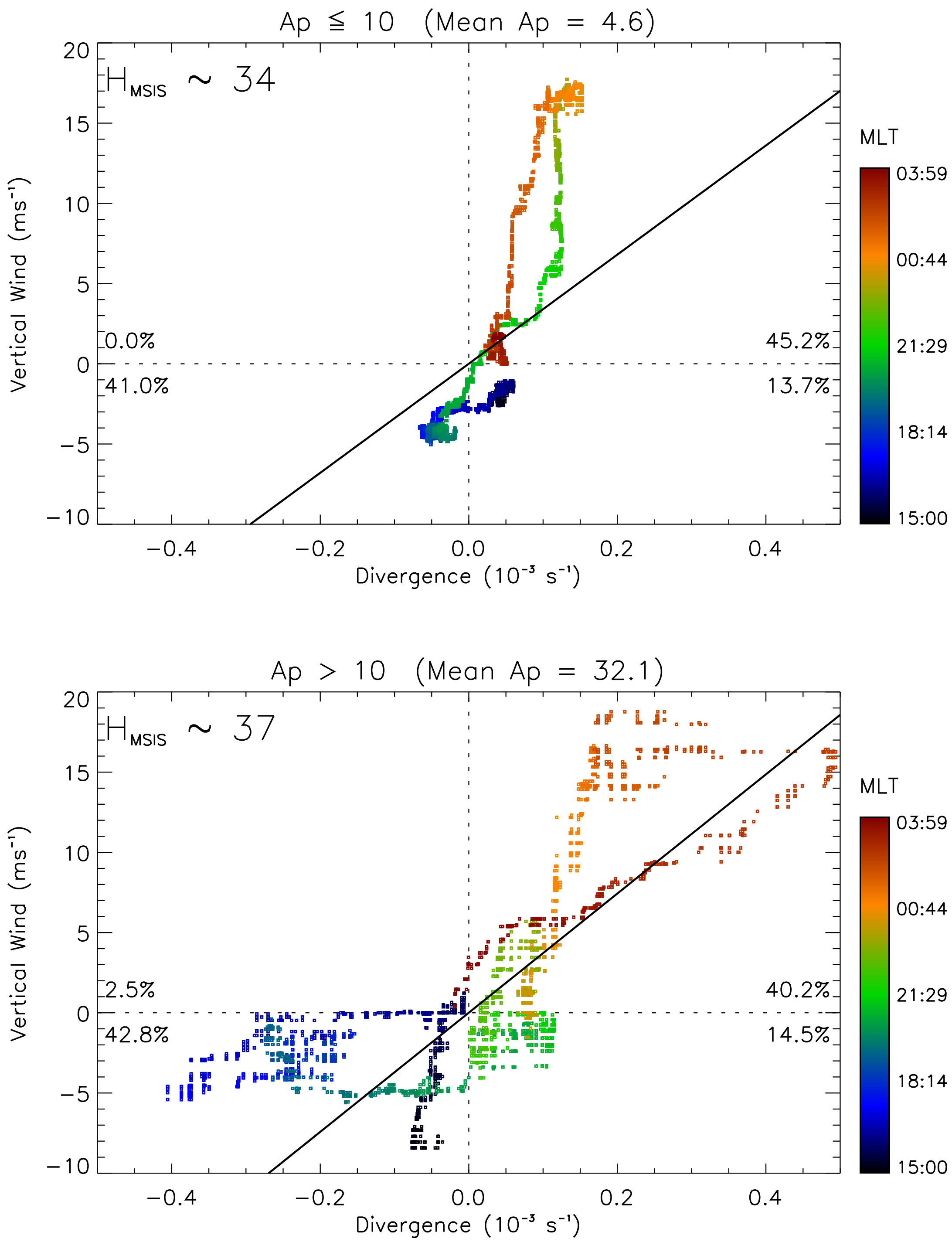


Figure 8.2.3: Scatter plot of the (median filtered) vertical wind as a function of (median filtered) divergence, from five years of Davis data. The color indicates magnetic local time, according to the scale bar on the right. The straight line represents the Burnside relation between vertical wind and horizontal divergence (Equation 8.1.1), where the scale height (H_{msis}) was calculated for 00:00 UT using the NRLMSISE-00 model, where the mean *ap* value (indicated on the top of each plot) was used as input to that model. Percentages indicate the fraction of data points lying within the given quadrant.

downward (negative) vertical winds were generally associated with negative divergence (convergence), as predicted by the Burnside relation. Indeed, over 80% of observations (for both levels of activity) fell into one of these two quadrants (the percentage of data points within a given quadrant is indicated in the corner of each quadrant). As seen in Figure 8.2.2, positive divergence was observed predominantly after $\sim$ 21:00 MLT, with convergence (and downward vertical winds) observed before this time on average.

The distribution of data points also showed significant differences between the two activity levels. Under quiet geomagnetic conditions, the data were confined to a relatively narrow band. Convergent average horizontal winds were associated with similar vertical winds independent of the time at which they were observed, while upward winds were associated with stronger divergence between 21:00 MLT and magnetic midnight, and relatively weaker divergence after this time. This can be seen in the different paths taken by the green and orange/red data points in the top-right quadrant. There were also a significant fraction ($\sim$ 14%) of observations of positive divergence associated with negative vertical winds in the early magnetic afternoon. From the location of the data points relative to the straight line predicted by Burnside, and assuming that the modelled scale height H_{msis} was reasonably accurate, it would appear that either the observed vertical winds were larger for a given value of divergence, or the horizontal divergence was smaller for a given vertical wind, than would be expected from the Burnside relation.

During active conditions there was a much greater spread in the data. The time-dependent splitting of data points in the top-right quadrant was much more pronounced under these conditions, although the sense of the split was opposite to that which was observed under quiet conditions. That is, stronger divergence was observed for a given vertical wind (or a weaker vertical wind for a given divergence) after magnetic midnight, while the converse was true before magnetic midnight. Some splitting was also evident in the early magnetic afternoon convergent winds, in the bottom-left quadrant, where negative divergence was observed associated with smaller (larger) than expected downward vertical winds in the early (late) magnetic afternoon. Again, a significant fraction ($\sim$ 14%) of active-time data was present in the lower-right quadrant, observed in the pre-magnetic midnight time sector. The percentage during active times was slightly higher than that during quiet times in this quadrant, but both indicated that downwellings were observed to be associated with divergent horizontal winds relatively frequently, which would not be expected from the Burnside relation. In comparison, the fraction of data appearing in the top-left quadrant was much less significant ($<$ 3% under active conditions and 0% under quiet conditions), thus on average convergent winds were only very rarely associated with upwellings (in accordance with the Burnside relation), suggesting that the heating was occurring far below the 630.0 nm airglow layer.

Considering Figure 6.2.9 (Section 6.2.3), the time period 21:00-03:00 MLT spans the region where the cross-polar jet would often be expected to emerge on the nightside, and where air parcels would be entrained to move eastward and westward by the convecting ions. Conde and Smith (1998) used an all-sky imaging Fabry-Perot spectrometer, similar to that used at Mawson in the present work, to investigate the spatial structure of the F-region wind field above Poker Flat, Alaska. These authors reported two frequently observed features of the wind field which were not predicted from generic runs of the NCAR-TIEGCM[2] model. The first was a period of very small wind speeds in the approximately one hour period prior to magnetic midnight (which the authors termed the 'doldrums'). The second feature was a wind shear which manifested as magnetically southeastward flow on the magnetic poleward side of the field-of-view and very small wind speeds of variable direction on the magnetic equatorward side. This shear, which swept across the field-of-view in the direction of the magnetic equator, was interpreted as the sharp equatorward edge of the cross-polar jet, which flowed magnetic southeastward, and was the dominant feature of the wind field after magnetic midnight. One possible explanation advanced to explain these features was an upwelling associated with auroral heating. Such an upwelling could act to decelerate the cross-polar jet (and thereby produce the observed wind shear) by advecting horizontal momentum (as lower altitude, slower horizontal wind flows are transported vertically upwards into the path of the cross-polar jet) or by producing meridionally divergent F-region winds which would flow poleward on the poleward edge of the region of auroral heating and equatorward on the equatorward edge.

It is possible that the recurrent upwelling and divergent horizontal winds observed above Davis are the signature of the station's regular passage underneath a region of heating similar to the type outlined above, although Davis is usually located poleward of the poleward edge of the auroral oval in this magnetic time sector. From the vertical wind activity maps of Innis and Conde (2002) (shown in Figure 7.4.2), the magnetic time sector between magnetic midnight and approximately 03:00 MLT is shown as having the highest levels of vertical wind standard deviation for magnetic latitudes poleward of and under the poleward edge of the nominal auroral oval. Thus the largest average vertical winds observed above Davis were seen to occur in the region of greatest vertical wind variability. During active periods at Davis, the upwelling was observed to span, on average, the magnetic local time sector between magnetic midnight and 04:00 MLT, in close agreement with the sector exhibiting greatest vertical wind activity in the analysis of Innis and Conde (2002). Interestingly, data from Mawson station, which is at a magnetic latitude 4° lower than Davis, and therefore often within the nominal auroral oval during the magnetic local times mentioned, have not shown an upwelling in the averages over two years of data, as for example in Figure 6.3.3. The maps of Innis and Conde (2002) were however averaged over 5° magnetic latitude bins, and thus the resolution is likely not sufficient to resolve the different features observed above Mawson and Davis.

8.3 Divergence and Vertical Wind above Mawson and Davis

In the previous section, the average upwelling observed above Davis was examined in relation to the average divergence of the horizontal wind field. Under quiet conditions

[2] See Table 2.3.

a clear relationship between the average vertical wind and the horizontal divergence was present, however that relationship was not well described by the Burnside relation given by Equation 8.1.1. In this section, a different approach is adopted to examine the vertical wind/horizontal divergence relationship and to compare that relationship between Mawson and Davis stations.

In the case of the Davis data, the procedure outlined in the previous section was used to extract sets of vertical wind and horizontal divergence from the five years (2004-2008) of data (where observations made under cloud levels of more than $3/8$ were rejected). However, the averaging process (sorting by magnetic local time and subsequent median filtering) was not performed in this instance. Instead, the data were sorted into 'quiet' and 'active' sets, based on prevailing geomagnetic conditions, where again these activity ranges were quantified by the *ap* index, with $ap \leq 10$ corresponding to quiet conditions and $ap > 10$ to active conditions. Each of these activity ranges were then further divided by magnetic local time into a 'pre-midnight' (11:00-24:00 MLT) and a 'post-midnight' (00:00-10:00 MLT) data set, and a third data set containing all magnetic local times for the given activity level. Within each of the resulting data sets, divergence bins of width 2×10^{-5} s^{-1} were created spanning the range $[-1, 1] \times 10^{-3}$ s^{-1}, and the median value of divergence and vertical wind within each divergence bin was calculated.

A similar procedure was performed for the Mawson data, however in this case the horizontal divergence was returned as part of the vector-fitting algorithm, and thus the divergence estimates were obtained simultaneously with vertical wind estimates, removing the need for interpolation. Two years of Mawson data were analysed (2007 and 2008), and cloudy data (cloud $> 3/8$) were rejected. The remaining data were sorted into activity levels and time ranges identical to those used for the Davis data. In the case of the Mawson data however it was found that the observed divergence covered a smaller range, and thus divergence bins of width 1×10^{-5} s^{-1} were created spanning the range $[-4, 4] \times 10^{-4}$ s^{-1}, and the median value of divergence and vertical wind within each divergence bin was calculated.

The results are shown in Figures 8.3.1 (Davis) and 8.3.2 (Mawson). In these figures, quiet conditions are shown in the left column, active conditions in the right column. The range of magnetic local times is shown above each plot, and the mean *ap* is indicated in the upper left corner. The error-bars in these figures are actually the ratio of the standard deviation of the vertical winds within a given divergence bin to the square-root of the number of data points within that bin, and are used to indicate relative uncertainty. In order to test the applicability of the Burnside relation to each of these data sets, a model curve of the form $y = mx$ was least-squares fitted to each set of data, and each data point given a weighting equal to the inverse of the error-bar value. The value of m returned by the fit, along with the 1σ uncertainty are shown in the lower right corner. Below these figures is shown the coefficient of determination (R^2) for the fit, giving an indication of the goodness of fit. In terms of the Burnside relation, the fitted value of m corresponds to a scale height, which has units of kilometers. This value has been used to judge the applicability of the Burnside relation, by comparing it to a modelled value of

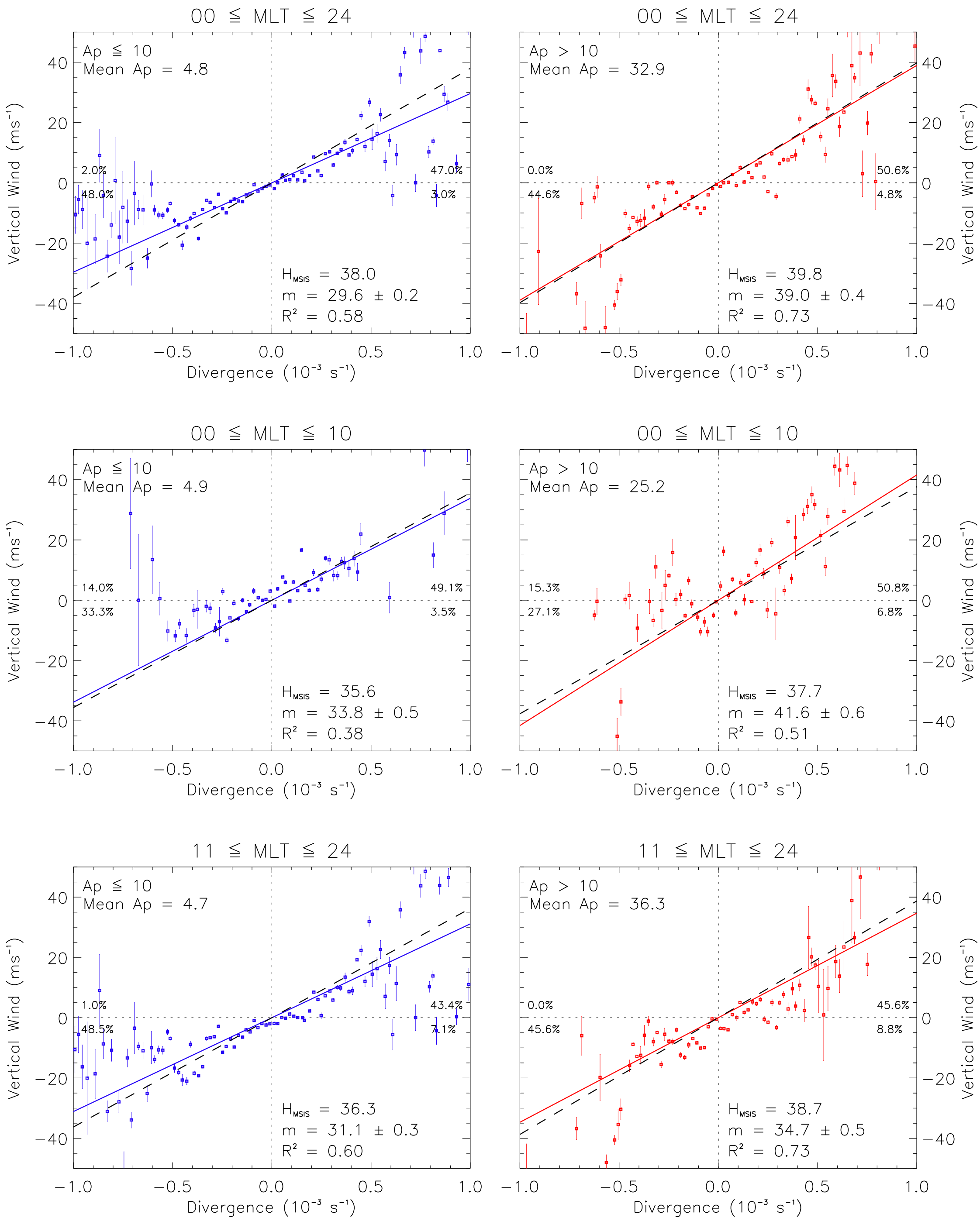


Figure 8.3.1: Scatter plot of vertical wind as a function of divergence, from five years of Davis data. The left-hand column shows data recorded during quiet conditions, the right-hand column during active conditions. Error-bars show the ratio of the standard deviation of the vertical winds within a given divergence bin to the square-root of the number of data points within that bin, indicating relative uncertainty. The magnetic local time range is indicated above each plot, and mean *ap* is shown in the top left corner. The solid straight line is a least-squares fit of the form $y = mx$, the dashed line is the Burnside relation for the given geomagnetic conditions. The scale height calculated from the NRLMSISE-00 model (and used to evaluate the Burnside relation) is shown in the lower right, below which are shown the value of m ($\pm 1\sigma$), and the coefficient of determination (R^2). Percentages indicate the fraction of data points lying within the given quadrant.

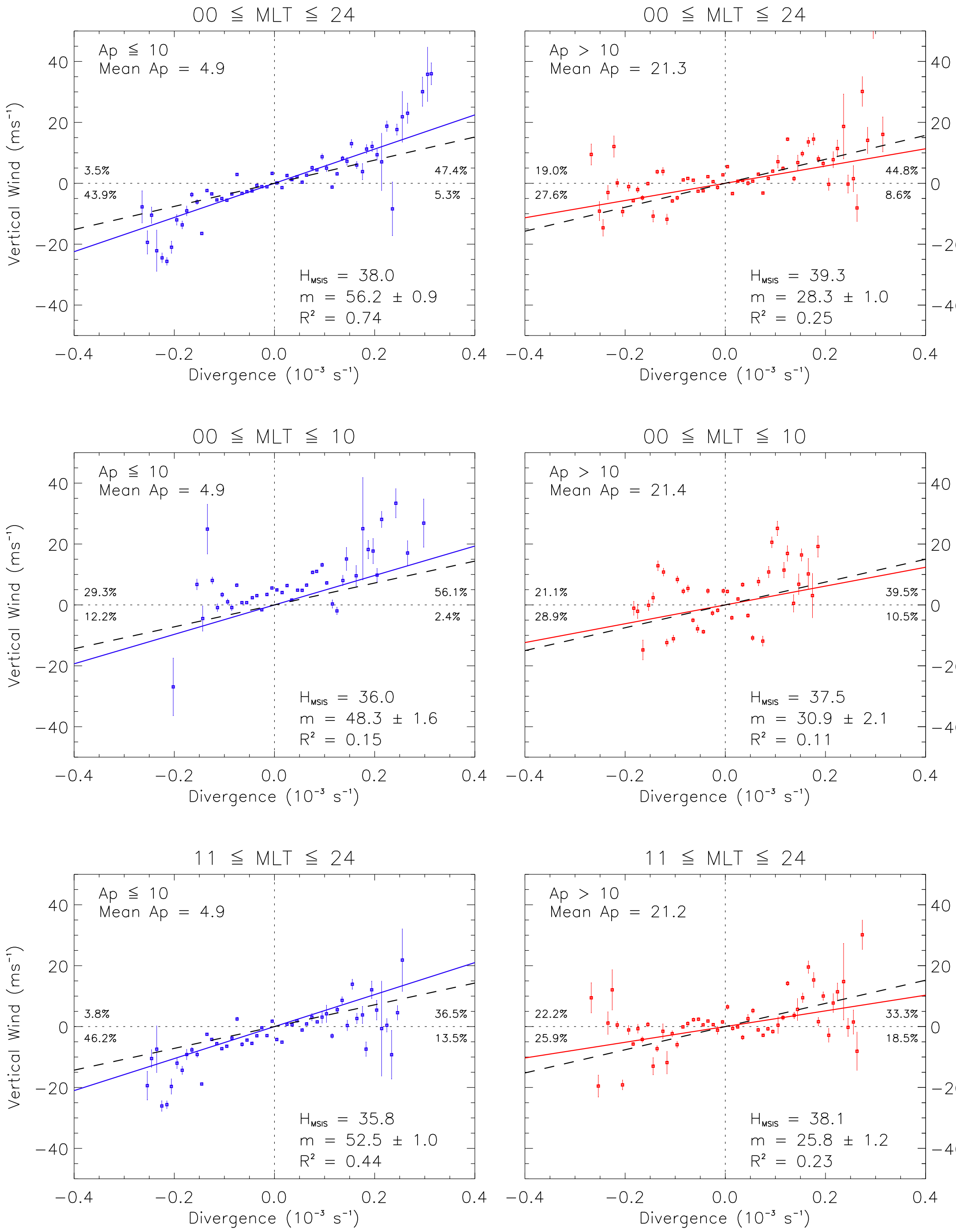


Figure 8.3.2: Same format as Figure 8.3.1, but for two years of data from Mawson.

the scale height, derived from the NRLMSISE-00 model using the mean temperature and geomagnetic activity level of the relevant data set (the modelled scale height is labelled H_{msis}). A value of unity for the ratio H_{msis}/m would therefore indicate that the observed horizontal divergence could be used as a proxy for the vertical wind using the Burnside relation, assuming that the scale height modelled using NRLMSISE-00 was reasonably accurate.

Considering data from all magnetic local times (top row in Figures 8.3.1 and 8.3.2), under quiet conditions the vertical winds above both Davis and Mawson appeared to respond approximately linearly to changes in the horizontal divergence, albeit over a limited range. At Davis this range was approximately between $\pm 5 \times 10^{-4}$ s^{-1}, while at Mawson the range was closer to $\pm 2 \times 10^{-4}$ s^{-1}. Outside of this range there appeared to be a systematic departure away from the linear trend, particularly for large negative values of divergence. The model fits to these data returned values for m of $\sim$ 30 km at Davis and almost twice this value ($\sim$ 56 km) at Mawson. Under active conditions the fitted 'scale height' at Davis was in very close agreement with the modelled value of $\sim$ 39 km ($H_{msis}/m \sim 1.02$), while the fitted scale height at Mawson was 28 km, approximately 10 km smaller than the modelled value ($H_{msis}/m \sim 1.4$). This trend in fitted scale heights was observed in all time ranges, with quiet-time scale heights calculated at Mawson equal to approximately 1.5 times the corresponding Davis scale height, while active-time values at Mawson were approximately 0.7-0.75 times the values derived at Davis. At Mawson the quiet-time fitted scale height values were always significantly higher than active-time values, while at Davis the quiet-time value in a given time sector was always lower than the corresponding active-time value, although the relative difference was not as large as at Mawson. Note that while the mean *ap* in the quiet-time data was similar for the two stations at a value of around 5 nT, the mean *ap* at Davis during active times was in the range of 25-36 nT, compared to $\sim$ 22 nT at Mawson.

In the post-midnight sector, between 00:00 and 10:00 MLT (middle row in the figures) there was a relative decrease in the number of downward winds associated with negative divergence (corresponding to the lower-left quadrant of the plots) compared to the data from all magnetic local times. This was particularly true at Mawson, where under quiet conditions the percentage of such events was approximately 4 times smaller than were observed in the 11:00-24:00 MLT time sector, although under active conditions there was actually a slightly larger percentage in this quadrant. In addition, the Mawson data in this time sector were noticeably biased toward positive vertical winds, thus the data are not accurately modelled by a function of the form $y = mx$. At Davis the effect was not as large, and the decrease in downward winds observed with convergent horizontal winds was comparable under both levels of activity. At both stations the decrease in the number of downwellings associated with convergent horizontal winds was offset by an increase in the number of upwellings associated with convergent winds (upper-left quadrant), and also by a generally smaller increase in the number of upwellings associated with divergent winds (upper-right quadrant). Post magnetic midnight data (bottom row) were very similar to the data from all magnetic local times at both stations, since the majority of data was

Davis		m	$\mathrm{H_{msis}}$	$\mathrm{H_{msis}}/m$
Quiet	(00-10 MLT)	33.8 ± 0.5	35.6	1.05
Quiet	(11-24 MLT)	31.1 ± 0.3	36.3	1.17
Quiet	(00-24 MLT)	29.6 ± 0.2	38.0	1.28
Active	(00-10 MLT)	41.6 ± 0.6	37.7	0.91
Active	(11-24 MLT)	34.7 ± 0.5	38.7	1.12
Active	(00-24 MLT)	39.0 ± 0.4	39.8	1.02
Mawson				
Quiet	(00-10 MLT)	48.3 ± 1.6	36.0	0.75
Quiet	(11-24 MLT)	52.5 ± 1.0	35.8	0.68
Quiet	(00-24 MLT)	56.2 ± 0.9	38.0	0.68
Active	(00-10 MLT)	30.9 ± 2.1	37.5	1.21
Active	(11-24 MLT)	25.8 ± 1.2	38.1	1.48
Active	(00-24 MLT)	28.3 ± 1.0	39.3	1.39

Table 8.1: Fitted and modelled scale heights and the ratio $\mathrm{H_{msis}}/m$, at Mawson and Davis.

obtained in this magnetic time sector.

The agreement between the fitted scale heights and those modelled using NRLMSISE-00 was quite good at Davis, under both quiet and active conditions. At Mawson the agreement was poor under both quiet and active conditions: quiet time fitted scale heights were $\sim 1.2 - 1.5$ times the modelled scale height, while under active conditions the trend was opposite, with modelled scale heights $\sim 1.2 - 1.5$ times the fitted scale heights. At Davis, quiet-time fitted scale heights were smaller than the modelled values, while at Mawson the opposite was true (these data are summarised in Table 8.1). Overall the data from Davis produced the best agreement with the modelled Burnside relation, at least within the range of divergence values for which the data responded approximately linearly to changes in divergence.

As mentioned previously the fitted lines were weighted by the inverse of the error-bar/uncertainty value, and thus the linear portion of each plot (for which the uncertainties were relatively low) contributed the most to the fit. Note also that while the Burnside relation depends upon the scale height, which can vary with local time as the temperature and thermospheric composition varies, the ratios shown in Table 8.1 are not affected by this variation, as the scale heights used in the comparison were derived from modelled atmospheric composition and temperature, which would show the same (or very similar) average local-time variations as the measured scale heights.

Smith and Hernandez (1995) carried out a similar investigation to that described above at South Pole station, which is at a similar magnetic latitude to Davis (75°). These authors presented results for a single night during a storm period, during which the daily *ap* index reached approximately 200 nT. On this night the relationship observed between vertical wind and divergence was opposite to what has been presented above, i.e. upward vertical winds were associated with negative divergence, and downward winds were associated with positive divergence. The range of divergence values observed was very similar to those presented in Figure 8.3.1. However these authors calculated a scale height of -78.8

km, which was more comparable (in magnitude at least) to the quiet-time scale heights calculated above Mawson in the present study.

Guo and McEwen (2003) examined data from five winters (1993-95, 1997-2000) of Fabry-Perot interferometer measurements from Eureka (magnetic latitude 88.9°). While no fitting was done to these data to calculate the 'scale height', the measurements suggested that the observed range of horizontal divergence was similar to that observed above Mawson, however much higher vertical wind speeds were observed by these authors, indicating that a much larger scale height would result. The sense of the data was however the same as was observed in the present study: upward vertical winds associated with positive divergence, downward vertical winds with negative divergence, in agreement with the Burnside relation.

It was pointed out in Section 4.9 that in order to derive line-of-sight wind estimates it is necessary to determine the spectral peak-position that corresponds to zero Doppler-shift. Since no convenient laboratory source exists for radiation at the wavelength of interest (630.0 nm) this Doppler-shift baseline was calculated by assuming that the average vertical wind was zero when averaged over an entire night. In the discussion of Section 2.4.2, is was seen that this condition (mean zenith wind equal to 0 m.s^{-1}) might not always be met, but that the error introduced by the assumption should be at most 10-20 m.s^{-1} (Aruliah and Rees, 1995), and likely much less.

Consider the case of a uniform (non-divergent) horizontal wind flowing above the observatory. The line-of-sight wind estimates from opposite look directions (directions separated by 180° of azimuth) will then be of equal magnitude but opposite sign (i.e. they will lie an equal distance above and below the zero Doppler-shift baseline). If for some reason the mean vertical wind did not correspond to the 'true' zero Doppler-shift baseline (if for example a diurnal component were present in the vertical wind, of which less than

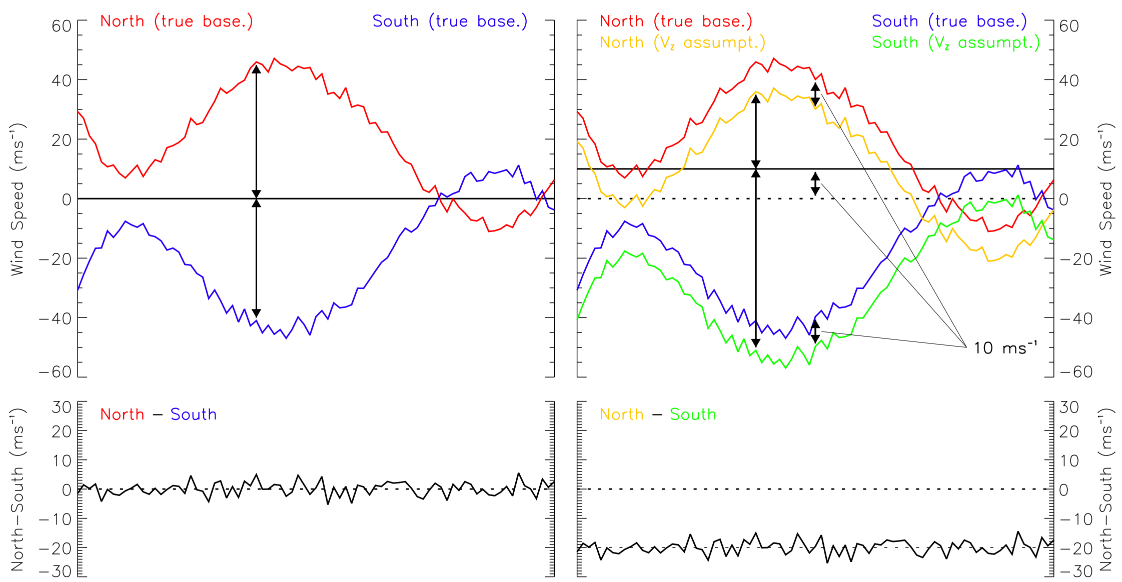


Figure 8.3.3: Schematic showing the effect that an incorrect Doppler baseline has on observed winds and the divergence calculated from them.

a full cycle were being sampled due to short observation periods), then the resulting wind estimates would have *un*equal magnitudes relative to the (incorrect) Doppler baseline, and as such the flow would appear divergent or convergent, depending on the relative shift between the true and assumed Doppler baselines. The magnitude of the apparent divergence/convergence would be proportional to twice the difference between the true and assumed baselines. Thus a systematic error in the assumed Doppler baseline would produce a systematic error in the calculated divergence, however that divergence would not be related to the sign (or magnitude) of the instantaneous vertical wind, since for each night the divergence offset due to the incorrect Doppler baseline would be (approximately) constant, while the vertical wind would not.

This situation is depicted schematically in Figure 8.3.3. In this example, red and blue lines in the top left panel represent northward and southward looking wind estimates, relative to the true Doppler baseline (solid black line). A positive wind is directed away from the observatory, a negative wind toward the observatory. Each of these lines thus indicates a northward meridional wind. The meridional divergence is therefore zero, within the limits of measurement noise (bottom left panel). In the top right panel, additional lines have been drawn showing the winds relative to an incorrect Doppler baseline based on an assumption of zero mean vertical wind (winds labelled 'V_z assumpt.'). As shown in the bottom right panel, the effect of the incorrect baseline is to produce a systematic apparent negative meridional divergence, equal to twice the difference between the true and assumed Doppler baselines, which in this example was chosen to be equivalent to 10 $m.s^{-1}$. If there were a significant, consistent error in the assumed Doppler baselines used to derive the data shown in Figures 8.3.1 and 8.3.2, the effect would be to shift the data points left or right along the divergence axis, according to whether the incorrect baseline resulted in a positive or negative apparent divergence offset. Visual inspection of the data in these figures suggests that any such effect is small.

From these data it appears that the horizontal divergence could be used more reliably as a proxy for the vertical wind (using the Burnside relation) during active geomagnetic conditions, where active denotes here an *ap* index of the range 20-30, which would at best be considered moderately active when compared to the range of activity observed over a normal solar cycle. Even under very quiet geomagnetic conditions, $ap \sim 5$, the Burnside condition could be used with reasonable accuracy for the Davis data, however at Mawson the evaluation of the Burnside relation would have required a scale height approximately 1.5 times greater than that predicted by the NRLMSISE-00 model in order to produce reasonable agreement between vertical winds calculated from the Burnside relation and those observed directly by the Mawson SDI. This difference in behaviour is likely related to the different location of each station relative to the auroral oval, since this represents a highly variable source of energy input into the F-region thermosphere at high-latitudes, and would likely invalidate many of the assumptions mentioned in Section 8.1. Mawson station is usually located equatorward of or underneath the nominal auroral oval, while Davis station is usually within the region bounded by the auroral oval, the polar cap. From the vertical wind variability maps of Innis and Conde (2002) it is clear that the nominal auroral oval represents an effective boundary between high and low vertical wind variability, and thus it is reasonable to expect that the relationship between divergence and vertical wind might similarly depend on location relative to the auroral oval.

As Cooper et al. (2009) showed, the applicability of the Burnside relation depends to some extent on the variability of the forces driving the vertical winds, and on the altitude range being sampled relative to the region of heating. Mawson station, which is often within (underneath) the auroral oval, often samples a wind field which is driven by highly spatially localised and rapidly time varying forces (as shown in Sections 8.4 and 8.5 below), and therefore would perhaps not be expected to observe vertical winds that are linked to the horizontal wind divergence in such a simple manner as given by the Burnside relation. As shown in Table 8.1, under different levels of activity the data from Mawson showed opposite trends with respect to the Burnside relation: during quite times the ratio $\mathrm{H}_{\mathrm{msis}}/m < 0.8$, while during active times $\mathrm{H}_{\mathrm{msis}}/m > 1.2$. At Davis station, where the percentage of time spent underneath the auroral oval is much lower, the data showed much better agreement with the Burnside relation, with values of $\mathrm{H}_{\mathrm{msis}}/m$ closer to unity. Also at Davis the active-time data (taking the 00-24 MLT ratio as representative of the data) was in better agreement with the Burnside relation than the quiet-time data. When geomagnetic conditions are active, the auroral oval would be expected to expand further equatorward on average, thus ensuring that Davis station would be well inside the polar cap, perhaps explaining why the agreement with the Burnside condition improved under these conditions.

8.4 Local Divergence and Vertical Wind

In Chapter 7, results from a campaign of bistatic observations between the Mawson and Davis spectrometers were presented. In that study, vertical winds from three common-volume regions lying between Mawson and Davis were derived from line-of-sight observations from both stations. The results revealed a vertical wind field that was highly variable, both spatially and temporally, but which was at times well correlated over large ($\sim$ 480-635 km) distances. One day in particular (April 7, 2008, day 98) showed strong, short-lived vertical motions in the common-volume regions associated with an auroral arc moving rapidly equatorward across the field-of-view (see Figure 8.4.1, which is reproduced from Chapter 7). An obvious question arising from these results is how such vertical motions were related to the horizontal winds, and in particular if there were any correspondence between the vertical winds in the common-volume regions and the divergence of the horizontal wind field in those same regions.

Figure 8.4.2 shows the evolution of the inferred horizontal wind field during the period 17:01-17:58 UT, during which time a bright auroral feature moved rapidly across the field-of-view from (geomagnetic) south to north, before decreasing in brightness and becoming a more diffuse structure centred overhead of the station. Prior to the start of this period the horizontal wind speed had been very low across the entire field-of-view, with an average magnitude < 20 m.s^{-1}. At around 17:15 UT, a vertical wind disturbance

(Figure 8.4.1), which was observed to begin in the common-volume regions approximately 15 minutes prior, had peaked, with a 150 m.s^{-1} downward vertical wind observed in region C, coincident with a small downwelling above Davis and a 50 m.s^{-1} upwelling in region A. Following this peak the vertical winds in all regions reversed direction, reaching magnitudes of 50-80 m.s^{-1}. This direction reversal was coincident with a magnetic equatorward expansion of the auroral oval. As the auroral brightness began to decrease, the vertical winds in all regions also decreased in magnitude.

The horizontal winds inferred from the line-of-sight measurements are shown in the first, third and fifth rows of Figure 8.4.2. Green arrows show the horizontal wind vectors (the wind scale is shown in the lower right corner) overlaid on the all-sky images recorded by the Mawson SDI (the uncalibrated intensity scale is shown below the vector wind scale). The vector in the central zone represents the average wind direction over the entire field-of-view. The Mawson SDI zones which contributed to the common-volume vertical wind calculations are outlined in blue, and magnetic northward (equatorward) is to the top of these images. The magnetic equatorward expansion of a bright loop-like auroral structure was clearly associated with a surge in horizontal wind speeds. A region of strong wind

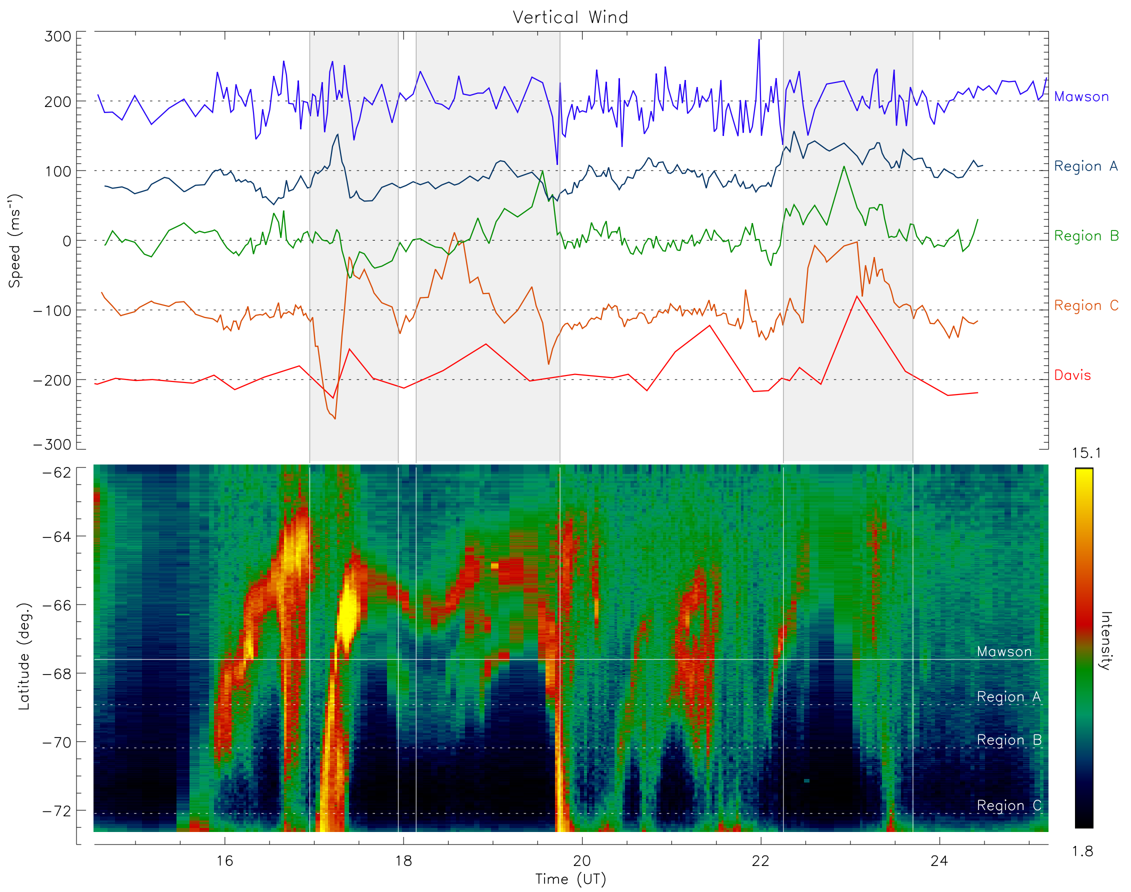


Figure 8.4.1: Top panel: vertical winds observed above Mawson and Davis and derived in the common-volume regions A (blue line), B (green line) and C (red line) on April 7, 2008 (day 98). Lower panel: keogram derived from vertical cross-sections through the all-sky images recorded by the Mawson SDI. Figure reproduced from Chapter 7.

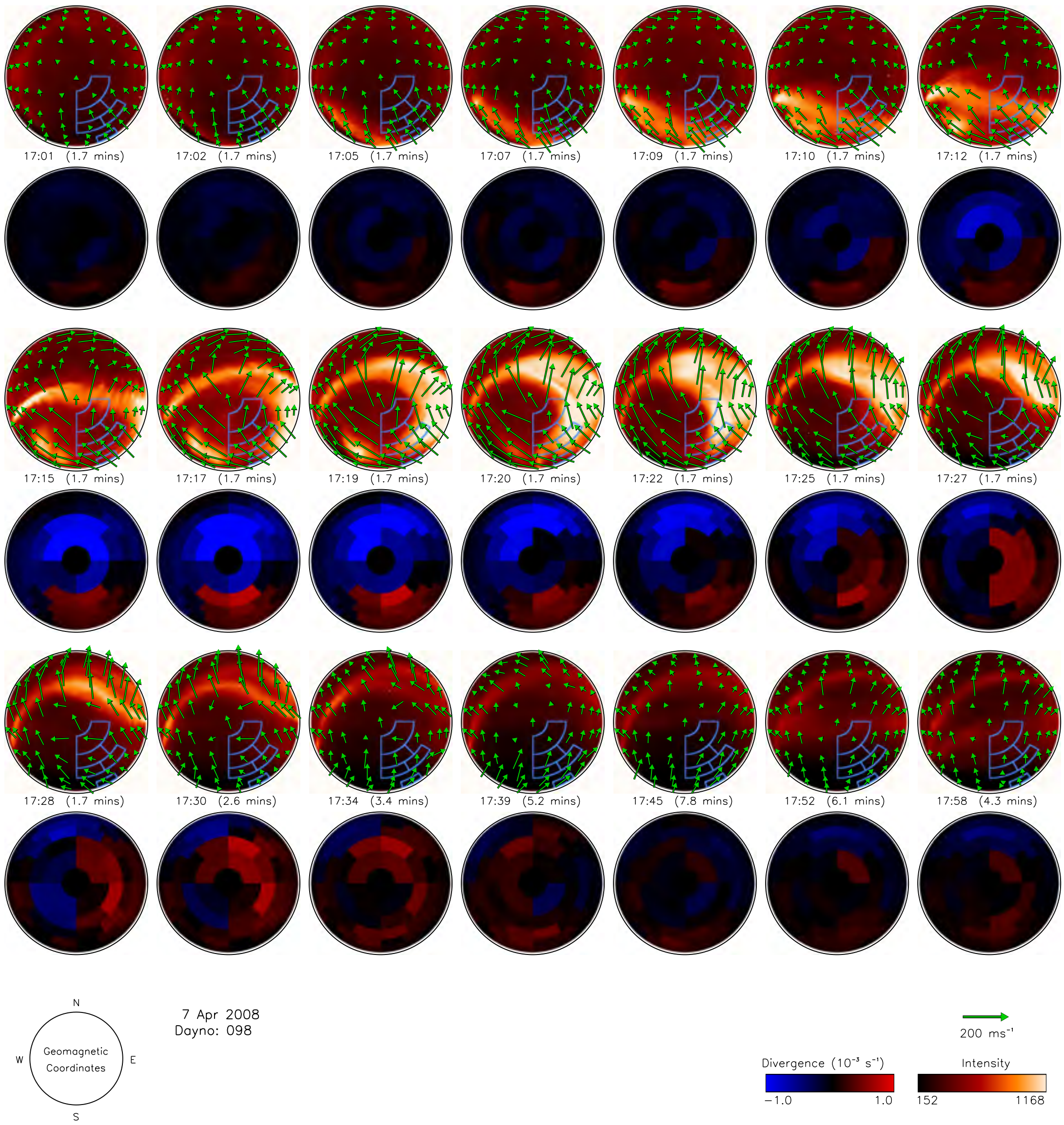


Figure 8.4.2: Vector wind fields, all-sky images and the divergence field on April 7, 2008, for the period 17:01-17:58 UT. The first, third and fifth rows show the fitted vector wind field (green arrows) overlaid on the all-sky images recorded by the Mawson SDI (and not projected onto the sky). The second, fourth and sixth rows show the divergence calculated in each zone (described in text). Wind, intensity and divergence scales are shown in the bottom right corner of the figure. The coordinate system for each wind field and divergence image is shown in the bottom left corner. Between each wind and divergence field are shown the universal time halfway through the exposure, and the integration time in parenthesis.

gradients in the meridional direction developed on the equatorward edge of the auroral arc and closely tracked the edge of the arc as a region of 100-200 m.s^{-1} winds developed on the poleward edge of the arc as it moved across the field-of-view. The wind gradient was large over a very small meridional extent, in most cases less than the zone separation of approximately 100 km. Such sharp gradients can be resolved because the vector wind fit retains all of the information contained in the directly measured line-of-sight wind component. In the magnetic northeast a region of wind shear was apparent between 17:07 and 17:20 UT, with winds directed magnetic eastward in the region, as opposed to the more northward/northwestward flow observed over most of the rest of the field-of-view.

The second, fourth and sixth rows of Figure 8.4.2 show estimates of the divergence calculated within each of the Mawson SDI's software-defined zones for each exposure. The divergence estimate within each zone was calculated as follows: for a given exposure, the horizontal component of the directly measured line-of-sight wind speed ($H_{\|}$) and the fitted component normal to it ($H_{\perp}$, described in Section 4.10) were interpolated across the entire field-of-view using smooth quintic polynomial interpolation in two dimensions, implemented by the IDL function "TRIGRID". Prior to the interpolation, the mean value

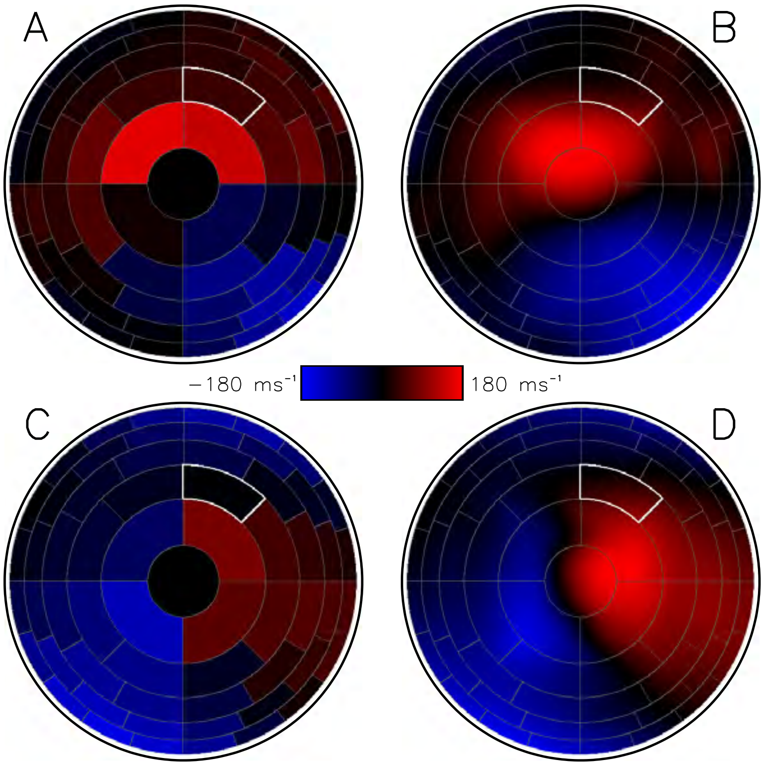


Figure 8.4.3: Panel A: horizontal line-of-sight winds ($H_{\|}$) measured by the Mawson SDI. Panel B: $H_{\|}$ from panel A interpolated over the field-of-view. Panel C: fitted horizontal winds normal to the line-of-sight ($H_{\perp}$). Panel D: $H_{\perp}$ from panel C interpolated over the field-of-view. Note that these maps are shown from the perspective of an observer looking down on the field-of-view from a point in space (thus the zone numbers increase *clockwise* here, compare with top panels of Figure 4.6.1), however they are not shown *projected* onto the sky, as are the zone maps in the lower panels of Figure 4.6.1.

of $H_{\parallel}$ and $H_{\perp}$ in the four zones of the first ring out from the centre were used as values of these wind components in the central zone, where horizontal components cannot be measured. Strictly speaking, the radial and normal components are not defined in the zenith. The use of average wind components from the first annulus as estimates for the entire central zone is thus simply a convenience to provide continuity across this region for the interpolation; no divergence estimate is actually obtained in the central zone.

Figure 8.4.3 shows the results of this interpolation, where the data for this example come from the exposure at 17:15 UT in Figure 8.4.2. For each zone (zone number 7 is used in this example, which is highlighted in Figure 8.4.3) the interpolated $H_{\parallel}$ within the zone were binned by radial distance r (projected onto the sky at an altitude of 240 km) from the centre of the field-of-view and averaged, to give the average line-of-sight velocity as a function of radial distance (shown in the top left panel of Figure 8.4.4). A similar procedure was carried out for $H_{\perp}$ (lower left panel in Figure 8.4.4), however the independent variable in this case was the distance in the direction of the unit vector normal to the line-of-sight (anti-clockwise, the distance being labelled n). This distance was calculated using the projected radius r and the angle in radians relative to one edge of the zone, thus zero distance corresponded to one azimuthal edge of the zone. To each

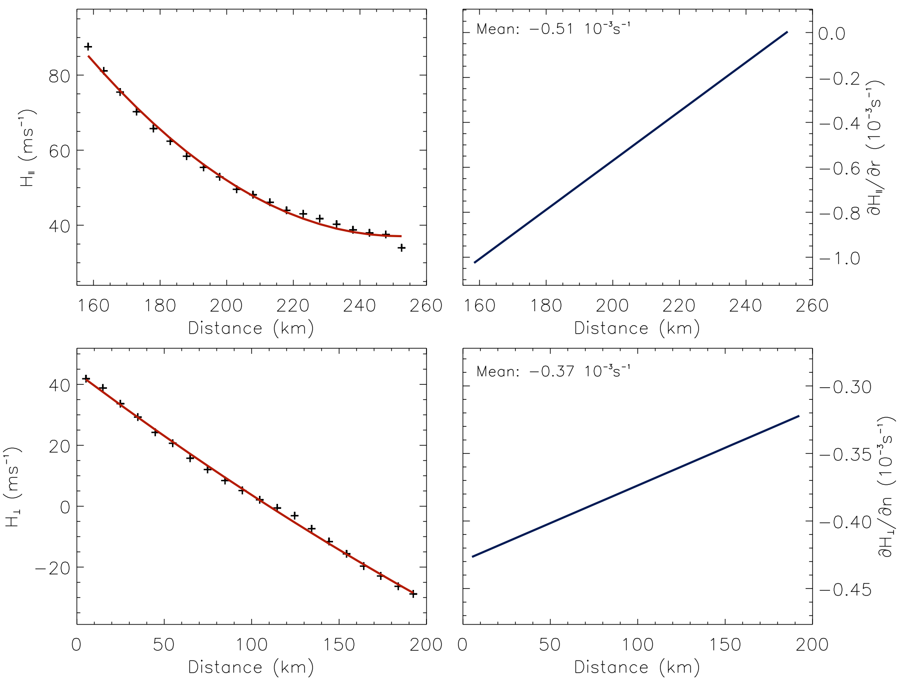


Figure 8.4.4: Top left: symbols show the average line-of-sight wind plotted against radial distance (projected onto the sky at 240 km altitude) from the centre of the field-of-view, for twenty distance bins in zone number 7. The solid red line shows the 2$^{\mathrm{nd}}$-order polynomial fitted to these line-of-sight winds. Top right: the solid blue line shows the divergence of the line-of-sight winds ($\partial H_{\parallel}/\partial r$) shown top left. The mean divergence is indicated in the top left corner of the plot. In the lower left panel is plotted (with symbols) the average fitted wind component normal to the line-of-sight ($H_{\perp}$) as a function of distance in the direction of the unit vector normal to the line-of-sight (this distance is labelled n), and the fitted 2$^{\mathrm{nd}}$-order polynomial is plotted as a solid red line. The lower right panel shows the divergence of the normal component ($\partial H_{\perp}/\partial n$).

of these curves was fitted a 2$^{\text{nd}}$-order polynomial function (solid red lines in the figure) as a local approximation to the global quintic interpolation across the field-of-view. The analytic derivative of this fitted polynomial was then calculated, as shown in the right column of Figure 8.4.4, $\partial H_{\parallel}/\partial r$ in the top panel and $\partial H_{\perp}/\partial n$ in the lower panel. Finally, the mean value of the derivative in each case was used as an estimate of the divergence in the given direction, and the total divergence for the zone was then simply the sum of the two divergence components:

$$D = \frac{\partial H_{\parallel}}{\partial r} + \frac{\partial H_{\perp}}{\partial n} \tag{8.4.1}$$

In Figure 8.4.2, the second, fourth and sixth rows of images show the divergence calculated for each zone using the method described above, where blue colours indicate negative divergence (convergent flow) and red colours indicate divergent flow, according to the scale shown in the bottom right corner of the figure. A point to note regarding the method outlined above is that only the $H_{\parallel}$ component is independent, since that is the component of the wind field which is measured directly. The normal component is fitted to the observed line-of-sight winds using the method described in Section 4.10, which requires certain assumptions to be made regarding the variation of the horizontal wind field across the field-of-view. Importantly, only linear gradients are used to describe the variation of the orthogonal wind components along the meridional and zonal directions, and these gradients are assumed constant across the field-of-view. The normal component of the wind is derived from the fitted vector field, which itself is calculated from the coefficients of a Fourier series expansion of the horizontal line-of-sight winds around each annulus. While the gradient terms describing the vector field are constant across the field-of-view, the background wind components are calculated separately in each annulus. The result of these assumptions is that the derivative of the normal wind component around each annulus will be described by a sinusoidal function of azimuthal angle, and will only differ between annuli due to the different background wind terms calculated in each annulus, in addition to a linear dependence on radial distance from the zenith.

Thus the spatial variability which is observed in the divergence plots of Figure 8.4.2 is due to the variability of the measured line-of-sight winds, superimposed on the smooth sinusoidal variation of $\partial H_{\perp}/\partial n$ in the azimuthal direction, and on the smaller radial variation of $\partial H_{\perp}/\partial n$ due to the contribution from the background wind and radial distance. While no new information about the horizontal wind field is gained from this type of analysis, it is almost essential to see the divergence highlighted in this way in order to investigate its spatial variation. While this information is essentially contained in the plotted vector wind fields, examples of which have been shown already, unless the divergence is very pronounced it can very difficult to 'map out' the divergence simply by visual inspection of these fields.

In Figure 8.4.2, the region of strong (-1×10^{-3} s^{-1}) convergence on the equatorward edge of the auroral arc is clear in both the vector wind fields and the divergence plots. The latter also show a region of divergence on the poleward side of the arc, which expanded,

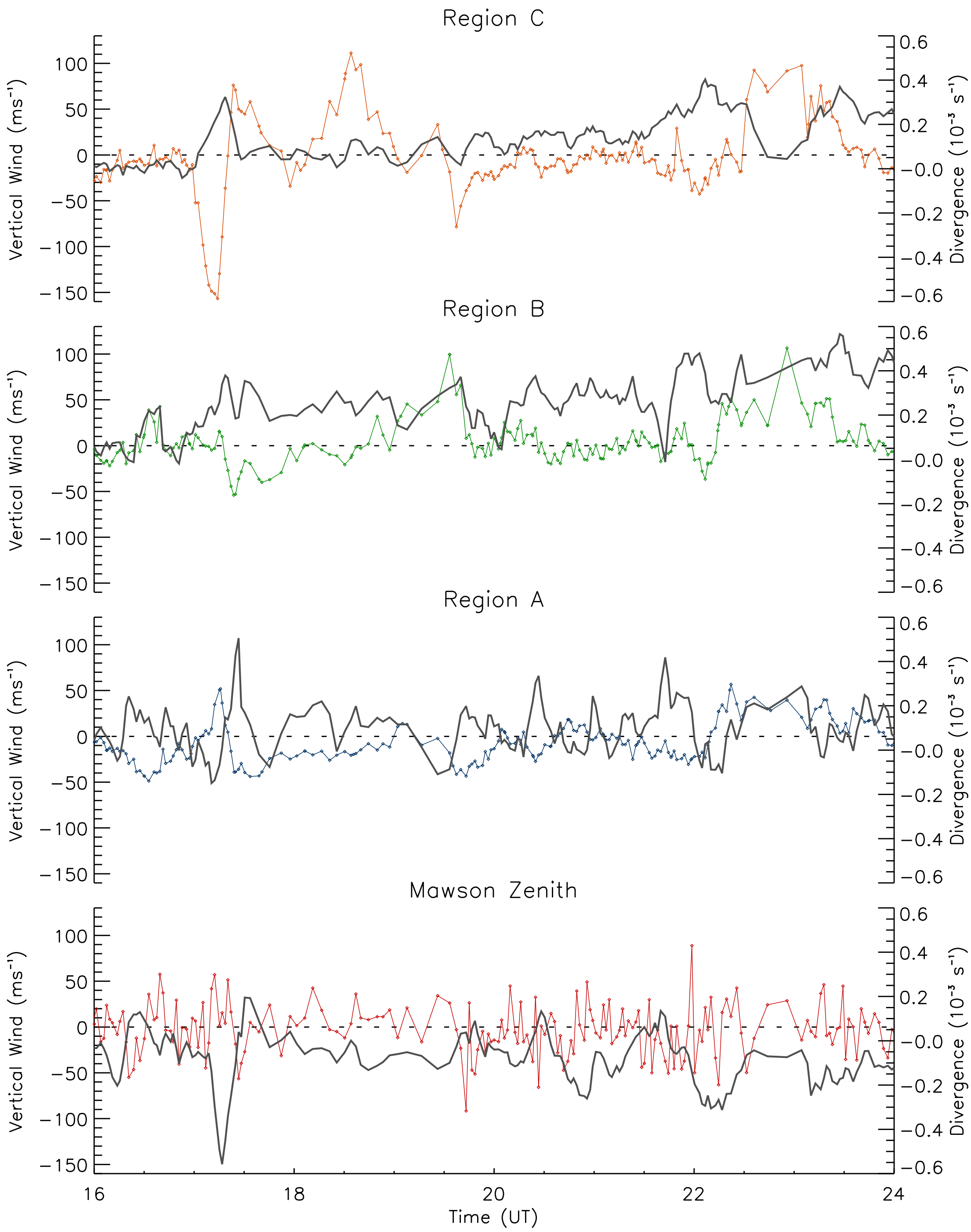


Figure 8.4.5: Vertical winds derived in the common-volume regions compared with the divergence inferred in those regions. Vertical winds are shown as coloured lines, relative to the left-hand-side y-axis, and divergences are shown as grey lines, and correspond to the right-hand-side y-axis. In the lowest panel, the vertical winds are those recorded in the Mawson zenith (and not bi-statically measured), while the divergence shown in this panel is that calculated from the fitted vector field, across the entire field-of-view.

from a localised region on the magnetic southward edge of the field-of-view, toward the magnetic east, and by 17:28 UT covered a large fraction of the field-of-view, as the auroral arc brightness had begun to decrease. Interestingly, the line dividing the convergent and divergent areas was aligned essentially parallel to the auroral arc (magnetically east-west) at 17:15 UT, while by 17:27 UT this dividing line was aligned magnetically north-south, perpendicular to the arcs. At the beginning of the time period shown in the figure the winds were very weak, and the divergence was correspondingly very small. By 17:58 UT, after the arc brightness had decreased appreciably, the divergence had also weakened across the field-of-view, while the horizontal winds had increased in magnitude relative to the start of the period, with speeds around 40 m.s^{-1} directed predominantly magnetic northward. The strong vertical wind disturbances in the common-volume zones were observed during a period of rapid temporal variations and strong spatial gradients in the horizontal wind field, associated with the equatorward motion of the bright auroral arc.

By calculating the divergence in each zone over the entire night, it was possible to compare the divergence estimates within the common-volume regions with the vertical winds derived there from the bistatic experiment described in Chapter 7. This comparison is shown in Figure 8.4.5. The top three panels show (from the top) the vertical winds derived in the common-volume regions C, B and A (coloured lines, the scale given by the left-hand-side y-axis) and the average divergence within those common-volume regions calculated using the method outlined above (grey lines, the scale given by the right-hand-side y-axis). Region C corresponds to Mawson zone number 54. The region B divergence represents the average divergence from Mawson zones 20 and 21, while the region A divergence is the average divergence in zones 3 and 10. The lowest panel shows the vertical winds (red line) recorded in the Mawson zenith (and was therefore not derived from bistatic measurements) and the divergence averaged over the two inner annuli (grey line) was used as the estimate of divergence in the zenith. The uncertainty in these calculated divergence estimates is difficult to quantify, however Monte Carlo simulations using input wind-fields with normally distributed noise added (typical of F-region wind uncertainties) suggest that the uncertainty in D is $\simeq 1.3 \times 10^{-4}$ s^{-1}. This is approximately 20% of the largest values shown in Figure 8.4.5.

There was often no clear correlation between the divergence calculated within each of the common-volume regions and the vertical winds derived there from the bistatic wind data. There were however some clear correlations between the different regions, as for example at approximately 17:15 UT, where all three common-volume regions showed a strong, positive divergence peak, in the range $2 - 5 \times 10^{-4}$ s^{-1}. A strong convergence was observed in the Mawson zenith at this time. The vertical wind above Mawson and in regions A and C at this time were all of opposite sign to the calculated divergence, contrary to what would be expected from the Burnside relation. In region A the divergence changed sign at the same time as the vertical wind switched from upwelling to downwelling. There was negligible divergence associated with the large upward wind event between 18:00 and 19:00 UT in region C, and also during the upwelling between 22:30 and 23:15 UT. Indeed between approximately 21:00 and 23:00 UT the divergence estimate in region C appeared anti-correlated with the vertical wind derived there. In region B the divergence remained positive and showed an essentially increasing trend during the night. This kind of divergence would be more easily sustained over small scales than it would, for example, over a region as large as the entire instrument field-of-view.

Of course this kind of analysis would be much improved if independent information on the wind component normal to the line-of-sight (with a spatial resolution comparable to the Mawson SDI) could be used instead of a normal component fitted to the azimuthal variation of the line-of-sight winds. As formulated here, only the line-of-sight wind variations contribute information to the calculation of divergence within each zone, and thus the usefulness of this analysis in the present circumstance is limited. Two scanning Doppler imagers with overlapping fields-of-view would be optimal for this purpose, although no independent horizontal wind information would be obtained along the line joining the two instruments, and since this is where direct bistatic measurements of the vertical wind would be possible, even this experimental set-up would have drawbacks. Only with three such instruments would it be possible to resolve the full three-dimensional wind field within the field-of-view overlap with sufficient spatial resolution.

8.5 Wind Shear

In the previous section a region of wind shear of limited spatial extent was identified in Figure 8.4.2 along the geomagnetic northern (equatorward) edge of the field-of-view. The magnitude of the wind shear was observed to increase as the equatorward edge of the loop-like auroral structure moved toward the shear region. Wind fields displaying quite strong shear have also been observed on a number of other nights, usually associated with bright auroral arcs within the field-of-view. The regions of wind shear were often observed to move in tandem with the auroral features.

In this section, four case studies will be presented showing strongly sheared wind fields associated with bright auroral arcs on four different nights. From these wind fields it was possible to derive an estimate of the amount of power that would have been required to drive the observed sheared wind fields, under the assumption that the only dissipative mechanism was viscosity. The vector fitting algorithm calculates the average of each gradient across the entire field-of-view, and these should therefore be considered lower limits to the actual gradients which may have been present, particularly in the immediate vicinity of the auroral arcs, where gradients could have been much sharper than it is possible to represent using all-sky averages of the gradient terms. In addition, in order to constrain the vector fit it has been assumed that the zonal gradient of the meridional wind was equal to zero. The power estimates presented here would therefore almost certainly represent a lower limit to the power required to drive the observed wind fields.

The rate of viscous heating (viscous dissipation) can be calculated from the following relation (Rees, 1989, p. 125):

$$\begin{aligned}\frac{\partial q_{\mathrm{v}}}{\partial t} &= \bar{\varepsilon} : \nabla\mathbf{u} \\ &= \sum_{ij} \varepsilon_{ij} \frac{\partial u_j}{\partial x_i} \qquad i,j = 0,1,2\end{aligned} \tag{8.5.1}$$

where $\nabla\mathbf{u}$ is the (tensor) gradient of the neutral velocity and $\bar{\varepsilon}$ is the viscous contribution to the stress tensor, whose components (in the Navier-Stokes formulation) are given by:

$$\varepsilon_{ij} = \mu \left[\left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i} \right) - \frac{2}{3} (\nabla \cdot \mathbf{u}) \right] \tag{8.5.2}$$

The quantity μ is the coefficient of dynamic viscosity (kg.m^{-1}.s^{-1}), and u_j is the component of the neutral velocity vector in the j^{th} direction. Equations 8.5.1 and 8.5.2 lead to the expression for the viscous heating per unit volume (W.m^{-3}):

$$\begin{aligned}\frac{\partial q_{\mathrm{v}}}{\partial t} &= \mu \left[2\left(\frac{\partial u_x}{\partial x}\right)^2 + 2\left(\frac{\partial u_y}{\partial y}\right)^2 + \left(\frac{\partial u_y}{\partial x} + \frac{\partial u_x}{\partial y}\right)^2 - \frac{2}{3}(\nabla \cdot \mathbf{u})^2 \right] \\ &= \mu \left[2\left(\frac{\partial u_x}{\partial x}\right)^2 + 2\left(\frac{\partial u_y}{\partial y}\right)^2 + \left(\frac{\partial u_y}{\partial x} + \frac{\partial u_x}{\partial y}\right)^2 - \frac{2}{3}\left(\frac{\partial u_x}{\partial x} + \frac{\partial u_y}{\partial y}\right)^2 \right]\end{aligned} \tag{8.5.3}$$

where the wind field is assumed to be independent of altitude ($\frac{\partial}{\partial z} = 0$), and the vertical wind is assumed constant over the field-of-view ($\frac{\partial u_z}{\partial x} = \frac{\partial u_z}{\partial y} = 0$). From Section 4.10, the gradient of the meridional wind in the (magnetic) zonal direction is equated to zero in order to adequately constrain the fitted wind field, and the wind components u_x and u_y are labelled u and v respectively, thus for this work:

$$\frac{\partial q_{\mathrm{v}}}{\partial t} = \mu \left[2\left(\frac{\partial u}{\partial x}\right)^2 + 2\left(\frac{\partial v}{\partial y}\right)^2 + \left(\frac{\partial u}{\partial y}\right)^2 - \frac{2}{3}\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right)^2 \right] \tag{8.5.4}$$

The coefficient of dynamic viscosity, μ, is given to a good approximation by (Rees, 1989):

$$\mu_k = A_k \left(\frac{T}{1K}\right)^{0.69} \quad (\mathrm{kg.m^{-1}.s^{-1}}) \tag{8.5.5}$$

in the temperature range 200-2000 K, and the coefficients A_k relate to the different atmospheric species, and are given in Table 8.2. The viscosity in a multi-component atmosphere can be approximated by the weighted sum:

$$\mu = \frac{\sum_k A_k n_k}{\sum_k n_k} \left(\frac{T}{1K}\right)^{0.69} \tag{8.5.6}$$

where n_k is the number density (m^{-3}) of the k^{th} species. In applying Equation 8.5.6, the NRLMSISE-00 atmospheric model was used to estimate the number densities of the neutral species listed in Table 8.2 at an altitude of 240 km. The mean temperature across all zones in each exposure was used as input to the model, as were the prevailing *ap* and $F_{10.7}$ index values. The mean all-sky temperature was also used directly in Equation 8.5.6. Along with the estimated coefficient of viscosity, the fitted gradients returned by the vector wind fitting algorithm allowed the viscous heating rate to be inferred.

Figure 8.5.1 shows the fitted vector wind fields (green arrows) overlaid onto the all-sky images recorded by the Mawson SDI (the intensity scale is shown in the lower right corner of the top panel, along with the vector scale). As before the vector in the central zone represents the mean wind direction and magnitude across all zones. In the lower panel, the group of four plots show the various parameters from which the rate of viscous heating was derived. Thus the mean (across the field-of-view) temperature, fitted wind gradients (black line - $\frac{\partial u}{\partial x}$, blue line - $\frac{\partial u}{\partial y}$, red line - $\frac{\partial v}{\partial y}$), the mean number density returned by the NRLMSISE-00 atmospheric model, and the dynamic viscosity calculated using Equation 8.5.5 are shown in this group. The calculated rate of viscous heat dissipation per unit volume is shown in the lowest panel. In all time-series plots the shaded area shows the time period displayed in the upper panel all-sky images.

The first of these figures (Figure 8.5.1) shows data from the night of April 21, 2007 (day number 111). This day was geomagnetically very quiet, with mean $ap = 2$ and mean Dst = 5 nT. The all-sky images showed an auroral arc at the magnetic southward edge of the field-of-view increase in brightness and move magnetically equatorward to a location slightly equatorward of the station. At the beginning of this series of images the wind was directed magnetically southeastward on the equatorward side of the auroral arc, with magnitudes around 100 m.s^{-1}, while poleward of the arc the wind speed was greatly reduced, and directed southward. As the arc moved through the field-of-view between 18:33 and 20:00 UT the wind equatorward of the arc remained generally southeastward, while the wind collocated with the arc and poleward of it rotated to flow approximately magnetically westward, parallel to the arc. Between 18:59 and 19:28 UT the winds collocated with the auroral arc were directed parallel to the arc, while the wind poleward of the arc was slightly rotated relative to the arc, with a southward component to the east and a northward component to the west. The region of wind shear between the southeastward flow equatorward of the arc and the westward flow within and poleward of the arc was observed to track the location of the arc as it moved equatorward. After 20:00 UT a second arc appeared equatorward of the first, and the previously clear boundary between the two flow regions equatorward and poleward of the arc became less well defined. By 20:21 UT the wind field over most of the field-of-view had rotated to flow generally westward, except in a small number of zones on the equatorward edge of the field-of-view, where a

k	A_k (kg.m^{-1}.s^{-1})
N_2	3.43×10^{-7}
O_2	4.03×10^{-7}
O	3.90×10^{-7}
He	3.84×10^{-7}

Table 8.2: Coefficients A_k in Equation 8.5.5, from (Rees, 1989, p.126).

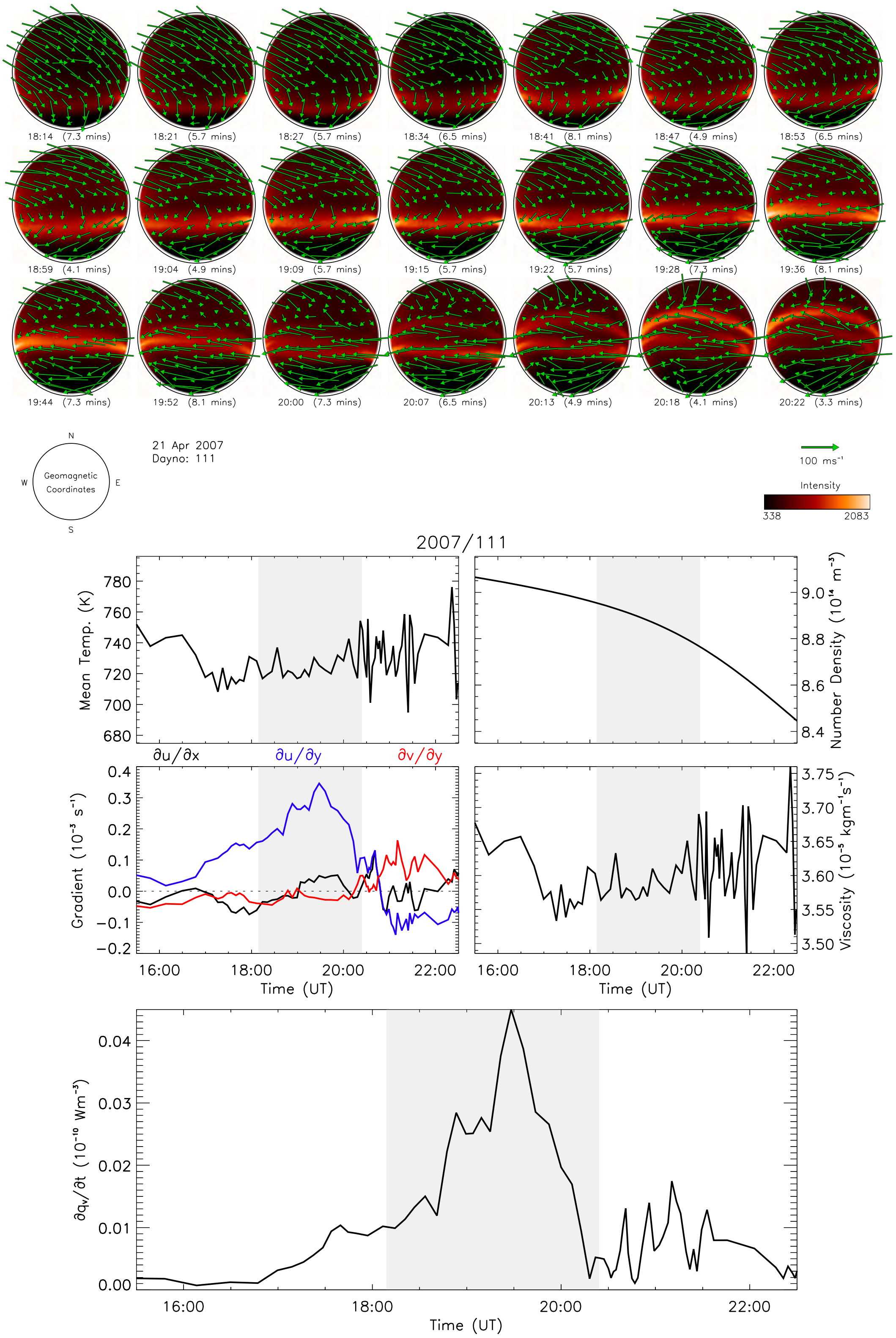


Figure 8.5.1: Top panel: fitted vector wind fields (green arrows) overlaid onto all-sky images recorded by the Mawson SDI on April 21, 2007. The vector in the central zone shows the average horizontal wind direction across the entire field-of-view. These vectors and all-sky images have not been projected onto the sky. The vector scale is shown in the lower right-hand corner, along with the image intensity scale. Lower panel: the group of four plots show the mean (all-sky) temperature, fitted gradients, modelled number density, and viscosity. The bottom plot shows the rate of viscous heating. The shaded areas show the time period displayed in the top panel.

large southward wind component observed. From the lower panel of Figure 8.5.1 the peak viscous heating occurred at 19:28 UT, with a magnitude of 0.045×10^{-10} W.m^{-3}, where the wind shear (quantified by the gradient $\partial u/\partial y$) reached its greatest magnitude due to the almost anti-parallel flow regions on opposite sides of the auroral arc.

The night of May 24, 2007 (day number 144) is shown in Figure 8.5.2. In contrast to day 111, day 144 was moderately active, with mean $ap = 30$ and mean Dst $= -28$ nT. The evolution of the wind field was however very similar to that observed on day 111. At 14:11 UT the wind was directed predominantly magnetically southward, with a slight rotation toward the magnetic west at the southward edge of the field-of-view. Wind speeds were higher than on day 111, with magnitudes around 150 m.s^{-1}, consistent with the greater level of magnetic activity (Chapter 6). By 14:43 UT an auroral arc was present in the southern half of the all-sky image, and the wind vectors collocated with this arc were rotated further westward, while the wind equatorward of the arc showed a small but noticeable rotation in the opposite direction, i.e. magnetic eastward. By 14:51 UT a strong wind shear was present along the equatorward boundary of the auroral arc, and the wind direction in the region of the arc was rotating to flow parallel to the arc. The data cadence had increased by this time due to the increased auroral brightness. The shear region tracked the motion of the auroral arc, and by 14:60 UT the auroral arc lay approximately overhead of the station, with wind flow on the equatorward edge directed approximately southeastward, while the winds within and poleward of the arc were flowing southwestward.

In the frame at 15:04 UT a localised region of increased auroral brightness was observed within the arc toward the magnetic west. Interestingly, the inferred wind field at this time shows a region of wind flow just equatorward of the auroral arc which appeared to flow around the brightened region, while within the arc (and within the region of enhanced brightness) the winds remained approximately parallel to the arc. By 15:07 UT the bright region appeared to have moved toward the westward edge of the field-of-view, and the feature of the wind field flowing around the position which had previously been collocated with the bright auroral feature was still apparent. From 15:12 UT until 15:25 UT the arc was observed to move equatorward and to increase in latitudinal extent, and the magnitude of the wind shear was reduced as the wind field became more uniform, flowing toward the magnetic southwest. The magnitude of the viscous heating rate reached slightly smaller levels than on day 111, with a maximum heating rate of around 0.033×10^{-10} W.m^{-3}. While the observed shear between the wind flows on the poleward and equatorward side of the auroral arc was noticeably less dramatic on day 144 than was observed on day 111 (where the two flow regions were close to being anti-parallel), the greater wind speeds on day 144 helped to offset this and produce similar viscous heating rates. As on day 111, the major contribution was from the zonal wind gradient in the meridional direction. The narrow spike in the viscous heating rate around 13:45 UT was also coincident with the appearance of an auroral arc along the poleward edge of the field-of-view, however this arc did not move as far into the field-of-view as the arc in the period described above. In this earlier event there was a relatively larger contribution from the zonal gradient of the

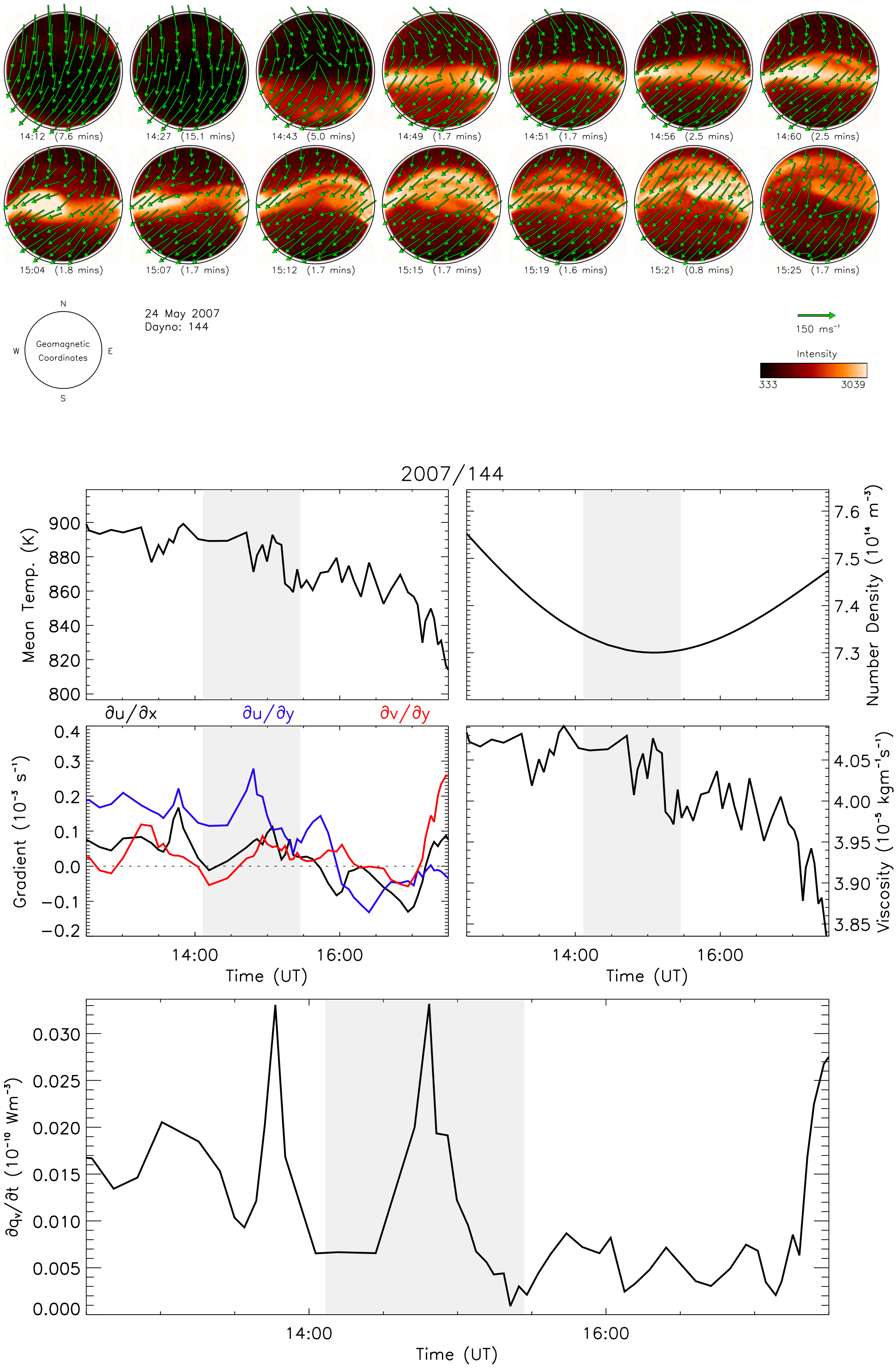


Figure 8.5.2: Same format as Figure 8.5.1, for May 24, 2007.

zonal wind (the zonal divergence).

The night of July 14, 2007 (day number 195) is shown in Figure 8.5.3. Geomagnetic activity on this night was also moderate, with mean ap = 19 and mean Dst = -20 nT. However wind speeds were much higher than on the two nights already described, with wind speeds of around 250 m.s^{-1} observed. At the start of the period, 12:19 UT, a meridional gradient in the inferred wind field was already present, with magnetic westward flow of $\sim$ 100 m.s^{-1} in the southern half of the field-of-view, and much smaller wind speeds toward the magnetic northward edge of the field-of-view, directed southwestward. An auroral arc was also present poleward of the station at this time. Over the next four exposures (12:23 - 12:36 UT) the auroral arc brightness increased, but the arc itself did not appear to move significantly. During this time the wind speed poleward of the arc steadily increased in magnitude, maintaining its westward flow direction. From 12:38 UT onwards the arc moved steadily northward. Beginning at 12:38 UT and continuing until 13:15 UT a narrow region of eastward wind flow was observed along the northern edge of the field-of-view, in the opposite direction to the flow poleward of the arc. Between 12:38 and 12:50 UT the region of shear on the equatorward edge of the arc was relatively well defined, as the wind just equatorward of the arc was generally of very small magnitude. After 12:50 UT the winds equatorward of the arc increased in speed, and by 13:02 UT a flow region connecting the dominant westward flow to the narrow band of eastward flow was apparent equatorward of the auroral arc. By 13:23 UT this narrow eastward flow region had been completely reversed, and by this time the wind field had become relatively uniform, flowing magnetic westward with speeds of around 250 m.s^{-1}.

The inferred viscous heating rate on this night peaked around 13:00 UT with a magnitude of 0.23×10^{-10} W.m^{-3}, an order of magnitude larger than either of the previous two nights which were described above. The largest contribution to the viscous heating rate was again the component describing the shear of the zonal wind, $\partial u/\partial y$. Near the end of the period displayed with the all-sky images there was an increase in the meridional divergence, which produced a small increase in the viscous heating rate around 13:35 UT, temporarily offsetting the reduction in wind shear as the wind field became more uniform. The order of magnitude increase in the heating rate was due primarily to the much higher wind speeds, which were approximately twice as large as were observed on days 111 and 144.

The final night to be presented is shown in Figure 8.5.4, the night of October 3, 2008 (day number 277). The lines which are visible in the all-sky images from this night were due to the zone map being overlaid onto the all-sky images after the completion of each exposure; these lines did not affect the spectral acquisition, and thus did not affect any subsequent analysis. There were relatively few clear examples of wind shear in the data from 2008 compared with that from 2007, which is reflected in the choice of data which has been presented above (2008 was a geomagnetically very quiet year). Day 277 showed a different type of behaviour to that which occurred on days 111, 144, and 195 of 2007. On those nights, the region of wind shear was observed on the equatorward edge of the auroral arc and seen to follow the equatorward motion of the arc. However, on day 277 of

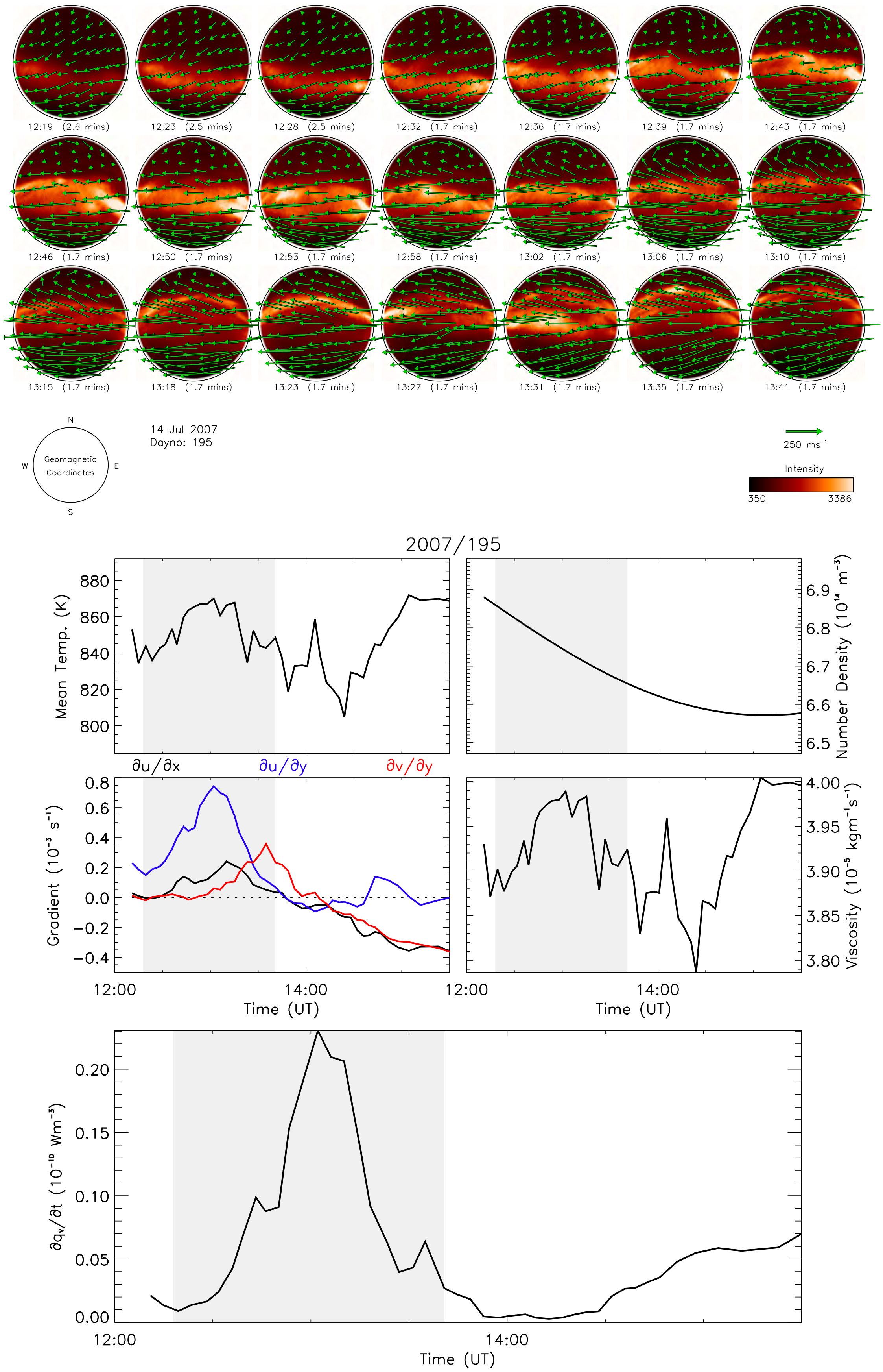


Figure 8.5.3: Same format as Figure 8.5.1, for July 14, 2007.

2008, the wind field was observed to evolve quite differently. At the start of the period, wind flow was magnetically westward and uniform across the field-of-view, with speeds of around 150 m.s^{-1}. Diffuse aurora was present over most of the field-of-view, however along the poleward edge of the images there was relatively little brightness. From 18:21 to 18:33 UT the wind in this region of low brightness was directed more westward than were the winds collocated with the diffuse aurora, while on the poleward edge of the region of diffuse aurora a brighter discrete arc was forming. By 18:41 UT the diffuse aurora which had previously covered most of the field-of-view had diminished in extent, while the auroral arc remained, located approximately overhead of the station. By this time the winds poleward of the arc had begun to decrease in magnitude, while the winds equatorward of the arc remained steady, directed approximately westward with a speed of around 150 m.s^{-1}. By 19:07 UT a bright auroral feature had appeared in the magnetic southeast, and the inferred wind field showed winds directed southward and southeastward along the poleward edge of the field-of-view, and very low wind speeds along the poleward edge of the arc. The situation was similar in the next exposure 19:10 UT, however the bright auroral feature had become less structured, and had decreased in brightness. Between 19:15 and 19:25 UT the wind field in the southern half of the field-of-view recovered, and the winds became more uniform across the entire field-of-view, with a predominantly westward flow.

The peak viscous heating rate on this night was comparable to the rates observed on days 111 and 144 of 2007. The peak rate of heating was 0.026×10^{-10} W.m^{-3}, again driven by the large meridional gradient in the zonal wind. However, in contrast to all three of the nights discussed previously, the meridional gradient of the zonal wind was negative during the period presented, with large wind speeds equatorward of the auroral features and low wind speeds poleward. The peak viscous heating rate at 19:07 UT was concurrent with the appearance of a bright auroral feature, when the inferred wind field was the least uniform. For the three nights presented from 2007, the region of faster wind flow was located within and poleward of the auroral arc, with lower wind speeds equatorward of the arc.

The viscous heating rates derived above provide a lower limit to the amount of power that must have been supplied to the atmosphere per unit volume in order to sustain the observed wind shears (assuming that all of the energy loss was due to viscous dissipation). From the wind fields displayed in this section, it would not be unreasonable to assume that the magnitude of the actual wind shear present in some cases could be at least as much as double the values that have been obtained by the vector fitting algorithm (and possibly even more), bearing in mind that this algorithm produces an average of each wind gradient across the entire field-of-view (and assumes $\partial v/\partial x = 0$). The effect on the heating rate of doubling the gradient $\partial u/\partial y$ depends on the relative contributions from each of the other gradient terms. At those times when $\partial u/\partial y$ is the dominant gradient contributing to the heating rate (which is the case when the wind fields are strongly sheared in the magnetic meridional direction), a doubling of this gradient can result in a heating rate which is almost four times higher due to the squared dependence on this term in Equation 8.5.4. At times when each of the gradient terms contributes equally, the increase in the

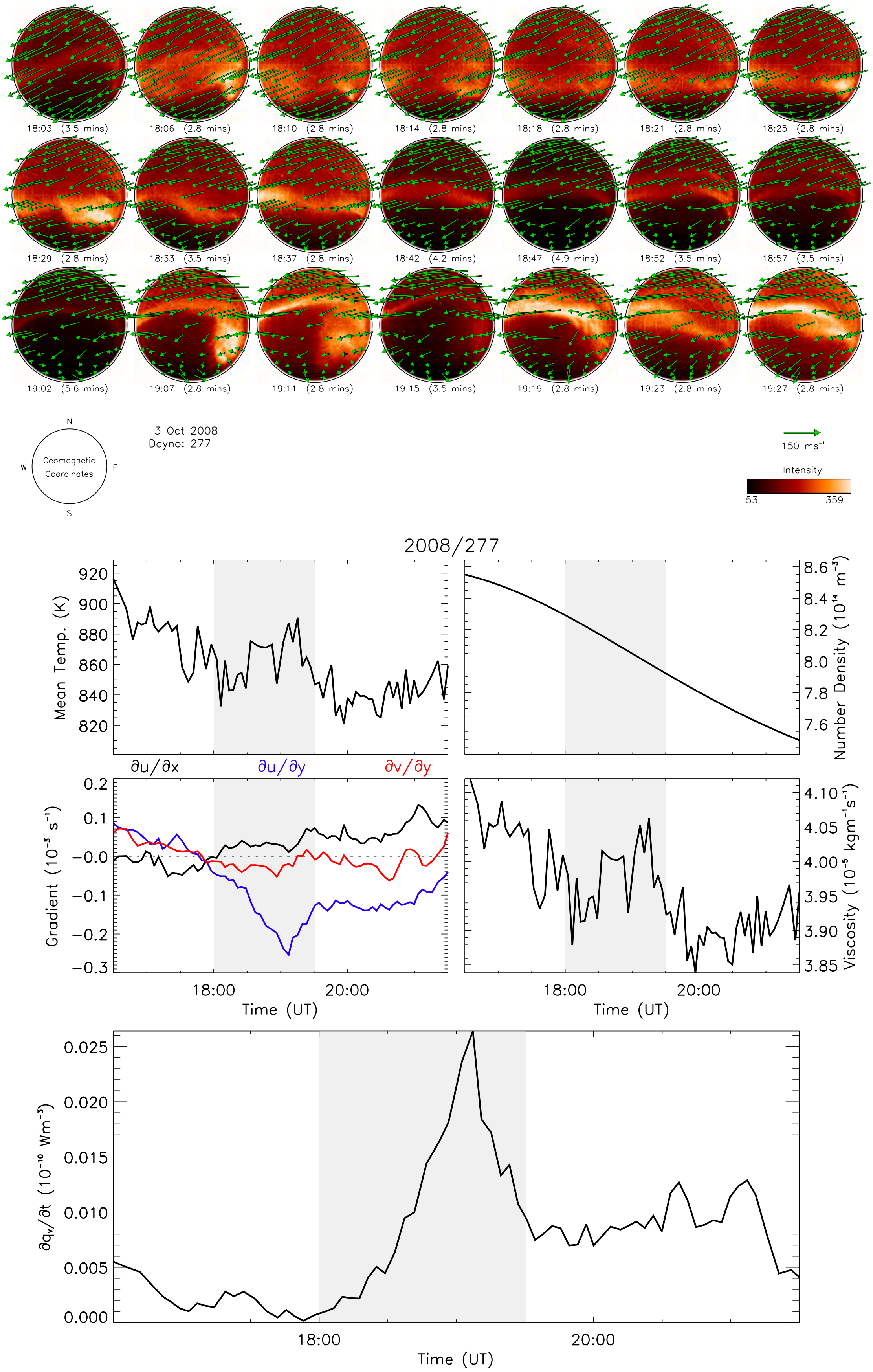


Figure 8.5.4: Same format as Figure 8.5.1, for October 3, 2008.

heating rate due to a doubling of $\partial u/\partial y$ is more modest.

The periods of sheared wind flow described above lasted between approximately 1 and 2 hours. In the context of other sources of high-latitude neutral gas heating, the derived viscous heating rates are not demanding (however these must be considered lower limits to the actual heating rates). For example, the heating rate due to solar insolation at an altitude of 250 km is on the order of 10^{-10} W.m^{-3} during solar minimum conditions (Killeen, 1987). Energy deposition due to Joule and particle heating is much more variable. Brekke and Rino (1978) derived height-resolved Joule heating rates during a disturbed (geomagnetically active) period, and obtained heating rates on the order of 10^{-7} to 10^{-6} W.m^{-3} at altitudes between 125 and 163 km. Similar Joule heating rates were calculated by Thayer (1998) during moderate to active geomagnetic conditions. These heating rates were calculated at E-region heights, where the volume energy deposition rate due to Joule heating maximises due to the enhanced conductivities in that region. The Joule heating rates (per unit volume) at F-region heights will in general be much lower than these values (however the heating rate per unit mass can actually maximise at these higher altitudes, due to the decrease in mass density with increasing altitude (Deng and Ridley, 2007)). Fujiwara et al. (2007) derived values of the Joule heating rate per unit mass in the dayside polar cap/cusp region of approximately 20 W.kg^{-1} at an altitude of $\sim$ 250 km. For an assumed mean mass density at this altitude of 2×10^{-11} kg.m^{-3}, the Joule heating rate is equal to approximately 4×10^{-10} W.m^{-3}, comparable to the solar heating rate. Fujiwara et al. (2007) also estimated the heating rate due to auroral electron and proton precipitation, and derived a heating rate per unit volume approximately one order of magnitude smaller than the Joule heating rate.

The largest rate of viscous heat dissipation derived in the present study was approximately 2.3×10^{-11} W.m^{-3}, comparable to the estimated auroral particle heating rate of Fujiwara et al. (2007). From the time at which this peak heating rate occurred (13:00 UT on day 195 of 2007) the station was located in the early magnetic afternoon sector, near 14:20 MLT, and thus was not located directly within the dayside cusp region. For three of the nights presented in the present study the peak viscous heating rate was approximately 3×10^{-12} W.m^{-3}, one order of magnitude smaller than the particle heating rate estimated by Fujiwara et al. (2007) and two orders of magnitude smaller than the Joule heating rate derived by these authors (op. cit.). Though again the values of viscous heating rate derived in the present work should be regarded as lower limits to the actual heating rate, which, due to the squared dependence on the magnitude of the wind shear ($\partial u/\partial y$), could easily have been a factor of four times greater than the derived values.

The periods of wind shear observed during the nights presented above were often highly spatially localised, particularly on nights 111 and 144 of 2007, where sharp reversals in wind direction were observed in adjacent zones, separated by 100 km or less (at an altitude of 240 km), and the most localised shear regions were observed on the magnetic equatorward edge of bright auroral arcs. Indeed the periods of strong, localised wind shear in the entire two-year data-set were almost exclusively observed to be coincident with the appearance of bright auroral arcs within the field-of-view. This relationship with bright aurora has also been observed by a SDI located at Poker Flat (Conde, M., private communication, 2010). This is not surprising given the localised nature of the thermospheric energy inputs into the F-region thermosphere in the vicinity of auroral arcs. These data show that strong localised wind shears can exist over spatial scales of less than $\sim$ 100 km, and that, in the lower limit, only a modest amount of power is required to drive them, relative to the available energy sources in the auroral F-region.

8.6 Summary

In this chapter results have been presented of an investigation into the relationship between vertical winds and the divergence of the horizontal wind field. Above Davis, an average upwelling observed between 21:00 and 03:00 MLT was compared with the divergence of the observed average horizontal winds (Section 8.2). The average time variation of the divergence was found to closely follow the variations in the average vertical wind during quiet conditions, while during active conditions the correlation between the average vertical wind and the horizontal divergence was not as good. In general there were more vertical wind/divergence data pairs which were in agreement with the Burnside relation during active conditions. During quiet times, the ratio of the average vertical wind to the average horizontal divergence was observed to be larger than predicted by Burnside, however there was less spread in the data during these quiet conditions. Comparison with vertical wind activity maps derived by Innis and Conde (2002) revealed that, at least during active conditions, the region of recurrent upwelling observed above Davis was collocated with the magnetic local time sector of maximum vertical wind activity (for the vertical wind activity maps which did not include data filtering by magnetic activity, altitude or solar zenith angle).

A comparison was also performed between vertical winds and horizontal divergence using five years of data from the Davis FPS and two years of data from the Mawson SDI (Section 8.3). In this study the data were binned by divergence and the median vertical wind within each divergence bin was calculated. From these data were derived the scale height which would be required by the Burnside relation to produce the given (best-fit) trend in the vertical wind as a function of the horizontal divergence. The comparison confirmed the observation of Section 8.2 that the Burnside relation was more accurate during periods of greater geomagnetic activity at Davis. In general the agreement between the observed vertical wind/divergence trend and that predicted by the Burnside relation was quite good at Davis, while at Mawson the agreement was poor, likely due to the fact that Mawson is often underneath the auroral oval, where energy inputs can be highly spatially localised and rapidly time varying. At both stations there was a small range of divergence values over which the trend appeared reasonably linear, while outside this range the scatter in the data increased significantly.

Section 8.4 presented an investigation into divergence on a smaller scale. Whereas the previous two sections had estimated the divergence over essentially the entire field-of-view, the aim of this section was to attempt to derive information on the spatial variability of the divergence above Mawson on one night for which common-volume vertical wind measurements were available from the bistatic campaign described in Chapter 7. Despite the fact the Mawson SDI could only uniquely resolve the component of horizontal wind divergence parallel to the radial line-of-sight (since the wind component perpendicular to this cannot in general be measured from a single site), a clear correspondence was observed between the inferred divergence maps and the morphology of a bright auroral feature which was observed to move rapidly through the field-of-view. The average horizontal divergence calculated within each of the common-volume regions was then compared to the vertical winds derived in those regions using the bistatic arrangement of Chapter 7. The result of this comparison was not conclusive. While some correlation was observed between the divergence calculated in different common-volume regions during a vertical wind disturbance which was coincident with the presence of the bright auroral feature mentioned above, the divergence within a given common-volume region did not show a clear relationship with the derived vertical winds within that region. This particular type of analysis was clearly limited by the inability to uniquely resolve the horizontal divergence in two orthogonal directions.

In the case study presented in Section 8.4 it was noted that a narrow region of wind shear was observed along the magnetic equatorward edge of the field-of-view which increased in magnitude as the auroral feature moved through the field-of-view. Wind shears of this type were observed on multiple nights, almost exclusively related to the appearance of bright auroral arcs within the field-of-view. In Section 8.5, four case studies were presented which showed examples of some of the types of sheared wind fields which were observed, and an estimate of the power dissipated through viscous interactions (and thus an estimate of the power input required to drive the sheared wind fields) was derived. The observed wind shears were often highly localised, spanning regions less than ~ 100 km in width at an assumed altitude of 240 km. The largest estimated rate of viscous heat dissipation was 2.3×10^{-11} W.m^{-3}, and it was noted that this in all likelihood represented a lower-limit estimate of the actual viscous heating rate, which could easily have been four times larger than the derived values. This lower limit was approximately one order of magnitude smaller than estimates of the solar heating rate from Killeen (1987) and the Joule heating rate derived by Fujiwara et al. (2007), and comparable to the heating rate due to particle precipitation estimated by Fujiwara et al. (2007).

Conclusions and Future Work

The experimental aim of this project was to investigate small-scale structure in thermospheric winds at southern auroral latitudes, in order to test the hypothesis that small spatial scales (< 500 km) play an important role in thermospheric dynamics. To facilitate this aim, a new all-sky, scanning imaging Fabry-Perot spectrometer was installed at Mawson station, Antarctica, and software was written to operate and automate the instrument's spectroscopic observations. Data from this instrument were used to infer two-dimensional horizontal wind fields, temperatures and emission intensities from observations of the 630.0 nm airglow line of atomic oxygen. These data, in conjunction with thermospheric wind measurements from a narrow-field imaging Fabry-Perot spectrometer operating at Davis station (approximately 635 km east of Mawson), were analysed and the results presented in this work.

The ability to simultaneously measure winds and temperatures at many tens of locations across the sky represents a significant advance in studying upper atmospheric dynamics, particularly at high latitudes where driving forces are expected to be strong and variable. Only two other comparable instruments are discussed in the literature, those operating at Poker Flat (e.g. Conde and Smith, 1998; Conde et al., 2001), and Svalbard (e.g. Griffin et al., 2008; Aruliah et al., 2010). The current work represents a major step forward in the analysis of data from these types of instruments. The principle findings from this study are summarised below.

Daily Variability and Average Behaviour

Generally, the large-scale thermospheric dynamics observed above Mawson (and presented in Chapter 6) were in agreement with previous studies of auroral-latitude neutral dynamics. With increasing geomagnetic activity, both wind magnitudes and temperatures were observed to increase, and the signature of ion-convection forcing often became dominant in the magnetic dusk-sector wind flow, where strong sunward flow was observed in contrast to the antisunward flow observed during quiet times. Equatorward flow was observed around magnetic midnight under all conditions, where ion-drag and pressure-gradient forcing would be expected to reinforce each other and drive the wind antisunward. These findings are in close agreement with the many previous studies cited in Chapter 6.

The signature of ion-forcing was also observed after magnetic midnight, a feature which has *not* often been observed above Mawson (one exception being Smith et al. (1998)). The lack or weakness of the dawnside circulation is often attributed to the influence of the Coriolis force, which opposes the formation of the dawnside convection cell while reinforcing it on the duskside (Thayer and Killeen, 1993). In the present work this feature was observed even under quiet geomagnetic conditions, and suggests that even under these conditions ion-neutral coupling was sufficiently strong to dominate the force balance in the early magnetic morning sector. This dawnside convection signature also showed a strong latitudinal variation, whereby the magnitude of the eastward wind component increased with decreasing absolute magnetic latitude (i.e. increased in the magnetic equatorward direction), from a negligible eastward flow poleward of Mawson to a dominantly eastward flow equatorward of the station.

Common-Volume Vertical Winds

Vertical winds present a problem for ground-based observatories, since the measurement of line-of-sight winds necessarily incorporates a component due to vertical motion, a component whose contribution increases with decreasing zenith angle. From a single station, measurement of the vertical wind is generally only possible in the station zenith, and therefore the removal of the vertical contribution from the off-zenith line-of-sight wind measurements requires some assumption to be made about the spatial variation of the vertical wind. One assumption frequently made (and made in the present work when inferring horizontal wind fields) is that in comparison with the horizontal wind speed, the vertical wind contribution can be safely neglected. However, observations of very large vertical winds (comparable in magnitude to the horizontal wind) have been accumulating for many years (see Chapter 7 and references therein). Therefore, at least in the auroral zone where strong sources of local heating are common, we must question the validity of this assumption. While the frequency of occurrence of very large vertical winds is low, they will still present a problem at isolated times. Relatively little is known about the spatial variability of the vertical wind field, and results from studies investigating the correlation between vertical winds measured from spatially separated locations have been mixed.

The bistatic experiment described in Chapter 7 has improved our understanding of the vertical wind field at auroral latitudes, by following the time evolution of the vertical wind at five locations along a baseline extending for approximately 635 km. These measurements showed that vertical winds could at times be correlated over horizontal scales of $\sim$ 160-480 km, while at other times there was little correlation over the smallest separations of $\sim$ 160 km. The presence of active aurora often heralded vertical wind disturbances (as was observed in the daily wind plots of Chapter 6, where intensity enhancements were very frequently associated with stronger and more variable vertical winds above Mawson). The strongest vertical winds measured during the bistatic experiment were observed in conjunction with bright, rapidly moving auroral arcs and strongly perturbed horizontal wind fields, and correspondingly large horizontal wind gradients.

The investigation of Chapter 7 also showed that the vertical winds above Mawson and Davis stations responded very differently on average to increased auroral brightness. At both stations, increased emission intensity corresponded to increased levels of geomagnetic activity. At Mawson, increased intensity was linked to larger vertical wind magnitudes on average, however at Davis, only the variability increased - average vertical wind magnitudes were not affected by increased brightness/activity. Thus between the two stations, separated by approximately 635 km (predominantly in geomagnetic latitude), quite dissimilar average vertical wind behaviour was observed as geomagnetic activity increased, presumably due to each station's location relative to the auroral oval.

Divergence and Local-Scale Structure

Given the difficulty of measuring vertical winds, there is a strong motivation to investigate indirect methods of inferring vertical winds from other, more accessible, measurements. The Burnside relation, given by Equation 8.1.1 in Chapter 8, is one such method, linking the vertical wind to the divergence of the horizontal wind, subject to a number of assumptions (outlined in Section 8.1). Using this relation, the vertical wind is expected to be directly proportional to the horizontal divergence (the constant of proportionality being the atmospheric scale height), and therefore upward vertical winds are expected to be associated with positive divergence in the horizontal wind, while downward vertical winds should be associated with convergent wind flow.

Two studies investigating the large-scale validity of the Burnside relation were presented in Chapter 8. Data from two years of Mawson observations and five years of Davis observations were analysed, providing a substantial data set at two separate locations. A quantitative estimate of the validity of the Burnside relation was made by comparing the gradients of straight lines fitted through vertical wind/divergence scatter plots with the 'true' atmospheric scale height estimated using the NRLMSISE-00 empirical model.

These results indicated that the average, large-scale ($\sim$ 800 km) behaviour of the vertical winds above both Mawson and Davis were in quite close agreement with the Burnside relation. Systematic departures from the Burnside relation were observed as the vertical wind magnitude increased. At Davis, observations made under increased levels of geomagnetic activity resulted in larger 'fitted' scale heights compared with quiet conditions, however the opposite was observed above Mawson. It is very likely that this difference in behaviour is related to the assumptions underlying the Burnside derivation, and the ways in which these assumptions depend for their validity on a station's location relative to auroral energy sources.

On a local scale, the Burnside relation was used to generate estimates of the vertical wind at the three common-volume locations described in Chapter 7. These estimates were then compared with the bistatic vertical wind measurements from one of the case studies presented in Chapter 7. In this case the vertical wind estimates generated using the Burnside relation were not in agreement with the measured vertical winds. While it is not possible to draw general conclusions from this single night comparison, it would appear that either the method used to infer wind gradients at this scale was limited by a lack of information about the wind gradient normal to the line-of-sight (since only the radial wind component was directly measured), or that the assumptions underlying the Burnside relation were not valid in this case, over these small spatial scales (equivalent to one of the Mawson SDI viewing zones, $\sim$ 100 km). In addition, the modelling results of

Cooper et al. (2009) might suggest that either the sampled wind altitude was too close to the altitude of maximum energy deposition (per unit mass), or that the momentum inputs to the thermosphere during this period were more rapidly time-varying than the Burnside relation could reasonably be expected to model (which is unlikely given that the disagreement persisted over the entire observation period). It is likely that a combination of these factors contributed to the poor agreement between the modelled and measured vertical wind, however a much larger data set than is currently available would be required to determine this.

The final study presented in Chapter 8 was an investigation into the strongly sheared wind fields which were sometimes observed above Mawson, almost exclusively in the presence of bright auroral arcs. The observed shears were often highly localised, spanning regions less than ~ 100 km in width at 240 km altitude. Under the assumption that the only dissipative process acting was due to viscosity, the viscous heating rate was calculated during four case studies. The calculated heating rates ranged from 2.6×10^{-12} W.m^{-3} to 2.3×10^{-11} W.m^{-3}, however these heating rates were almost certainly lower limits to the actual rate of (viscous) energy dissipation, since the measured shears were all-sky averages, and thus the true magnitude of the shear could easily have been much larger (and heating rates correspondingly higher). Thus viscous dissipation could be a much more important term in the energy balance than is typically appreciated.

Conclusion

From the results presented in this work it is evident that spatial scales smaller than ~ 500 km do indeed play an important role in thermospheric dynamics. Structure in the wind and temperature fields was often observed down to at least the separation between adjacent zones (~ 100 km). Strongly sheared wind fields were observed in the presence of bright, latitudinally narrow auroral arcs, and the heating rate of such sheared wind fields was estimated to be on the same order as that due to particle precipitation. The vertical wind field was observed to show varying degrees of correlation between spatially separated locations; sometimes correlated over distances of $\sim 160-480$ km, at other times showing no correlation over the smallest separations (~ 160 km). The 635 km separation between Mawson and Davis station was sufficient to produce very different behaviour in the average vertical winds measured at each station under different levels of geomagnetic activity. While the large-scale (~ 800 km), average behaviour of the vertical wind at each station was reasonably well modelled by the Burnside relation, the ability of this relation to model the observed vertical winds on the scale of a single SDI viewing zone (~ 100 km) was poor.

These results all indicate that an understanding of the small-scale spatial structure of the thermosphere is very important for improving our understanding of the upper atmosphere as a whole. In particular, as the spatial resolution of complex first-principles based atmospheric models is improved, the availability of local-scale thermospheric observations will be crucial for testing our ability to characterise the important physical processes occurring in the real atmosphere.

8.7 Future Work

Future work will continue to focus on investigating the small-scale features of the wind and temperature fields. One promising area of investigation is in comparing thermospheric dynamics at two different altitudes. The Mawson SDI uses a filter-wheel to select the airglow emission wavelength of interest. Data have already been acquired from the Mawson SDI at a wavelength of 557.7 nm, which is emitted from atomic oxygen in the lower thermosphere. The emission height of the 557.7 nm line is highly variable in the auroral zone, and horizontal winds in the lower thermosphere often show strong vertical gradients, thus complicating the analysis of data at this wavelength. However, if the height characteristics of the emission are known or can be inferred, it is possible to directly compare the response of the wind-field to auroral forcing at two different altitudes by interleaving observations of the 630.0 and 557.7 nm emission lines. Such a comparison has recently been presented using data from the Mawson SDI (Kosch et al., 2010).

In Alaska, the deployment of multiple scanning Doppler imagers with overlapping fields-of-view has opened up new possibilities for inferring vector wind fields, through the use of bistatic and tristatic wind estimates. In the bistatic case, vertical wind measurements at multiple locations between the instruments are possible, with high time resolution. Horizontal wind estimates can also be obtained, through suitable assumptions about the local vertical wind. With tristatic data, all three components of the vector wind are estimated directly, requiring no assumptions to be made. This will allow further investigation into the validity of the vector wind fitting method outlined in the present work, through independent measurement of the vector field in regions of overlap between instrument fields-of-view.

The presence of multiple SDI's will greatly extend the spatial coverage of wind and temperature measurements, and allow for more detailed investigations into the climatology of both lower and upper thermospheric winds and temperatures. In addition, it will be possible to track the propagation of gravity waves by comparing time-series of essentially simultaneous wind measurements from many independent viewing directions ($>$ 200 for the two instruments currently operating in Alaska). Parameterizing the contribution of gravity waves to thermospheric dynamics, energetics and composition will be an important area of future research.

Bibliography

Abreu, V. J., Schmitt, G. A., Hays, P. B., and Dachev, T. P.: Volume emission rate profiles of the 6300-Å tropical nightglow obtained from the AE-E satellite: Latitudinal and seasonal variations, J. Geophys. Res., 87, 6364–6352, 1982.

Adkins, C. J.: Equilibrium Thermodynamics, European Physics Series, McGraw-Hill, second edn., 1975.

Ahn, B.-H., Chen, G. X., Sun, W., Gjerloev, J. W., Kamide, Y., Sigwarth, J. B., and Frank, L. A.: Equatorward expansion of the westward electrojet during magnetically disturbed periods, J. Geophys. Res., 110, A01 305, doi:10.1029/2004JA010553, 2005.

Aikio, A. T., Pitkänen, T., Kozlovsky, A., and Amm, O.: Method to locate the polar cap boundary in the nightside ionosphere and application to a substorm event, Ann. Geophys., 24, 1905–1917, 2006.

Akasofu, S.-I.: Energy coupling between the solar wind and the magnetosphere, Space Sci. Rev., 28, 121–190, 1981.

Akasofu, S.-I.: Several 'controversial' issues on substorms, Space Sci. Rev., 113, 1–40, 2004.

Aksnes, A., Stadsnes, J., Lu, G., Østgaard, N., Vondrak, R. R., Detrick, D. L., Rosenberg, T. J., Germany, G. A., and Schulz, M.: Effects of energetic electrons on the electrodynamics in the ionosphere, Ann. Geophys., 22, 475–496, 2004.

Alfvén, H.: Cosmical electrodynamics, Oxford University Press, New York, 1950.

Amm, O., Donovan, E. F., Frey, H., Lester, M., Nakamura, R., Wild, J. A., Aikio, A., Dunlop, M., Kauritie, K., Marchaudon, A., McCrea, I. W., Opgenoorth, H. J., and Stromme, A.: Coordinated studies of the geospace environment using Cluster, satellite and ground-based data: an interim review, Ann. Geophys., 23, 2129–2170, 2005.

Anderson, H. and Vondrak, R. R.: Observations of Birkeland currents at auroral latitudes, Rev. Geophys. Space Phys., 13, 243–262, 1975.

Angelopoulos, V., McFadden, J. P., Larson, D., Carlson, C. W., Mende, S. B., Frey, H., Phan, T., Sibeck, D. G., Glassmeier, K.-H., Auster, U., Donovan, E., Mann, I. R., Rae, I. J., Russell, C. T., Runov, A., Zhou, X.-Z., and Kepko, L.: Tail Reconnection Triggering Substorm Onset, Science, 321, 931–935, 2008.

Aruliah, A. L. and Griffin, E.: Evidence of meso-scale structure in the high-latitude thermosphere, Ann. Geophys., 19, 37–46, 2001.

Aruliah, A. L. and Rees, D.: The trouble with thermospheric vertical winds: geomagnetic, seasonal and solar cycle dependence at high latitudes, J. Atmos. Terr. Phys., 57, 597–609, 1995.

Aruliah, A. L., Rees, D., and Fuller-Rowell, T. J.: The combined effect of solar and geomagnetic activity on high latitude thermospheric neutral winds. Part I. Observations, J. Atmos. Terr. Phys., 53, 467–483, 1991.

Aruliah, A. L., Griffin, E. M., Aylward, A. D., Ford, E. A. K., Kosch, M. J., Davis, C. J., Howells, V. S. C., Pryse, S. E., Middleton, H. R., and Jussila, J.: First direct evidence of meso-scale variability on ion-neutral dynamics using co-located tristatic FPIs and EISCAT radar in Northern Scandinavia, Ann. Geophys., 23, 147–162, 2005.

Aruliah, A. L., Griffin, E. M., Yiu, H.-C. I., McWhirter, I., and Charalambous, A.: SCANDI - an all-sky Doppler imager for studies of thermospheric spatial structure, Ann. Geophys., 28, 549–567, 2010.

Baker, D. N., Klimas, A. J., Vassiliadis, D., Pulkkinen, T. I., and McPherron, R. L.: Reexamination of driven and unloading aspects of magnetospheric substorms, J. Geophys. Res., 102, 7169–7177, 1997.

Baker, D. N., Pulkkinen, T. I., Büchner, J., and Klimas, A. J.: Substorms: A global instability of the magnetosphere-ionosphere system, J. Geophys. Res., 104, 14,601–14,611, 1999.

Baker, D. N., Barth, C. A., Mankoff, K. E., Kanekal, S. G., Bailey, S. M., Mason, G. M., and Mazur, J. E.: Relationships between precipitating auroral zone electrons and lower thermospheric nitric oxide densities: 1998-2000, J. Geophys. Res., 106, 24,465–24,480, 2001.

Baker, D. N., Peterson, W. K., Eriksson, S., Li, X., Blake, J. B., Burch, J. L., Daly, P. W., Dunlop, M. W., Korth, A., Donovan, E., Friedel, R., Fritz, T. A., Frey, H. U., Mende, S. B., Roeder, J., and Singer, H. J.: Timing of magnetic reconnection initiation during a global magnetospheric substorm onset, Geophys. Res. Lett., 29, 2190, doi:10.1029/2002GL015539, 2002.

Baker, J. B. H., Zhang, Y., Greenwald, R. A., Paxton, L. J., and Morrison, D.: Height-integrated Joule and auroral particle heat in the night side high latitude thermosphere, Geophys. Res. Lett., 31, doi:10.1029/2004GL019535, 2004.

Baker, J. B. H., Greenwald, R. A., Ruohoniemi, J. M., Oksavik, K., Gjerloev, J. W., Paxton, L. J., and Hairston, M. R.: Observations of ionospheric convection from the Wallops SuperDARN radar at middle latitudes, J. Geophys. Res., 112, A01 303, doi:10.1029/2006JA011982, 2007.

Banks, P. M.: Ionosphere-Magnetosphere coupling 2. Electric fields, Rev. Geophys. Space Phys., 13, 874–878, 1975.

Barth, C. A., Mankoff, K. D., Bailey, S. M., and Solomon, S. C.: Global observations of nitric oxide in the thermosphere, J. Geophys. Res., 108, 1027, doi:10.1029/2002JA009458, 2003.

Bates, D. R.: Forbidden oxygen and nitrogen lines in the nightglow, Planet. Space Sci., 26, 897–912, 1978.

Batten, S. and Rees, D.: Thermospheric winds in the auroral oval: Observations of small scale structures and rapid fluctuations by a Doppler imaging system, Planet. Space Sci., 38, 675–694, 1990.

Batten, S., Fuller-Rowell, T. J., and Rees, D.: A Numerical Data Base for VAX and Personal Computers for Storage, Reconstruction, and Display of Global Thermospheric and Ionospheric Models, Planet. Space Sci., 35w, 1167, 1987.

Baumgardner, J., Wrotten, J., Semeter, J., Kozyra, J., Buonsanto, M., Erickson, P., and Mendillo, M.: A very bright SAR arc: implications for extreme magnetosphere-ionosphere coupling, Ann. Geophys., 25, 25932608, 2007.

Beer, T.: Atmospheric Waves, Halsted Press, 1974.

Bilitza, D.: International reference ionosphere 2000, Radio Sci., 36, 261–275, 2001.

Biondi, M. A.: Measured vertical motion and converging and diverging horizontal flow of the midlatitude thermosphere, Geophys. Res. Lett., 11, 84–87, 1984.

Biondi, M. A., Sipler, D. P., Zipf, M. E., and Baumgardner, J. L.: All-sky Doppler interferometer for thermospheric dynamics studies, Appl. Opt., 34, 1646–1654, 1995.

Birkeland, K.: The Norwegian Aurora Polaris Expedition 1902-1903, vol. 1, Christiania, Norway, 1908.

Birn, J. and Hesse, M.: The substorm current wedge and field-aligned currents in MHD simulations of magnetotail reconnection, J. Geophys. Res., 96, 1611–1618, 1991.

Blockx, C., Gérard, J.-C., Coumans, V., Hubert, B., and Meurant, M.: Contributions of the driven process and the loading-unloading process during substorms: A study based on the IMAGE-SI12 imager, J. Geophys. Res., 114, A02 209, doi:10.1029/2008JA013280, 2009.

Boyle, C. B., Reiff, P. H., and Hairston, M. R.: Empirical polar cap potentials, J. Geophys. Res., 102, 111–125, 1997.

Brekke, A.: Physics of the Upper Polar Atmosphere, Wiley-Praxis Series in Atmospheric Physics, John Wiley & Sons Ltd, 1997.

Brekke, A. and Rino, C. L.: High-resolution altitude profiles of the auroral zone energy dissipation due to ionospheric currents, J. Geophys. Res., 83, 2517–2524, 1978.

Browning, K. A. and Wexler, R.: The determination of kinematic properties of a wind field using Doppler radar, J. Appl. Meteorol., 7, 105–113, 1968.

Burns, A. G., Solomon, S. C., Wang, W., and Killeen, T. L.: The ionospheric and thermospheric response to CMEs: Challenges and successes, J. Atmos. Solar-Terr. Phys., 69, 77–85, doi:doi:10.1016/j.jastp.2006.06.010, 2007.

Burnside, R. G., Herrero, F. A., Meriwether Jr., J. W., and Walker, J. C. G.: Optical observations of thermospheric dynamics at Arecibo, J. Geophys. Res., 86, 5532–5540, 1981.

Cannon, P. S., Reinisch, B. W., Buchau, J., and Bullett, T. W.: Response of the Polar Cap F Region Convection Direction to Changes in the Interplanetary Magnetic Field: Digisonde Measurements from Northern Greenland, J. Geophys. Res., 96, 1239–1250, 1991.

Chua, D., Parks, G., Brittnacher, M., Germany, G., and Spann, J.: Auroral substorm timescales: IMF and seasonal variations, J. Geophys. Res., 109, A03 207, doi: 10.1029/2003JA009951, 2004.

Chun, F. K., Knipp, D. J., McHarg, M. G., Lu, G., Emery, B. A., Vennerstrøm, S., and Troshichev, O. A.: Polar cap index as a proxy for hemispheric Joule heating, Geophys. Res. Lett., 26, 1101–1104, 1999.

Chun, F. K., Knipp, D. J., McHarg, M. G., Lacey, J. R., Lu, G., and Emery, B. A.: Joule heating patterns as a function of polar cap index, J. Geophys. Res., 107, 1119, doi:10.1029/2001JA000246, 2002.

Codrescu, M. V., Fuller-Rowell, T. J., and Foster, J. C.: On the importance of E-field variability for Joule heating in the high-latitude thermosphere, Geophys. Res. Lett., 22, 2393–2396, doi:doi:10.1029/95GL01909, 1995.

Cole, K. D.: Joule heating of the upper atmosphere, Aust. J. Phys., 15, 223–235, 1962.

Cole, K. D.: Stable auroral red arcs, sinks for energy of D_{st} main phase, J. Geophys. Res., 70, 1689–1706, 1965.

Conde, M.: Analysis of Fabry-Perot spectra of lidar backscatter echoes, in: ANARE Reports, edited by Morris, R. J. and Wilkinson, P. J., 146, pp. 91–114, Antarctic Division, Department of the Environment and Heritage, 2001.

Conde, M.: Deriving Wavelength Spectra from Fringe Images from a Fixed-Gap Single-Etalon Fabry-Perot Spectrometer, Appl. Opt., 41, 2672–2678, 2002.

Conde, M. and Dyson, P. L.: Thermospheric horizontal winds above Mawson, Antarctica, Adv. Space Res., 16, 41–52, 1995a.

Conde, M. and Dyson, P. L.: Thermospheric vertical winds above Mawson, Antarctica, J. Atmos. Terr. Phys., 57, 589–596, 1995b.

Conde, M. and Smith, R. W.: 'Phase compensation' of a separation scanned, all-sky imaging Fabry-Perot spectrometer for auroral studies, Appl. Opt., 36, 5441–5450, 1997.

Conde, M. and Smith, R. W.: Spatial structure in the thermospheric horizontal wind above Poker Flat, Alaska, during solar minimum, J. Geophys. Res., 103(A5), 9449–9472, 1998.

Conde, M., Craven, J. D., Immel, T., Hoch, E., Stenbaek-Nielsen, H., Hallinan, T., Smith, R. W., Olson, J., Sun, W., Frank, L. A., and Sigwarth, J.: Assimilated observations of thermospheric winds, the aurora, and ionospheric currents over Alaska, J. Geophys. Res., 106, 10,493–10,508, 2001.

Conde, M. G.: Spectroscopic Observations of the λ630 nm Thermospheric Emission from Mawson, Antarctica, Under Daylight, Twilight and Night-time Observing Conditions, Ph.D. thesis, The Mawson Institute for Antarctic Research, University of Adelaide, 1990.

Cooper, S. L., Conde, M., and Dyson, P.: Numerical simulations of thermospheric dynamics: divergence as a proxy for vertical winds, Ann. Geophys., 27, 2491–2502, 2009.

Cosgrove, R. B. and Thayer, J. P.: Parametric dependence of electric field variability in the Sondrestrom database: A linear relation with Kp, J. Geophys. Res., 111, A10 313, doi:10.1029/2006JA011658, 2006.

Crickmore, R. I., Dudeney, J. R., and Rodger, A. S.: Vertical thermospheric winds at the equatorward edge of the auroral oval, J. Atmos. Terr. Phys., 53, 485–492, 1991.

Crooker, N. U., Shodhan, S., Gosling, J. T., Simmerer, J., Lepping, R. P., Steinberg, J. T., and Kahler, S. W.: Density extremes in the solar wind, Geophys. Res. Lett., 27, 3769–3772, 2000.

Crowley, G., Schoendorf, J., Roble, R. G., and Marcos, F. A.: Cellular structures in the high-latitude thermosphere, J. Geophys. Res., 101, 211–223, 1996.

Culot, F. C., Lathuillére, C., and Lilensten, J.: Influence of geomagnetic activity on the OI 630.0 and 557.7 nm dayglow, J. Geophys. Res., 110, A01 304, doi: 10.1029/2004JA010667, 2005.

Cyr, O. C. S., Howard, R. A., Sheeley Jr., N. R., Plunkett, S. P., Michels, D. J., Paswaters, S. E., Koomen, M. J., Simnett, G. M., Thompson, B. J., Gurman, J. B., Schwenn, R., Webb, D. F., Hildner, E., and Lamy, P. L.: Properties of coronal mass ejections: SOHO LASCO observations from January 1996 to June 1998, J. Geophys. Res., 105, 18,169–18,185, 2000.

de Deuge, M. A., Greet, P. A., and Jacka, F.: Optical observations of gravity waves in the auroral zone, J. Atmos. Terr. Phys., 56, 617–629, 1994.

de Meneses, F. C., Muralikrishna, P., and Clemesha, B. R.: Height profiles of OI 630 nm and OI 557.7 nm airglow intensities measured via rocket-borne photometers and estimated using electron density data: a comparison, Geofis. Int., 47, 161–166, 2008.

Deng, Y. and Ridley, A. J.: Possible reasons for underestimating Joule heating in global models: E field variability, spatial resolution, and vertical velocity, J. Geophys. Res., 112, A09 308, doi:10.1029/2006JA012006, 2007.

Dickinson, R. E., Ridley, E. C., and Roble, R. G.: A Three-dimensional General Circulation Model of the Thermosphere, J. Geophys. Res., 86, 1499–1512, doi: 10.1029/JA086iA03p01499, 1981.

Djuth, F. T., Sulzer, M. P., Gonzáles, S. A., Mathews, J. D., Elder, J. H., and Walterscheid, R. L.: A continuum of gravity waves in the Arecibo thermosphere?, Geophys. Res. Lett., 31, L16 801, doi:10.1029/2003GL019376, 2004.

Dobbin, A. L., Aylward, A. D., and Harris, M. J.: Three-dimensional GCM modeling of nitric oxide in the lower thermosphere, J. Geophys. Res., 111, A07 314, doi: 10.1029/2005JA011543, 2006a.

Dobbin, A. L., Griffin, E. M., Aylward, A. D., and Millward, G. H.: 3-D GCM modelling of thermospheric nitric oxide during the 2003 Halloween storm, Ann. Geophys., 24, 2403–2412, 2006b.

Elphinstone, R. D., Murphree, J. S., and Cogger, L. L.: What is a global auroral substorm?, Rev. Geophys., 34, 169–232, 1996.

Emmert, J. T., Faivre, M. L., Hernandez, G., Jarvis, M. J., Meriwether, J. W., Niciejewski, R. J., Sipler, D. P., and Tepley, C. A.: Climatologies of nighttime upper thermospheric winds measured by ground-based Fabry-Perot interferometers during geomagnetically quiet conditions: 1. Local time, latitudinal, seasonal, and solar cycle dependence, J. Geophys. Res., 111, 2006a.

Emmert, J. T., Hernandez, G., Jarvis, M. J., Niciejewski, R. J., Sipler, D. P., and Vennerstrom, S.: Climatologies of nighttime upper thermospheric winds measured by ground-based Fabry-Perot interferometers during geomagnetically quiet conditions: 2. High-Latitude circulation and interplanetary magnetic field dependence, J. Geophys. Res., 111, 2006b.

Etemadi, A., Cowley, S. W. H., Lockwood, M., Bromage, B. J. I., Willis, D. M., and Lühr, H.: The dependence of high-latitude dayside ionospheric flows on the North-South component of the IMF: A high time resolution correlation analysis using EISCAT Polar and AMPTE UKS and IRM data, Planet. Space Sci., 36, 471–498, doi:10.1016/0032-0633(88)90107-9, 1988.

Fagundes, P. R., Aruliah, A. L., Rees, D., and Bittencourt, J. A.: Gravity wave generation and propagation during geomagnetic storms over Kiruna (67.8°N, 20.4°E), Ann. Geophys., 13, 358–366, 1995.

Forbes, J. M.: Dynamics of the Thermosphere, J. Meteorol. Soc. Japan, 85B, 193–213, 2007.

Forbes, J. M. and Garrett, H. B.: Theoretical Studies of Atmospheric Tides, Rev. Geophys. Space Phys., 17, 1951–1981, 1979.

Forbes, J. M., Hagan, M. E., Zhang, X., and Hamilton, K.: Upper atmospheric tidal oscillations due to latent heat release in the tropical troposphere, Ann. Geophys., 15, 1165–1175, 1997.

Ford, E. A. K., Aruliah, A. L., Griffin, E. M., and McWhirter, I.: High time resolution measurements of the thermosphere from Fabry-Perot Interferometer measurements of atomic oxygen, Ann. Geophys., 25, 1269–1278, 2007.

Ford, E. A. K., Aruliah, A. L., Griffin, E. M., and McWhirter, I.: Satistical analysis of thermospheric gravity waves from Fabry-Perot Interferometer measurements of atomic oxygen, Ann. Geophys., 26, 29–45, 2008.

Förster, M., Haaland, S. E., Paschmann, G., Quinn, J. M., Torbert, R. B., Vaith, H., and Kletzing, C. A.: High-latitude plasma convection during northward IMF as derived from in-situ magnetospheric Cluster EDI measurements, Ann. Geophys., 26, 2685–2700, 2008a.

Förster, M., Rentz, S., Köhler, W., Liu, H., and Haaland, S. E.: IMF dependence of high-latitude thermospheric wind pattern derived from CHAMP cross-track measurements, Ann. Geophys., 26, 1581–1595, 2008b.

Foster, J. C., St.-Maurice, J.-P., and Abreu, V. J.: Joule heating at high latitudes, J. Geophys. Res., 88, 4885–4896, 1983.

Frank, L. A., Paterson, W. R., and Sigwarth, J. B.: Observations of plasma injection into the ring current during substorm expansive phase, J. Geophys. Res., 107, 1343, doi:10.1029/2001JA000169, 2002.

Friedrich, E., Samson, J. C., Voronkov, I., and Rostoker, G.: Dynamics of the substorm expansive phase, J. Geophys. Res., 106, 13,145–13,163, 2001.

Fujiwara, H., Kataoka, R., Suzuki, M., Maeda, S., Nozawa, S., Hosokawa, K., Fukunishi, H., Sato, N., and Lester, M.: Electromagnetic energy deposition rate in the polar upper thermosphere derived from the EISCAT Svalbard radar and CUTLASS Finland radar observations, Ann. Geophys., 25, 2393–2403, 2007.

Fukunishi, H.: Dynamic relationship between proton and electron auroral substorms, J. Geophys. Res., 80, 553–574, 1975.

Fuller-Rowell, T. J.: A two-dimensional, high-resolution, nested-grid model of the thermosphere 2. Response of the thermosphere to narrow and broad electrodynamic features, J. Geophys. Res., 90, 6567–6586, 1985.

Fuller-Rowell, T. J. and Rees, D.: A Three-dimensional Time-dependent Global Model of the Thermosphere, J. Atmos. Sci., 37, 2545–2567, 1980.

Fuller-Rowell, T. J., Rees, D., Quegan, S., Moffet, R. J., Codrescu, M. V., and Millward, G. H.: A coupled thermosphere-ionosphere model (CTIM), in: STEP Handbook on Ionospheric Models, edited by Schunk, R. W., p. 217, Utah State University, 1996.

Gavrilov, N. M. and Roble, R. G.: The effect of gravity waves on the global mean temperature and composition structure of the upper atmosphere, J. Geophys. Res., 99, 25,773–25,780, 1994.

Gonzalez, W. D., Joselyn, J. A., Kamide, Y., Kroehl, H. W., Rostoker, G., Tsurutani, B. T., and Vasyliunas, V. M.: What is a geomagnetic storm?, J. Geophys. Res., 99, 5771–5792, 1994.

Gordiets, B. F., Kulikov, Y. N., Markov, M. N., and Marov, M. Y.: Numerical modelling of the thermospheric heat budget, J. Geophys. Res., 87, 4504–4514, 1982.

Gorney, D. J.: Solar cycle effects on the near-Earth space environment, Rev. Geophys., 28, 315–336, 1990.

Gosling, J. T. and Hansen, R. T.: Solar Wind Speed Distributions: 1962-1970, J. Geophys. Res., 76, 1811–1815, 1971.

Greet, P. A., Conde, M. G., Dyson, P. L., Innis, J. L., Breed, A. M., and Murphy, D. J.: Thermospheric wind field over Mawson and Davis, Antarctica; simultaneous observations by two Fabry-Perot spectrometers of λ630 nm emission, J. Atmos. Solar-Terr. Phys., 61, 1025–1045, 1999.

Greet, P. A., Innis, J. L., and Dyson, P. L.: Thermospheric Vertical Winds in the Auroral Oval/Polar Cap Region, Ann. Geophys., 20, 1987–2001, 2002.

Griffin, E. M., Aruliah, A. L., McWhirter, I., Yiu, H.-C. I., Charalambous, A., and McCrea, I.: Upper thermospheric neutral wind and temperature measurements from an extended spatial field, Ann. Geophys., 26, 2649–2655, 2008.

Grocott, A. and Yeoman, T. K.: SuperDARN observations of ionospheric convection during magnetospheric substorms, Int. Conf. Substorms-8, pp. 81–86, 2006.

Groves, G. V. and Forbes, J. M.: Equinox tidal heating of the upper atmosphere, Planet. Space Sci., 32, 447, 1984.

Gueymard, C. A.: The sun's total and spectral irradiance for solar energy applications and solar radiation models, 76, 423–453, 2004.

Gundlach, J. P., Larsen, M. F., and Mikkelsen, I. S.: A simple model describing the nonlinear dynamics of the dusk/dawn asymmetry in the high-latitude thermospheric flow, Geophys. Res. Lett., 15, 307–310, 1988.

Guo, W. and McEwen, D. J.: Vertical winds in the central polar cap, Geophys. Res. Lett., 30, 1725, doi:10.1029/2003GL017124, 2003.

Haaland, S. E., Paschmann, G., Förster, M., Quinn, J. M., Torbert, R. B., McIlwain, C. E., Vaith, H., Puhl-Quinn, P. A., and Kletzing, C. A.: High-latitude plasma convection from Cluster EDI measurements: method and IMF-dependence, Ann. Geophys., 25, 239–253, 2007.

Hargreaves, J. K.: The Solar Terrestrial Environment, Cambridge University Press, 1992.

Hays, P. B., Rusch, D. W., Roble, R. G., and Walker, J. C. G.: The O I (6300Å) Airglow, Rev. Geophys. Space Phys., 16, 225–232, 1978.

Hays, P. B., Killeen, T. L., Spencer, N. W., Wharton, L. E., Roble, R. G., Emery, B. A., Fuller-Rowell, T. J., Rees, D., Frank, L. A., and Craven, J. D.: Observations of the Dynamics of the Polar Thermosphere, J. Geophys. Res., 89, 5597–5612, 1984.

Hecht, E.: Optics, Addison-Wesley, second edn., 1974.

Hedin, A. E.: Neutral thermospheric composition and thermal structure, Rev. Geophys. Space Phys., 17, 477–485, 1979.

Hedin, A. E.: MSIS-86 Thermospheric Model, J. Geophys. Res., 92, 4649, 1987.

Hedin, A. E.: Extension of the MSIS Thermosphere Model into the Middle and Lower Atmosphere, J. Geophys. Res., 96, 1159–1172, 1991.

Hedin, A. E., Spencer, N. W., and Killeen, T. L.: Empirical Global Model of Upper Thermosphere Winds Based on Atmosphere and Dynamics Explorer Satellite Data, J. Geophys. Res., 93, 9959–9978, 1988.

Hedin, A. E., Biondi, M. A., Burnside, R. G., Hernandez, G., Johnson, M., Killeen, T. L., Mazaudier, C., Meriwether, J. W., Salah, J. E., Sica, R. J., Smith, R. W., Spencer, N. W., Wickwar, V. B., and Virdi, T. S.: Revised Global Model of Thermosphere Winds Using Satellite and Ground-Based Observations, J. Geophys. Res., 96, 7657–7688, 1991.

Heelis, R. A., McEwen, D., and Guo, W.: Ion and neutral motions observed in the winter polar upper atmosphere, J. Geophys. Res., 107, 1476, doi:10.1029/2002JA009359, 2002.

Hernandez, G.: Fabry-Perot Interferometers, Cambridge University Press, 1986.

Hernandez, G. and Roble, R. G.: Thermospheric nighttime neutral temperature and winds over Fritz Peak Observatory: Observed and calculated solar cycle variation, J. Geophys. Res., 100, 14,647–14,659, 1995.

Hernandez, G., Smith, R. W., Roble, R. G., Gress, J., and Clark, K. C.: Thermospheric Dynamics at the South Pole, Geophys. Res. Lett., 17, 1255–1258, 1990.

Hicks, T. R., Reay, N. K., and Scaddan, R. J.: A servo-controlled Fabry-Perot interferometer using capacitance micrometers for error detection, 7, 27–30, 1974.

Hill, T. W.: Origin of the plasma sheet, Rev. Geophys. Space Phys., 12, 379–388, 1974.

Hocke, K. and Schlegel, K.: A review of atmospheric gravity waves and travelling ionospheric disturbances: 1982-1995, Ann. Geophys., 14, 917–940, 1996.

Holmes, J. M., Conde, M., Deehr, C., and Lummerzheim, D.: Morphology of evening sector aurorae in λ557.7nm Doppler temperatures, Geophys. Res. Lett., 32, 2005.

Holzer, R. E. and Slavin, J. A.: Magnetic flux transfer associated with expansions and contractions of the dayside magnetosphere, J. Geophys. Res., 83, 3831–3839, 1978.

Holzworth, R. H. and Meng, C.-I.: Mathematical representation of the auroral oval, Geophys. Res. Lett., 2, 377–380, 1975.

Hsu, T.-S. and McPherron, R. L.: Occurence frequencies of IMF triggered and nontriggered substorms, J. Geophys. Res., 108, 1307, doi:10.1029/2002JA009442, 2003.

Hundhausen, A. J.: The Solar Wind, In: Introduction to Space Physics, chap. 4, pp. 91–128, Cambridge University Press, 1995.

Hunsucker, R. D.: Atmospheric gravity waves generated in the high-latitude ionosphere: A review, Rev. Geophys. Space Phys., 20, 293–513, 1982.

Iijima, T. and Potemra, T. A.: The amplitude distribution of field-aligned currents at northern high latitudes observed by triad, J. Geophys. Res., 81, 2165–2174, 1976.

Iijima, T. and Potemra, T. A.: Large-scale characteristics of field-aligned currents associated with substorms, J. Geophys. Res., 83, 599–615, 1978.

Innis, J. L. and Conde, M.: High-latitude thermospheric vertical wind activity from Dynamics Explorer 2 wind and temperature observations: Indications of a source region for polar cap gravity waves, J. Geophys. Res., 107, 1172, doi:10.1029/2001JA009130, 2002.

Innis, J. L., Greet, P. A., and Dyson, P. L.: Fabry-Perot spectrometer observations of the auroral oval/polar cap boundary above Mawson, Antarctica, J. Atmos. Terr. Phys., 58, 1973–1988, 1996.

Innis, J. L., Greet, P. A., Murphy, D. J., Conde, M. G., and Dyson, P. L.: A large vertical wind in the thermosphere at the auroral oval/polar cap boundary seen simultaneously from Mawson and Davis, Antarctica, J. Atmos. Solar-Terr. Phys., 61, 1047–1058, 1999.

Innis, J. L., Greet, P. A., and Dyson, P. L.: Evidence for thermospheric gravity waves in the southern polar cap from ground-based vertical velocity and photometric observations, Ann. Geophys., 19, 533–543, 2001.

Ishii, M.: Relationship between thermospheric vertical wind and the location of ionospheric current in the polar region, Adv. Polar Upper Atmos. Res., 19, 63–70, 2005.

Ishii, M., Oyama, S., Nozawa, S., Fujii, R., Sagawa, E., Watari, S., and Shinagawa, H.: Dynamics of Neutral Wind in the Polar Region Observed with two Fabry-Perot Interferometers, Earth Planets Space, 51, 833–844, 1999.

Ishii, M., Conde, M., Smith, R. W., Krynicki, M., Sagawa, E., and Watari, S.: Vertical wind observations with two Fabry-Perot interferometers at Poker Flat, Alaska, J. Geophys. Res., 106, 10,537–10,551, 2001.

Ishii, M., Kubota, M., Conde, M., Smith, R. W., and Krynicki, M.: Vertical wind distribution in the polar thermosphere during Horizontal E Region Experiment (HEX) campaign, J. Geophys. Res., 109, A12 311, doi:10.1029/2004JA010657, 2004.

Issautier, K.: Some Basic Aspects of the Solar Wind, in: Solar and Heliospheric Origins of Space Weather Phenomena, edited by Rozelot, J.-P., Lect. Notes Phys., chap. 3, pp. 25–49, Springer, 2006.

Jones, N., Wardill, P., and Jacka, F.: Dynamics of the Thermosphere over Mawson, Antarctica: II. Dependence on the Y component of the interplanetary magnetic field, in: Anare Res. Notes, edited by Burns, G. B. and Craven, M., 48, Antarctic Division, Department of Science, 1987.

Kallenrode, M.-B.: Space Physics, Springer-Verlag, third edn., 2004.

Kamide, Y., Perreault, P. D., Akasofu, S.-I., and Winningham, J. D.: Dependence of substorm occurence probability on the interplanetary magnetic field and on the size of the auroral oval, J. Geophys. Res., 82, 5521–5528, 1977.

Kamide, Y., Sun, W., and Akasofu, S.-I.: The average ionospheric electrodynamics for the different substorm phases, J. Geophys. Res., 101, 99–109, 1996.

Kamide, Y., Baumjohann, W., Daglis, I. A., Gonzalez, W. D., Grande, M., Joselyn, J. A., McPherron, R. L., Phillips, J. L., Reeves, E. G. D., Rostoker, G., Sharma, A. S., Singer, H. J., Tsurutani, B. T., and Vasyliunas, V. M.: Current understanding of magnetic storms: Storm-substorm relationships, J. Geophys. Res., 103, 17,705–17,728, 1998.

Kamide, Y., Kokubun, S., Bargatze, L. F., and Frank, L. A.: The size of the polar cap as an indicator of substorm energy, Phys. Chem. Earth, 24, 119–127, 1999.

Kaufmann, R. L.: Substorm currents: Growth phase and onset, J. Geophys. Res., 92, 7471–7486, 1987.

Kella, D., Vejby-Christensen, L., Johnson, P. J., Pedersen, H. B., and Andersen, L. H.: The source of green light emission determined from a heavy-ion storage ring experiment, Science, 276, 1530–1533, doi:10.1126/science.276.5318.1530, 1997.

Keskinen, M. J. and Satyanarayana, P.: Nonlinear unstable auroral-arc driven thermospheric winds in an ionosphere-magnetosphere coupled model, Geophys. Res. Lett., 20, 2687–2690, 1993.

Khan, H. and Cowley, S. W. H.: Observations of the response time of high-latitude ionospheric convection to variations in the interplanetary magnetic field using EISCAT and IMP-8 data, Ann. Geophys., 17, 1306–1335, 1999.

Killeen, T. L.: Energetics and Dynamics of the Earth's Thermosphere, Rev. Geophys., 25, 433–454, 1987.

Killeen, T. L. and Roble, R. G.: Thermosphere dynamics: Contributions from the first 5 years of the Dynamics Explorer program, Rev. Geophys., 26, 329–367, 1988.

Killeen, T. L., Craven, J. D., Frank, L. A., Ponthieu, J.-J., Spencer, N. W., Heelis, R. A., Brace, L. H., Roble, R. G., Hays, P. B., and Carignan, G. R.: On the relationship between dynamics of the polar thermosphere and morphology of the aurora: global-scale observations from Dynamics Explorers 1 and 2, J. Geophys. Res., 93, 2675–2692, 1988.

Killeen, T. L., Won, Y.-I., Niciejewski, R. J., and Burns, A. G.: Upper thermosphere winds and temperatures in the geomagnetic polar cap: Solar cycle, geomagnetic activity, and interplanetary magnetic field dependencies, J. Geophys. Res., 100, 21,327–21,342, 1995.

Kivelson, M. G. and Russell, C. T., eds.: Instroduction to Space Physics, Cambridge University Press, 1995.

Klostermeyer, J.: Thermospheric heating by atmospheric gravity waves, J. Atmos. Terr. Phys., 35, 2267–2275, 1973.

Knipp, D. J., Tobiska, W. K., and Emery, B. A.: Direct and indirect thermospheric heating sources for solar cycles 21-23, Solar Phys., 224, 495–505, 2004.

Kosch, M. J. and Nielsen, E.: Coherent radar estimates of average high-latitude ionospheric Joule heating, J. Geophys. Res., 100, 12,201–12,215, 1995.

Kosch, M. J., Ishii, M., Kohsiek, A., Rees, D., Schlegel, K., Hagfors, T., and Cierpka, K.: A comparison of vertical thermospheric winds from Fabry-Perot interferometer measurements over a 50 km baseline, Adv. Space Res., 26, 985–988, 2000.

Kosch, M. J., Cierpka, K., Rietveld, M. T., Hagfors, T., and Schlegel, K.: High-latitude ground-based observations of the thermospheric ion-drag time constant, Geophys. Res. Lett., 28, 1395–1398, 2001.

Kosch, M. J., Anderson, C., Makarevich, R. A., Carter, B. A., Fiori, R. A. D., Conde, M., Dyson, P. L., and Davies, T.: First E region observations of mesoscale neutral wind interaction with auroral arcs, J. Geophys. Res., 115, A02 303, doi:10.1029/2009JA014697, 2010.

Kozyra, J. U., Nagy, A. F., and Slater, D. W.: High-altitude enery source(s) for stable auroral red arcs, Rev. Geophys., 35, 155–190, 1997.

Larsen, M. F.: Winds and shears in the mesosphere and lower thermosphere: Results from four decades of chemical release wind measurements, J. Geophys. Res., 107, 1215, doi:10.1029/2001JA000218, 2002.

Larsen, M. F. and Odom, C. D.: Observations of altitudinal and latitudinal E-region neutral wind gradients near sunset at the magnetic equator, Geophys. Res. Lett., 24, 1711–1714, 1997.

Larsen, M. F. and Walterscheid, R. L.: Modified geostrophy in the thermosphere, J. Geophys. Res., 100, 17,321–17,329, 1995.

Larsen, M. F., Marshall, T. R., Mikkelsen, I. S., Emery, B. A., Christensen, A., Kayser, D., Hecht, J., Lyons, L., and Walterscheid, R.: Atmospheric Response in Aurora experiment: Observations of E and F region neutral winds in a region of postmidnight diffuse aurora, J. Geophys. Res., 100, 17,299–17,308, 1995.

Lee, D.-H.: On the generation mechanism of Pi 2 pulsations in the magnetosphere, Geophys. Res. Lett., 25, 583–586, 1998.

Leinert, C., Bowyer, S., Haikala, L. K., Hanner, M. S., Hauser, M. G., Levassuer-Regourd, A. C., Mann, I., Mattila, K., Reach, W. T., Schlosser, W., Staude, H. J., Toller, G. N., Weiland, J. L., Weinberg, J. L., and Witt, A. N.: The 1997 reference of diffuse night sky brightness, Astron. & Astrophys. Suppl. Ser., 127, 1–99, 1998.

Lilensten, J. and Kretzschmar, M.: The solar spectrum in the UV, EUV, and X ranges: observations, modelling, and effects on the Earth upper atmosphere in the frame of space weather, in: Solar and Heliospheric Origins of Space Weather Phenomena, edited by Rozelot, J.-P., Lect. Notes Phys., chap. 4, pp. 53–88, Springer, 2006.

Link, R.: A rocket observation of the 6300-Å/5200-Å intensity ratio in the dayside aurora: Implications for the production of $O(^1D)$ via the reaction $N(^2D) + O_2 \rightarrow NO + O(^1D)$, Geophys. Res. Lett., 10, 225, 1983.

Link, R.: $N(^2D) + O_2$: A source of thermospheric 6300 Å emission?, Planet. Space Sci., 40, 699–705, 1992.

Liou, K., Newell, P. T., Sibeck, D. G., Meng, C.-I., Brittnacher, M., and Parks, G.: Observation of IMF and seasonal effects in the location of auroral substorm onset, J. Geophys. Res., 106, 5799–5810, 2001.

Lopez, R. E., Koskinen, H. E. J., Pulkkinen, T. I., Bösinger, T., McEntire, W., and Potemra, T. A.: Simultaneous observation of the poleward expansion of substorm electrojet activity and the tailward expansion of current sheet disruption in the near-earth magnetotail, J. Geophys. Res., 98, 9285–9295, 1993.

Lühr, H., Rother, M., Köhler, W., Ritter, P., and Grunwaldt, L.: Thermospheric upwelling in the cusp region: Evidence from CHAMP observations, Geophys. Res. Lett., 31, L06 805, doi:10.1029/2003GL019314, 2004.

Lühr, H., Rentz, S., Ritter, P., Liu, H., and Häusler, K.: Average thermospheric wind patterns over the polar regions, as observed by CHAMP, Ann. Geophys., 25, 1093–1101, 2007.

Lui, A. T. Y.: Current Disruption in the Earth's Magnetosphere: Observations and Models, J. Geophys. Res., 101, 13,067–13,088, 1996.

Lui, A. T. Y.: Current controversies in magnetospheric physics, Rev. Geophys., 39, 535–563, 2001.

Lui, A. T. Y., Volwerk, M., Dunlop, M. W., Alexeev, I. V., Fazakerley, A. N., Walsh, A. P., Lester, M., Grocott, A., Mouikis, C., Henderson, M. G., Kistler, L. M., Shen, C., Shi, J. K., Zhang, T. L., and Rème, H.: Near-Earth substorm features from multiple satellite observations, J. Geophys. Res., 113, A07S26, doi:10.1029/2007JA012738, 2008.

Lyons, L. R.: Formation of auroral arcs via magnetosphere-ionosphere coupling, Rev. Geophys., 30, 93–112, 1992.

Lyons, L. R., Killeen, T. L., and Walterscheid, R. L.: The neutral wind 'flywheel' as a source of quiet-time, polar-cap currents, Geophys. Res. Lett., 12, 101–104, 1985.

Lyons, L. R., Blanchard, G. T., Samson, J. C., Lepping, R. P., Yamamoto, T., and Moretto, T.: Coordinated observations demonstrating external substorm triggering, J. Geophys. Res., 102, 27,039–27,051, 1997.

MacDougall, J. W., Andre, D. A., Sofko, G. J., Huang, C.-S., and Koustov, A. V.: Travelling ionospheric disturbance properties deduced from Super Dual Auroral Radar measurements, Ann. Geophys., 18, 1550–1559, 2001.

Maeda, S., Fuller-Rowell, T. J., and Evans, D. S.: Zonally averaged dynamical and compositional response of the thermosphere to auroral activity during September 18-24, 1984, J. Geophys. Res., 94, 16,869–16,883, 1989.

Maeda, S., Fuller-Rowell, T. J., and Evans, D. S.: Heat budget of the thermosphere and temperature variations during the recovery phase of a geomagnetic storm, J. Geophys. Res., 97, 14,947–14,957, 1992.

Mantas, G. P. and Carlson, H. C.: Reexamination of the O(^{3}P$\rightarrow^1$D) excitation rate by thermal electron impact, Geophys. Res. Lett., 18, 159–162, 1991.

Mayr, H. G., Harris, I., and Spencer, N. W.: Some Properties of Upper Atmosphere Dynamics, Rev. Geophys. Space Phys., 16, 539–565, 1978.

Mayr, H. G., Harris, I., Spencer, N. W., Hedin, A. E., Wharton, L. E., Porter, H. S., Walker, J. C. G., and Carlson Jr., H. C.: Tides and the midnight temperature anomaly in the thermosphere, Geophys. Res. Lett., 6, 447–450, 1979.

Mayr, H. G., Harris, I., Varosi, F., and Herrero, F. A.: Global Excitation of Wave Phenomena in a Dissipative Multiconstituent Medium 1. Transfer Function of the Earth's Thermosphere, J. Geophys. Res., 89, 10,929–10,959, 1984a.

Mayr, H. G., Varosi, F., and Herrero, F. A.: Global Excitation of Wave Phenomena in a Dissipative Multiconstituent Medium 2. Impulsive Perturbations in the Earth's Thermosphere, J. Geophys. Res., 89, 10,961–10,986, 1984b.

Mayr, H. G., Harris, I., Herrero, F. A., and Varosi, F.: Winds and composition changes in the thermosphere using the transfer function model, J. Atmos. Solar-Terr. Phys., 59, 691–709, 1997.

McCormac, F. G. and Smith, R. W.: The influence of the interplanetary magnetic field y component on ion and neutral motions in the polar thermosphere, Geophys. Res. Lett., 1, 935–938, 1984.

McCormac, F. G., Killeen, T. L., Gombosi, E., Hays, P. B., and Spencer, N. W.: Configuration of the high-latitude thermosphere neutral circulation for IMF $\mathrm{B_y}$ negative and positive, Geophys. Res. Lett., 12, 155–158, 1985.

McPherron, R. L.: Magnetospheric Substorms, Rev. Geophys. Space Phys., 17, 657–681, 1979.

McPherron, R. L., Russell, C. T., and Anlory, M. P.: Satellite studies of magnetospheric substorms on August 15, 1968: A phenomenological model for substorms, J. Geophys. Res., 78, 3131, 1973.

Meloni, A., De Michelis, P., and Tozzi, R.: Geomagnetic storms, dependence on solar and interplanetary phenomena: a review, 76, 882–887, 2005.

Meng, C.-I.: Dynamic variation of the auroral oval during intense magnetic storms, J. Geophys. Res., 89, 227–235, 1984.

Millward, G. H., Moffet, R. J., Quegan, S., and Fuller-Rowell, T. J.: A coupled thermosphere-ionosphere-plasmasphere model CTIP, in: STEP Handbook on Ionospheric Models, edited by Schunk, R. W., p. 239, Utah State University, 1996.

Morley, S. K. and Freeman, M. P.: On the association between northward turnings of the interplanetary magnetic field and substorm onsets, Geophys. Res. Lett., 34, L08 104, doi:10.1029/2006GL028891, 2007.

Müller-Wodarg, I. and Aylward, A. D.: The influence of tides on composition of the thermosphere, Adv. Space Res., 21, 807–810, 1998.

Nakai, H., Kamide, Y., Hardy, D. A., and Gussenhoven, M. S.: Time scales of expansion and contaction of the auroral oval, J. Geophys. Res., 91, 4437–4450, 1986.

Nappo, C. J.: An introduction to atmospheric gravity waves, vol. 85 of *International Geophysics*, Academic Press, 2002.

Niciejewski, R. J., Killeen, T. L. R., and Won, Y.: Observations of neutral winds in the polar cap during northward IMF, J. Atmos. Terr. Phys., 56, 285–295, 1994.

Niciejewski, R. J., Killeen, T. L., and Solomon, S. C.: Observations of thermospheric horizontal neutral winds at Watson Lake, Yukon Territory ($\Lambda = 65°$N), J. Geophys. Res., 101, 241–259, 1996.

Ogawa, T., Balan, N., Otsuka, Y., Shiokawa, K., Ihara, C., Shimomai, T., and Saito, A.: Observations and modeling of 630 nm airglow and total electron content associated with traveling ionospheric disturbances over Shigaraki, Japan, 54, 45–56, 2002.

Ohtani, S., Yamaguchi, R., Nosé, M., Kawano, H., Engebretson, M., and Yumoto, K.: Quiet time magnetotail dynamics and their implications for the substorm trigger, J. Geophys. Res., 107, 1030, doi:10.1029/2001JA000116, 2002.

Olsson, A., Janhunen, P., Karlsson, T., Ivchenko, N., and Blomberg, L. G.: Statistics of Joule heating in the auroral zone and polar cap using Astrid-2 satellite Poynting flux, Ann. Geophys., 22, 4133–4142, 2004.

Østgaard, N., Germany, G., Stadsnes, J., and Vondrak, R. R.: Energy analysis of substorms based on remote sensing techniques, solar wind measurements and geomagnetic indices, J. Geophys. Res., 107, 1233, doi:10.1029/2001JA002002, 2002a.

Østgaard, N., Vondrak, R. R., Gjerloev, J. W., and Germany, G.: A relation between the energy deposition by electron precipitation and geomagnetic indices during substorms, J. Geophys. Res., 107, doi:10.1029/2001JA002003, 2002b.

Pallamraju, D., Chakrabarti, S., and Valladares, C. E.: Magnetic storm-induced enhancement in neutral composition at low latitudes as inferred by $O(^1D)$ dayglow measurements from Chile, Ann. Geophys., 22, 3241–3250, 2004.

Palmroth, M., Janhunen, P., Germany, G., Lummerzheim, D., Liou, K., Baker, D. N., Barth, C., Weatherwax, A. T., and Watermann, J.: Precipitation and total power consumption in the ionosphere: Global MHD simulation results compared with Polar and SNOE observations, Ann. Geophys., 24, 861–872, 2006.

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

Press, W. H., Flannery, B. P., Teukolsky, S. A., and Vetterling, W. T.: Numerical recipes: the art of scientific computing, Cambridge University Press, 1986.

Price, G. D., Smith, R. W., and Hernandez, G.: Simultaneous measurements of large vertical winds in the upper and lower thermosphere, J. Atmos. Terr. Phys., 57, 631–643, 1995.

Priest, E. R.: The Sun and its Magnetohydrodynamics, In: Introduction to Space Physics, chap. 3, pp. 58–90, Cambridge University Press, 1995.

Reddy, C. A. and Mayr, H. G.: Storm-time penetration to low latitudes of magnetospheric-ionospheric convection and convection-driven thermospheric winds, Geophys. Res. Lett., 25, 3075–3078, 1998.

Rees, D., McWhirter, I., Hays, P. B., and Dines, T.: A stable, rugged, capacitance-stabilised piezoelectric scanned Fabry-Perot etalon, 14, 1320–1325, 1981.

Rees, D., Fuller-Rowell, T. J., Gordon, R., Killeen, T. L., Hays, P. B., Wharton, L. E., and Spencer, N. W.: A Comparison of Wind Observations of the Upper Thermosphere From the *Dynamics Explorer* Satellite with the Predictions of a Global Time-Dependent Model, Planet. Space Sci., 31, 1299–1314, 1983a.

Rees, D., Greenaway, A. H., Gordon, R., McWhirter, I., Charleton, P. J., and Åke Steen: The Doppler imaging system: initial observations of the auroral thermosphere, Planet. Space Sci., 32, 273–285, 1984a.

Rees, D., Smith, R. W., Charleton, P. J., McCormac, F. G. abd Lloyd, N., and Åke Steen: The generation of vertical winds and gravity waves at auroral latitudes - I. Observations of vertical winds, Planet. Space Sci., 32, 667–684, 1984b.

Rees, D., Lloyd, N. D., Fuller-Rowell, T. J., and Steen, A.: Observations of the variations of thermospheric winds in northern Scandinavia between 1980 and 1986: A study of geomagnetic activity effects during the last solar cycle, Adv. Space Res., 7, 39–47, 1987.

Rees, D., Smith, R. W., Signernes, F., Henriksen, K., Brandstrom, U., Harris, M., and Maskall, G.: Observations of thermospheric neutral winds within the polar cusp and the auroral oval using a Doppler imaging system (DIS), Ann. Geophys., 16, 1461–1474, 1998.

Rees, M. H.: Physics and chemistry of the upper atmosphere, Cambridge atmospheric and space science series, Cambridge University Press, 1989.

Rees, M. H., Emery, B. A., Roble, R. G., and Stamnes, K.: Neutral and Ion Gas Heating by Auroral Electron Precipitation, J. Geophys. Res., 88, 6289–6300, 1983b.

Reeves, G. D. and Henderson, M. G.: The storm-substorm relationship: Ion injections in geosynchronous measurements and composite energetic neutral atom images, J. Geophys. Res., 106, 5833–5844, 2001.

Richmond, A. D., Ridley, E. C., and Roble, R. G.: A thermosphere/ionosphere general circulation model with coupled electrodynamics, Geophys. Res. Lett., 19, 601–604, 1992.

Ridley, A. J., Deng, Y., and Tóth, G.: The global ionosphere-thermosphere model, J. Atmos. Solar-Terr. Phys., 68, 839–864, 2006.

Rishbeth, H. and Garriott, O. K.: Introduction to Ionospheric Physics, Academic Press, 1969.

Rishbeth, H. and Müller-Wodarg, I. C. F.: Vertical circulation and thermosheric composition: a modelling study, Ann. Geophys., 17, 794–805, 1999.

Rishbeth, H., Müller-Wodarg, I. C. F., Zou, L., Fuller-Rowell, T. J., Millward, G. H., Moffet, R. J., Idenden, D. W., and Aylward, A. D.: Annual and semiannual variations in the ionospheric F2-layer: II. Physical discussion, Ann. Geophys., 18, 945–956, 2000.

Rishbeth, H., Heelis, R. A., and Müller-Wodarg, I. C. F.: Variations of thermospheric composition according to AE-C data and CTIP modelling, Ann. Geophys., 22, 441–452, 2004.

Roble, R. G.: Dynamics of the Earth's Thermosphere, Rev. Geophys. Space Phys., 21, 217–233, 1983.

Roble, R. G.: Energetics of the Mesosphere and Thermosphere, in: The Upper Mesosphere and Lower Thermosphere: A Review of Experiment and Theory, edited by Johnson, R. M. and Killeen, T. L., no. 87 in Geophysical Monograph, chap. 1, pp. 1–21, American Geophysical Union, 1995.

Roble, R. G. and Emery, B. A.: On the global mean temperature of the thermosphere, Planet. Space Sci., 31, 597, 1983.

Roble, R. G., Ridley, E. C., and Dickinson, R. E.: On the global mean structure of the thermosphere, J. Geophys. Res., 92, 8745–8758, 1987.

Roble, R. G., Ridley, A. D., Richmond, A. D., and Dickinson, R. E.: A coupled thermosphere/ionosphere general circulation model, Geophys. Res. Lett., 15, 1325–1328, 1988.

Rodger, A. S., Wells, G. D., Moffet, R. J., and Bailey, G. J.: The variability of Joule heating, and its effects on the ionosphere and thermosphere, Ann. Geophys., 19, 773–781, 2001.

Rostoker, G.: Polar magnetic substorms, Rev. Geophys. Space Phys., 10, 157–211, 1972.

Rostoker, G., Akasofu, S.-I., Foster, J., Greenwald, R. A., Kamide, Y., Kawasaki, K., Lui, A. T. Y., McPherron, R. L., and Russell, C. T.: Magnetospheric substorms-Definition and signatures, J. Geophys. Res., 85, 1663–1668, 1980.

Rottman, G. J.: Observations of solar UV and EUV variability, Adv. Space Res., 8, 53–66, 1988.

Ruohoniemi, J. M., Shepherd, S. G., and Greenwald, R. A.: The response of the high-latitude ionosphere to IMF variations, J. Atmos. Solar-Terr. Phys., 64, 159–171, 2002.

Rusch, D. W., Sharp, W. E., and Gérard, J.-C.: The reaction of $N(^2D)$ with O_2 as a source of $O(^1D)$ atoms in aurorae, Geophys. Res. Lett., 5, 1043–1046, 1978.

Salby, M. L.: Fundamentals of Atmospheric Physics, vol. 61 of *International Geophysical Series*, Academic Press, 1995.

Sazykin, S., Fejer, B. G., Galperin, Y. I., Zinin, L. V., Grigoriev, S. A., and Mendillo, M.: Polarization jet events and excitation of weak SAR arcs, Geophys. Res. Lett., 29, doi:10.1029/2001GL014388, 2002.

Schlegel, K., Lühr, H., St.-Maurice, J.-P., Crowley, G., and Hackert, C.: Thermospheric density structures over the polar regions observed with CHAMP, Ann. Geophys., 23, 1659–1672, 2005.

Schoendorf, J., Crowley, G., and Roble, R. G.: Neutral density cells in the high latitude thermosphere - 2. Mechanisms, J. Atmos. Terr. Phys., 58, 1769–1781, 1996a.

Schoendorf, J., Crowley, G., Roble, R. G., and Marcos, F. A.: Neutral density cells in the high latitude thermosphere - 1. Solar maximum cell morphology and data analysis, J. Atmos. Terr. Phys., 58, 1751–1768, 1996b.

Sekar, R., Gurubaran, S., and Sridharan, R.: All sky imaging Fabry-Perot spectrometer for optical investigation of the upper atmosphere, Indian Journal of Radio & Space Physics, 22, 197–204, 1993.

Sergeev, V. A., Apatenkov, S. V., Angelopoulos, V., McFadden, J. P., Larson, D., Bonnell, J. W., Kuznetsova, M., Partamies, N., and Honary, F.: Simultaneous THEMIS observations in the near-tail portion of the inner and outer plasma sheet flux tubes at substorm onset, J. Geophys. Res., 113, A00C02, doi:10.1029/2008JA013527, 2008.

Sheehan, C. H. and St.-Maurice, J.-P.: Dissociative recombination of N_2^+, O_2^+, and NO^+: Rate coefficients for ground state vibrationally excited ions, J. Geophys. Res., 109, doi: 10.1029/2003JA010132, 2004.

Shematovich, V., Gérard, J.-C., Bisikalo, D. V., and Hubert, B.: Thermalization of $O(^1D)$ atoms in the thermosphere, J. Geophys. Res., 104, 4287–4295, 1999.

Shue, J.-H. and Weimer, D. R.: The Relationship between Ionospheric Convection and Magnetic Activity, J. Geophys. Res., 99, 401–415, 1994.

Sica, R. J., Hernandez, G., Romick, G. J., Rees, M. H., and Roble, R. G.: Auroral Zone Thermospheric Dynamics 2. Individual Nights, J. Geophys. Res., 91, 13,593–13,611, 1986a.

Sica, R. J., Rees, M. H., Romick, G. J., Hernandez, G., and Roble, R. G.: Auroral Zone Thermospheric Dynamics 1. Averages, J. Geophys. Res., 91, 3231–3244, 1986b.

Sipler, D. P., Biondi, M. A., and Roble, R. G.: F-region neutral winds and temperatures at equatorial latitudes: measured and predicted behavior during geomagnetically quiet conditions, Planet. Space Sci., 31, 53–66, 1983.

Sipler, D. P., Biondi, M. A., and Zipf, M. E.: Vertical winds in the midlatitude thermosphere from Fabry-Perot Interferometer measurements, J. Atmos. Terr. Phys., 57, 621–629, 1995.

Slavin, J. A., Fairfield, D. H., Kuznetsova, M. M., Owen, C. J., Lepping, R. P., Taguchi, S., Mukai, T., Saito, Y., Yamamoto, T., Kokubun, S., Lui, A. T. Y., and Reeves, G. D.: ISTP observations of plasmoid ejection: IMP 8 and Geotail, J. Geophys. Res., 103, 119–133, 1998.

Slavin, J. A., Fairfield, D. H., Lepping, R. P., Hesse, M., Ieda, A., Tanskanen, E., Østgaard, N., Mukai, T., Nagai, T., Singer, H. J., and Sutcliffe, P. R.: Simultaneous observations of earthward flow bursts and plasmoid ejection during magnetospheric substorms, J. Geophys. Res., 107, 1106, doi:10.1029/2000JA003501, 2002.

Smith, R. W.: Vertical winds: a tutorial, J. Atmos. Solar-Terr. Phys., 60, 1425–1434, 1998.

Smith, R. W.: The global-scale effect of small-scale thermospheric disturbances, J. Atmos. Solar-Terr. Phys., 62, 1623–1628, 2000.

Smith, R. W. and Hernandez, G.: Vertical winds in the thermosphere within the polar cap, J. Atmos. Terr. Phys., 57, 611–620, 1995.

Smith, R. W., Rees, D., and Stewart, R. D.: Southern hemisphere thermospheric dynamics: a review, Rev. Geophys., 26, 591–622, 1988.

Smith, R. W., Hernandez, G., Roble, R. G., Dyson, P. L., Conde, M., Crickmore, R., and Jarvis, M.: Observation and simulations of winds and temperatures in the Antarctic thermosphere for August 2-10, 1992, J. Geophys. Res., 103, 9473–9480, 1998.

Sobral, J. H., Takahashi, A. H., Abdu, M. A., Muralikrishna, P., Sahai, Y., and Zmlutti, C. J.: O(^1S) and O(^1D) quantum yields from rocket measurements of electron densities and 557.7 nm and 630 nm emissions in the nocturnal airglow, Planet. Space Sci., 40, 607–619, 1992.

Solomon, S. C., Hays, P. B., and Abreu, V. J.: The auroral 6300 Å emission: Observations and modelling, J. Geophys. Res., 93, 9867–9882, 1988.

Spencer, Wharton, L. E., Carignan, G. R., and Maurer, J. C.: Thermosphere zonal winds, vertical motions and temperature as measured from Dynamics Explorer, Geophys. Res. Lett., 9, 953–956, 1982.

Spencer, N. W., Wharton, L. E., Niemann, H. B., Hedin, A. E., Carignan, G. R., and Maurer, J. C.: The Dynamics Explorer wind and temperature spectrometer, Space Science Instrumentation, 5, 417–428, 1981.

Stern, D. P.: A brief history of magnetospheric physics during the space age, Rev. Geophys., 34, 1–32, 1996.

Streit, G. E., Howard, E. C. J., Schmeltekopf, J. J., Davidson, J. A., and Schiff, H. I.: Temperature dependence of O(^{1}D) rate constants for reactions with O_2, N_2, CO_2, O_3 and H_2O, 65, 4761–4764, 1976.

Sun, Y. and Dalgarno, A.: Collisional excitation of metastable O(^{1}D) atoms, 96, 5017–5019, 1992.

Tachiev, G. I. and Froese Fischer, C.: Breit-Pauli energy levels and transition rates for nitrogen-like and oxygen-like sequences, Astron. & Astrophys., 385, 716–723, doi:10.1051/0004-6361:20011816, 2002.

Tanskanen, E., Pulkkinen, T. I., Koskinen, H. E. J., and Slavin, J. A.: Substorm energy budget during low and high solar activity: 1997 and 1999 compared, J. Geophys. Res., 107, 1086, doi:10.1029/2001JA900153, 2002.

Testud, J.: Gravity waves generated during magnetic substorms, J. Atmos. Terr. Phys., 32, 1793–1805, 1970.

Thayer, J. P.: Height-resolved Joule heating rates in the high-latitude *E* region and the influence of neutral winds, J. Geophys. Res., 103, 471–487, 1998.

Thayer, J. P. and Killeen, T. L.: Vorticity and Divergence in the High-Latitiude Upper Thermosphere, Geophys. Res. Lett., 18, 701–704, 1991.

Thayer, J. P. and Killeen, T. L.: A kinematic analysis of the high-latitude thermospheric neutral circulation pattern, J. Geophys. Res., 98, 11,549–11,565, 1993.

Thayer, J. P. and Semeter, J.: The convergence of magnetospheric energy flux in the polar atmosphere, J. Atmos. Solar-Terr. Phys., 66, 807–824, 2004.

Todd, H., Cowley, S. W. H., Lockwood, M., Willis, D. M., and Lühr, H.: Response time of the high-latitude dayside ionosphere to sudden changes in the north-south component of the IMF, Planet. Space Sci., 36, 1415–1428, doi:10.1016/0032-0633(88)90008-6, 1988.

Torr, M. R. and Torr, D. G.: The role of metastable species in the thermosphere, Rev. Geophys. Space Phys., 20, 91–144, 1982.

Torr, M. R., Richards, P. G., and Torr, D. G.: A new determination of the ultraviolet heating efficiency of the thermosphere, J. Geophys. Res., 85, 6819–6826, 1980.

Tu, C.-Y. and Marsch, E.: Two-fluid model for heating of the solar corona and acceleration of the solar wind by high-frequency Alfvén waves, Solar Phys., 171, 363–391, 1997.

Vadas, S. L. and Fritts, D. C.: Thermospheric responses to gravity waves: Influences of increasing viscosity and thermal diffusivity, J. Geophys. Res., 110, D15 103, doi:10.1029/2004JD005574, 2005.

Vasyliunas, V. M. and Song, P.: Meaning of ionospheric Joule heating, J. Geophys. Res., 110, A02 301, doi:10.1029/2004JA010615, 2005.

Vasyliunas, V. M. and Wolf, R. A.: Magnetospheric substorms: Some problems and controversies, Rev. Geophys. Space Phys., 11, 181–189, 1973.

Verma, M. K.: Anomalous Viscosity, Resistivity, and Thermal Diffusivity of the Solar Wind Plasma, in: eprint arXiv:chao-dyn/9509002, 1995.

Vlasov, M. N., Nicolls, M. J., Kelley, M. C., Smith, S. M., Aponte, N., and González, S. A.: Modeling of airglow and ionospheric parameters at Arecibo during quiet and disturbed periods in October 2002, J. Geophys. Res., 110, A07 303, doi:10.1029/2005JA011074, 2005.

Volland, H.: Atmospheric Tidal and Planetary Waves, Atmospheric sciences library, Kluwer Academic Publishers, 1988.

Walterscheid, R. L. and Lyons, L. R.: The neutral circulation in the vicinity of a stable auroral arc, J. Geophys. Res., 97, 19,489–19,499, 1992.

Wang, H., Lühr, H., Ridley, A., Ritter, P., and Yu, Y.: Storm time dynamics of auroral electrojets: CHAMP observation and the Space Weather Modelling Framework, Ann. Geophys., 26, 555–570, 2008a.

Wang, W., Killeen, T. L., Burns, A. G., and Roble, R. G.: A high-resolution, three-dimensional, time dependent, nested grid model of the coupled thermosphere-ionosphere, J. Atmos. Solar-Terr. Phys., 61, 385–397, doi:10.1016/S1364-6826(98)00079-0, 1999.

Wang, W., Burns, A. G., Wiltberger, M., Solomon, S. C., and Killeen, T. L.: Altitude variations of the horizontal thermospheric winds during geomagnetic storms, J. Geophys. Res., 113, A02 301, doi:10.1029/2007JA012374, 2008b.

Wardill, P. and Jacka, F.: Vertical motions in the thermosphere over Mawson, Antarctica, J. Atmos. Terr. Phys., 48, 289–292, 1986.

Wardill, P. and Jacka, F.: Dynamics of the Thermosphere over Mawson, Antarctica: III. Horizontal Divergence of the Wind Field, in: Anare Res. Notes, edited by Burns, G. B. and Craven, M., 48, pp. 131–137, Antarctic Division, Department of Science, 1987.

Wardill, P., Jones, N., and Jacka, F.: Dynamics of the Thermosphere over Mawson, Antarctica: I. Diurnal Variation and Geomagnetic Dependence, in: Anare Res. Notes, edited by Burns, G. B. and Craven, M., 48, pp. 114–130, Antarctic Division, Department of Science, 1987.

Weimer, D. R.: Models of high-latitude electric potentials with a least error fit of spherical harmonic coefficients, J. Geophys. Res., 100, 19 595–19 607, 1995.

Weimer, D. R.: A flexible IMF dependent model of high-latitude electric potentials having "space weather" applications, Geophys. Res. Lett., 23, 2549–2552, 1996.

Weimer, D. R.: Improved ionospheric electrodynamic models and application to calculating Joule heating rates, J. Geophys. Res., 110, A05 306, doi:10.1029/2004JA010884, 2005.

Wells, G. D., Rodger, A. S., Moffet, R. J., Bailey, G. J., and Fuller-Rowell, T. J.: The effects of nitric oxide cooling and the photodissociation of molecular oxygen on the thermosphere/ionosphere system over the Argentine Islands, Ann. Geophys., 15, 355–365, 1997.

Whitten, R. C. and Poppoff, I. G.: Fundamentals of Aeronomy, Space Science Text, John Wiley & Sons, Inc., 1971.

Wilcox, J. M., Schatten, K. H., and Ness, N. F.: Influence of Interplanetary Magnetic Field and Plasma on Geomagnetic Activity during Quiet-Sun Conditions, J. Geophys. Res., 72, 19–26, 1967.

Wilksch, P. A.: Instrument function of the Fabry-Perot spectrometer, Appl. Opt., 24, 1502–1511, 1985.

Williams, P. J. S., Virdi, T. S., Lewis, R. V., Lester, M., Rodger, A. S., McCrea, I. W., and Freeman, K. S. C.: Worldwide atmospheric gravty-wave study in the European sector 1985-1990, J. Atmos. Terr. Phys., 55, 683–696, 1993.

Wilson, G. R., Weimer, D. R., Wise, J. O., and Marcos, F. A.: Response of the thermosphere to Joule and particle precipitation, J. Geophys. Res., 111, A10 314, doi: 10.1029/2005JA011274, 2006.

Wimmer-Schweingruber, R. F.: Interplanetary Disturbances, in: Space Weather, edited by Scherer, K., Fichtner, H., Heber, B., and Mall, U., Lect. Notes Phys., chap. 4, pp. 71–122, Springer, 2005.

Witasse, O., Lilensten, J., Lathuillére, C., and Blelly, P. L.: Modeling the OI 630.0 and 557.7 nm thermospheric dayglow during EISCAT-WINDII coordinated measurements, J. Geophys. Res., 104conde, 24,639–24,655, 1999.

Yee, J.-H.: Non-thermal distribution of O(^1D) atoms in the night-time thermosphere, Planet. Space Sci., 36, 89–97, 1988.

Yeh, K. C. and Liu, C. H.: Acoustic-gravity waves in the upper atmosphere, Rev. Geophys. Space Phys., 12, 193–216, 1974.

Yokoyama, N., Kamide, Y., and Miyaoka, H.: The size of the auroral belt during magnetic storms, Ann. Geophys., 16, 566–573, 1998.

Zhang, S. P. and Shepherd, G. G.: Solar influence on the O(^1D) dayglow emission rate: Global-scale measurements by WINDII and UARS, Geophys. Res. Lett., 31, L07 804, doi:10.1029/2004GL019447, 2004.