The Synoptic Variability of
Thermospheric and Mesospheric Winds
Observed Using a Fabry-Perot Interferometer

by
Anasuya Lohini Aruliah
Department of Physics and Astronomy,
University College London

Thesis submitted for the degree of Doctor of Philosophy
of the University of London
September 1991
This thesis is dedicated to
Panchanayagam and Naomi Arasakumari,
my beloved parents
Abstract

Thermospheric neutral winds have been routinely observed in Northern Scandinavia since 1981 using ground-based Fabry-Perot Interferometers (FPI) to measure the Doppler shifts of the atomic oxygen emission at 6300Å. Data have been collected each winter period between August and April.

The database from Kiruna (67.8°N, 20.4°E), consisting of nine successive winters of observations, has been used to study the long-term behaviour of thermospheric neutral winds at high-latitudes in response to changes in the solar cycle, season and geomagnetic activity. Three significant characteristics were noticed. First, both the seasonal and geomagnetic response of the neutral winds showed a solar cycle variation which appears to be solely a high-latitude phenomenon. Second, the seasonal analysis showed a significant asymmetry between the spring and autumn equinox periods. This is attributed mainly to an equinoctial asymmetry in the coupling between the solar wind and the magnetosphere and consequently to the high-latitude ionospheric convection. Third, the difference in the neutral wind response for solar maximum and minimum indicates that the $K_p$ index is not a sufficient and unique indicator of the magnetospheric excitation of the thermosphere.

Two smaller Scandinavian databases exist, from Kilpisjarvi (69.1°N, 20.8°E) and Svalbard (78.2°N, 15.6°E). Case studies have been compiled for the winter of 1987-1988 using neutral wind vectors obtained simultaneously from all three auroral sites. The results have been compared with the UCL-Sheffield three-dimensional, time-dependent global model of the thermosphere/ionosphere. Simultaneous data from three sites impose more stringent bounds for comparison with the model.

Incorporation of a state-of-the-art Gallium Arsenide imaging detector into the FPI has made possible observation of the weak, infra-red emission line of the hydroxyl transition at 8430Å and thereby achieved continuous monitoring of the upper mesosphere. A strong semidiurnal variation and signatures of gravity waves were observed.
Contents

Abstract
List of Figures and Tables

Chapter 1. The Physics of the Upper Atmosphere
1.1 Introduction
1.2 Thesis Outline
1.3 The Earth’s Upper Atmosphere
1.4 The Physics of the Upper Atmosphere
1.5 The Thermosphere
1.6 Solar Heating
1.7 Solar Cycle Variations in Solar Flux
1.8 Ionospheric Convection at High-Latitudes
1.9 High-Latitude Heating
1.10 Large Vertical Winds
1.11 Auroral and Airglow Emissions Observed With the UCL FPI
   1.11.1 Atomic Oxygen Emission at 6300Å
   1.11.2 Hydroxyl Radical Emission at 8430Å

Chapter 2. Instrument Details and the Analysis of Data
2.1 Introduction
2.2 The UCL Fabry-Perot Interferometer
2.3 Data Analysis
   2.3.1 Etalon Theory
   2.3.2 Collecting Data
   2.3.3 Correcting Data
   2.3.4 Fitting Peaks
   2.3.5 Calculating Wind Speeds
2.4 Problems Encountered During Data Analysis
   2.4.1 Instrument Temperature Drift
   2.4.2 Changes in the Refractive Index of the Etalon Gap Due to Pressure Leaks
   2.4.3 Unstable Free Spectral Ranges
   2.4.4 Problems Particular to the Observation of the 8430Å Emission
   2.4.4.1 Determination of the 8430Å Free Spectral Range
   2.4.4.2 Determination of the 8430Å Baseline

4
2.4.4.3 Contamination of the 8430Å Emission
2.4.5 The Effect of Cloud on the Doppler Shifts
2.4.6 The Effect of an Error in the Zenith Angle
2.4.7 The Effect of an Error in the Orientation of the FPI
2.4.7.1 What Can be Deduced About the Instrument Inclination from the Data Collected so Far?
2.4.8 Converging and Diverging Wind Fields
2.4.9 Large Vertical Winds

Chapter 3. Seasonal and Solar Cycle Variations in Thermospheric Winds
3.1 Introduction
3.2 Description of the Database
3.3 Seasonal Variation in Thermospheric Winds
3.4 Seasonal and Solar Cycle Variation in Thermospheric Winds
3.5 Height of the 6300Å Emission Peak
3.6 Seasonal and Solar Cycle Variation in Ion Velocities
3.7 Considering the Effect of a Long-Term Phase Lag in the Response of the Thermosphere to Solar Heating
3.8 Considering the Influence of Solar Wind-Magnetospheric Coupling
   3.8.1 The Semi-Annual Variation in Geomagnetic Activity
   3.8.2 Seasonal Variation in the Orientation of the Earth to the IMF
   3.8.3 Possible Explanation of the Equinoctial Asymmetry in High-Latitude Thermospheric Winds
   3.8.4 Further Asymmetry Introduced by $B_{my}$
   3.8.5 Predictions from this Model
3.9 Summary

Chapter 4. Geomagnetic and Solar Cycle Variations in High-Latitude Thermospheric Winds
4.1 Introduction
4.2 Description of the Database
4.3 Geomagnetic and Solar Cycle Variations
4.4 Discussion
Chapter 5. Simultaneous Observations of Thermospheric Winds from Three Stations in Northern Scandinavia

5.1 Introduction
5.2 Thermospheric Winds at Auroral Latitudes
5.3 Comparing Model Data With FPI Data

Chapter 6. Mesospheric Winds Observed at a Mid-Latitude Site

6.1 Introduction
6.2 Tidal and Gravity Waves in the Middle Atmosphere
   6.2.1 Tides
   6.2.2 Acoustic-Gravity Waves
   6.2.3 Influence of Tidal and Gravity Waves on the Upper Atmosphere
6.3 Initial Results from the Utah FPI
   6.3.1 Standard Behaviour of Mid-Latitude Mesospheric Winds
   6.3.2 Possible Evidence of Gravity Waves

Chapter 7. Conclusion

7.1 Review
7.2 Future Work

References
List of Tables

Chapter 1.
1.1 Location and Operation of UCL FPIs.

Chapter 2.
2.1 Rate of drift of peak positions due to etalon pressure leak.
2.2 Calculated free spectral ranges for the 8430Å data.

Chapter 3.
3.1 Average parameters associated with seasonally averaged thermospheric winds.
3.2 Seasonally averaged thermospheric winds.
3.3 Conversion of $K_p$ into $A_p$ indices.
3.4 Average parameters associated with seasonally averaged thermospheric winds for solar maximum $F_{10.7} > 110$ and solar minimum $F_{10.7} < 110$.
3.5 Seasonally averaged thermospheric winds for solar maximum.
3.6 Seasonally averaged thermospheric winds for solar minimum.
3.7 Estimation of the height of the 6300Å emission peak from the time lag in the observation of the Harang discontinuity to the east and west of Kiruna.
3.8 Average ion velocities for solar maximum 1988-1990.
3.9 Average ion velocities for solar minimum 1984-1987.

Chapter 4.
4.1 Average parameters associated with $K_p$ averaged thermospheric winds for solar maximum $F_{10.7} > 110$ and solar minimum $F_{10.7} < 110$.
4.2 $K_p$ Averaged thermospheric winds for solar maximum.
4.3 $K_p$ Averaged thermospheric winds for solar minimum.

Chapter 6.
6.1 Important semidiurnal Hough Functions $(m, n)$ where $m$=longitudinal mode and $n$=latitudinal mode.
6.2 Prediction for the Observed Wind Speed $V_{obs}$ for Various Values of the Half Width Half Maximum $\Delta t$ where $V_o = 50$ m/s.
List of Figures

Chapter 1
1.1 Nomenclature of the upper atmosphere based on temperature, composition, gaseous escape and ionization. [Hargreaves, 1979]
1.2 Typical vertical profiles of the mid-latitude ionosphere. [Hargreaves, 1979]
1.3 Distribution of major gases in the thermosphere for an exospheric temperature of 800 K. [COSPAR International Reference Atmosphere, 1972]
1.4 The altitude at which the rate of absorption of solar radiation is at a maximum. The principle atmospheric constituents that absorb the radiation in the different wavelength bands are indicated. [Handbook of Geophysics and the Space Environment, 1985]
1.5 Global temperature distribution in the thermosphere at equinox in conditions of medium solar activity. No account is taken of high-latitude heat sources due to energetic particles. [Jacchia, 1965]
1.6 Winds at 300 km computed from the temperature distribution of Figure 1.5. [Kohl and King, 1967]
1.7 The ratio of the solar spectral irradiance near solar maximum during January 1979 to near solar minimum during July 1976 for solar cycle 21. The ratios are plotted for 25Å intervals of wavelength. [Hinteregger, 1981]
1.8 Variation of F10.7 solar flux between 1980 and 1991.
1.9 The structure of the sun, internal and atmospheric. [Akasofu and Chapman, 1972]
1.10 The 27-day variation in several UV emission lines observed from the OSO 3 satellite in May and June 1967. The 10.7 cm flux for the same period is also given for comparison. [Hall and Hinteregger, 1970]
1.11 Global mean heating rate profiles. [Roble and Emery, 1983]
1.12 Variation in the size of the auroral oval with activity as denoted by the magnetic index Q. Coordinate system is CG latitude and CG local time. [Feldstein and Starkov, 1967]
1.13 The substorm in the luminous aurora. (a) T = 0, (b) T = 0 – 5 minutes, (c) T = 5 – 10 minutes, (d) T = 10 – 30 minutes, (e) T = 30 min–1 hour, (f) T = 1 – 2 hours. [Akasofu, 1968]
1.14 Comparison of model 6300Å emission profiles with and without auroral electron precipitation. [Sica et al., 1986]
1.15 Computed altitude profiles of the OI 6300Å volume emission rate produced by an incident Maxwellian electron spectrum with characteristic energy of 0.4 keV and total energy input rate of 2 ergs/cm²/s. The numerals identify the various processes that contribute to the total volume emission rate: 1. dissociative recombination of O₂⁺ ions; 2. electron impact by energetic auroral electrons; 3. electron impact by the thermal population; 4. dissociation of O₂ molecules; 5. cascading from the O(1S) level; and 6. the reaction between N(²D) atoms and O₂ molecules. The reaction between N⁺ ions and O₂ molecules is shown by a dashed line, 7., to indicate a tentative contribution. [Rees and Roble, 1986]

Chapter 2
2.1 Schematic diagram of the Fabry-Perot Interferometer [Lloyd, 1985]
2.2 Fabry-Perot fringes [Born and Wolf, 1987]
2.3 Airy function [Born and Wolf, 1987]
2.4 Three full 2-D images obtained with the FPI at Kilpisjarvi during the winter of 1990-1991. (a) calibration lamp image, (b) Flat-Field image, (c) Thermionic Emission image, and (d) Flat-Field image from an IPD at the end of its lifespan, measured at Longyearbyen in the winter of 1988-1989
2.5 An example of the reduced to radius-squared image of the calibration lamp obtained from the FPI at Kilpisjarvi.
2.6 Examples of the variation of peak position with time in the eight line-of-sight directions for the nights 8-11th February 1991.
2.7 An extreme example of the effect of instrument temperature drift on peak positions as observed at Kilpisjarvi on the 24th January 1987.
2.8 Illustration of the effect on the peak positions of a drop in temperature during the evening, typical for a mid-continent site such as Utah, which is too large and too rapid to be controlled by the FPI temperature control.
2.9 (a)Variation of the position of the first and second peaks with time due to a pressure leak in the etalon at Kiruna during the winter of 1987-1988, (b)Variation of the Free Spectral Range with time due to a pressure leak in the etalon over the same period of time.
2.10 An extreme example of detector distortion
2.11 Emission produced by the neon lamp used to calibrate the 8430Å data.
2.12 Contamination of 8430Å emission peaks.
2.13 Illustration of the effect of the presence of cloud on the perceived line-of-sight winds.
2.14 Geometry illustrating an FPI with its axis tilted at an angle \( \alpha \) from the vertical.

2.15 Figure demonstrating the error produced by a significant vertical wind component, \( \mathbf{V}_z \), during the calculation of the horizontal zonal wind, \( \mathbf{V}_e \) which assumes a zero vertical wind. \( \mathbf{V} \) is the actual horizontal zonal wind and \( \theta \) is the zenith angle for the line-of-sight observation.

Chapter 3

3.1 Seasonally averaged thermospheric winds 1981-1990: autumn
3.2 Seasonally averaged thermospheric winds 1981-1990: winter
3.3 Seasonally averaged thermospheric winds 1981-1990: spring
3.4 Seasonal comparison of thermospheric winds at solar maximum
3.5 Seasonal comparison of thermospheric winds at solar minimum
3.6 Representative electron density profiles at high-, middle- and low-latitudes for three levels of solar activity (\( R = 0, 100, 200 \)) based partly on bottomside ionograms and partly on theory. [Wright, 1962]
3.7 Seasonal variation in plasma drifts at Tromsø at solar maximum. [Farmer and Jarvis, 1991]
3.8 Seasonal variation in plasma drifts at Tromsø at solar minimum. [Farmer and Jarvis, 1991]
3.9 Monthly variation of solar flux and position of solar terminator at latitude 68°N.
3.10 Simple model of the daily variation of photoionization for an (a) optically thin and (b) optically thick medium.
3.11 Monthly variation of \([\text{O}] / [\text{N}_2]\) and daylight hours at pressure level 11 for solar maximum and solar minimum at latitude 68°N.
3.12 Monthly variation of \([\text{O}] \) and \([\text{N}_2]\) at pressure level 11 for solar maximum and solar minimum at latitude 68°N.
3.13 Monthly variation of photoionization/(\([\text{O}] / [\text{N}_2]\)) at pressure level 11 for solar maximum minimum at latitude 68°N for (a) an optically thick atmosphere and (b) an optically thin atmosphere.
3.14 Typical butterfly diagram showing the yearly variation in distribution of sunspots. During a sunspot cycle the distribution migrates toward the solar equator. [Handbook of Geophysics, 1985]
3.15 Annual variation of daily \( Ap \) index for 1937-1956. (a) weekly mean maximum, (b) weekly mean, (c) weekly mean minimum. [McIntosh, 1959]
3.16 Annual variation of the daily $U$ index for active (top curve), intermediate (middle curve) and quiet (bottom curve) years averaged over the years 1872-1930. [Chapman and Bartels, 1940]

3.17 (a) Northward IMF resulting in a closed magnetosphere, (b) Southward IMF resulting in an open magnetosphere. [Russell, 1972]

3.18 Satellite observations of the Interplanetary Magnetic Field. The orientation is given in terms of the Geocentric Solar Ecliptic coordinate system.

3.19 View of the ecliptic plane illustrating the average orientation of the Interplanetary Magnetic Field vector and the position of the Earth's North Geographic Pole at the solstices and the equinoxes.

3.20 Schematic illustration of the warped solar current sheet which divides interplanetary space into regions of magnetic field pointing towards or away from the sun. The region illustrated by the diagram covers approximately 20 astronomical units.

3.21 A contour plot of the diurnal and annual variation of the effective average southward component, $B_{m,z}$, of the IMF in the GSM coordinate system due to the towards and away sectors along an ideal spiral field. The more negative the contour the more geomagnetically active the interval should be on average [Russell and McPherron, 1973].

3.22 Average orientation of the IMF with respect to the Earth's magnetosphere in spring and autumn.

3.23 Seasonal variation of the effective average southward component, $B_{m,z}$, of the IMF in the GSM coordinate system calculated from the one-hour averages of satellite data collected between 1963 and 1990. [Wild, 1991]

3.24 Seasonal variation of the effective average southward component, $B_{m,z}$, of the IMF for six 4-hour intervals, average over the years 1963-1990. [Wild, 1991]

3.25 Normalised reconnection voltage between closed geomagnetic field and open IMF vs. IMF clock angle, $\theta$. Also shown for reference are the curves $\sin(\theta/2)$ and $\sin^2(\theta/2)$. [Fedder et al., 1991]

3.26 Equipotential convection contours derived from Millstone Hill radar observations for four levels of high-latitude particle precipitation. A 2kV contour spacing is shown in polar apex magnetic latitude/local time coordinates. [Foster et al., 1986]

3.27 The dependence of the location of the northern and southern edges of the auroral belt on geomagnetic activity as represented by the $Kp$ index, in the period around midnight (22-02 hours) in Local Geomagnetic Time. $\psi'$ is the
corrected geomagnetic latitude. The dashed line represents the centre line of the auroral belt. [Feldstein and Starkov, 1967]

3.28 Monthly and diurnal variation in electron densities at solar minimum measured by the EISCAT radar at Tromsø. [Farmer, 1990]

3.29 Four models of the high-latitude electric field for:
   (a) $B_y < 0, B_z > 0$,
   (b) $B_y > 0, B_z > 0$,
   (c) $B_y < 0, B_z < 0$,
   (d) $B_y > 0, B_z < 0$.

Chapter 4

4.1 $Kp$ averaged neutral winds observed at Kiruna at solar maximum for:
   (a) $0^\circ \leq Kp < 2^\circ$
   (b) $2^\circ \leq Kp < 5^\circ$
   (c) $5^\circ \leq Kp < 8^\circ$

4.2 $Kp$ averaged neutral winds observed at Kiruna at solar minimum for:
   (a) $0^\circ \leq Kp < 2^\circ$
   (b) $2^\circ \leq Kp < 5^\circ$
   (c) $5^\circ \leq Kp < 8^\circ$

4.3 Thermospheric wind vectors observed at College, Alaska for (a) low, (b) moderate and (c) high geomagnetic activity. [Sica et al., 1986]

4.4 Calculated contours of mass flow stream function for December solstice during (a) solar minimum and (b) solar maximum. [Roble et al., 1977]

4.5 (a) $Kp$ averaged neutral vectors observed at Kiruna at solar maximum and (b) $Kp$ averaged neutral vectors observed at Kiruna at solar minimum.

4.6 Comparison of average zonal winds for $1 < Kp < 2$ at solar maximum with $2 < Kp < 3$ at solar minimum.

4.7 Comparison of average zonal winds for $2 < Kp < 3$ at solar maximum with $4 < Kp < 5$ at solar minimum.

4.8 The global network of observatories involved in measuring the $Kp$ index in 1988. [Menvielle and Berthellier, 1991]

4.9 Electron collision frequency, $\nu_e$, and gyrofrequency, $\omega_e$, for the height region $50 - 150$ km. The ion collision frequency is $10 - 100$ times smaller than $\nu_e$, and the ions gyrofrequency is $30 - 50 \times 10^3$ times smaller than $\omega_e$, depending on the nature of the ion. [Thrane and Piggott, 1966]
4.10 Height dependence of the mean ionization rate per unit volume (m⁻³) per unit incident energy flux (1 mW/m²) for precipitating electrons with spectra with peaks in five different energy bands, where band 2 has a characteristic energy of 0.379 keV and band 22 has a characteristic energy of 17.479 keV. [Fuller-Rowell and Evans, 1987]

Chapter 5

5.1 Neutral wind vectors observed on the night of the 11th January 1988.

5.2 Line-of-sight neutral winds observed on the night of the 11th January 1988 at (a) Longyearbyen, (b) Kilpisjarvi and (c) Kiruna.

5.3 Neutral wind vectors observed on the night of the 15th December 1987.

5.4 Line-of-sight neutral winds observed on the night of the 15th December 1987 at (a) Longyearbyen, (b) Kilpisjarvi and (c) Kiruna.

5.5 Neutral wind vectors observed on the night of the 21st December 1987.

5.6 Line-of-sight neutral winds observed on the night of the 21st December 1987 at (a) Longyearbyen, (b) Kilpisjarvi and (c) Kiruna.

5.7 Neutral wind vectors observed on the night of the 14th January 1988.

5.8 Line-of-sight neutral winds observed on the night of the 14th January 1988 at (a) Longyearbyen, (b) Kilpisjarvi and (c) Kiruna.

5.9 Plasma velocities observed with the EISCAT radar for the period 15-21UT on the 14th January 1988.

5.10 Plasma velocities observed with the EISCAT radar for the period 03-09UT on the 15th January 1988.

5.11 UCL-Sheffield model simulation LJ3 of thermospheric winds observed through 24 hours at a longitude of 18°E, over a latitude range of 60°N to 82°N for geomagnetically quiet conditions at solar minimum.

5.12 UCL-Sheffield model simulation KA3 of thermospheric winds observed through 24 hours at a longitude of 18°E, over a latitude range of 60°N to 82°N for geomagnetically moderate conditions at solar minimum.

Chapter 6

6.1 Hough functions for the two antisymmetric and symmetric modes of lowest order for the semidiurnal solar tide. [Hargreaves, 1979]

6.2 Net upward flux of diurnal tide energy. The net downward flux of solar radiation in bands absorbable in the thermosphere is shown for comparison. [Lindzen, 1967]
6.3 Comparison of geomagnetic records from the auroral zone with electron content data from mid-latitudes, showing the onset of a magnetic substorm and the subsequent large-scale TIDs [Davies, 1971]

6.4 Zonal winds at 50°N with and without semidiurnal lower boundary forcing in the University College London TGCM. The simulations correspond to equinox solar minimum ($F_{10.7} = 75$) conditions. Pressure level 0 is the model lower boundary, assumed at an altitude of 80 km and pressure of 1 Pascal. The vertical pressure intervals are one pressure scale height. For the conditions of this model, pressure level 12 is at an altitude of approximately 280 km. [Forbes, 1987]

6.5 Comparison of the reduced to radius-squared fringe profile obtained for the emission at 6300Å with (a) an S-25 detector and (b) a state-of-the-art Gallium Arsenide (GaAs) detector to show the vast improvement in the signal to noise ratio.

6.6 Meridional wind components for Utah 8430Å.

6.7 Zonal wind components for Utah 8430Å.

6.8 Profiles of the W-E component of the wind as measured from 36 rockets which carried chemical trail experiments all near 30°N latitude between the 15th November and the 14th December or near 30°S latitude between the 15th May and the 14th June. [COSPAR 1972]

6.9 Solar semidiurnal westerly winds for December solstice from the Forbes model. [Forbes, 1987]

6.10 Solar diurnal westerly winds for December solstice from the Forbes model. [Forbes, 1987]

6.11 Rocket measurements of the altitude distribution of the hydroxyl emissions with wavelengths: (a) 7280Å, (b) 7600-10,400Å and (c) 9100-10,700Å. [Krasovskij and Šefov, 1965]

6.12 A simplified hydroxyl emission altitude distribution used to determine the average integrated winds in the upper mesosphere as observed with an FPI at Utah.

6.13 Mesospheric winds observed on the 1st September 1989 to the north-east and south-west of Utah.

6.14 Mesospheric winds observed on the 9th September 1989 to the south and south-west of Utah.
1.1 Introduction

Research into the Earth’s upper atmosphere is a relatively new branch of physics. Meteorology and astronomy are ancient sciences, evident in even the most primitive of civilisations, but the swathe of sky separating the clouds from the stars remained a mystery until this century. Then at the turn of the 20th century the discovery by Marconi of the ionosphere, which reflected radio waves back to the Earth, created a new interest in this part of the atmosphere. The exploitation of this discovery has provided some of the most important remote sensing techniques available, such as ionospheric sounders, coherent and incoherent radar.

After the second world war captured German V2 rockets enabled American scientists to make the first direct measurements of the upper atmosphere. This gave a tremendous boost to our knowledge but the data was limited in location and time, determined by the path and duration of the rocket flight. Then the artificial satellite era begun in the late fifties allowed long-term direct measurements to be made as well as indirect measurements through satellite drag. Consequently the sixties and seventies produced a wealth of understanding such that now the general behaviour of the thermosphere in particular has been well resolved [e.g. Rishbeth, 1989, Killeen, 1987, Rees and Fuller-Rowell, 1987, Smith et al., 1988a].

The Atmospheric Physics Laboratory at University College London has specialised in an optical remote sensing method using ground-based Fabry-Perot Interferometers (FPI). There are seven instruments deployed at four different sites. The instruments used, their locations and the emission lines observed are given in Table 1.1. Six contain a single Fabry-Perot etalon and the seventh contains a triple etalon. There are two further instruments built at UCL and operated by other research teams: one at Halley Bay in the South Pole operated by the British Antarctic Survey [Stewart et al., 1985, Smith et al., 1988a] and the other at Saskatoon in Canada, operated by the University of Saskatchewan [Lloyd et al., 1989]. The FPIs are used to measure the Doppler shifts of various emission lines from the airglow and aurora at heights within the upper atmosphere. The
bulk motion of the air i.e. winds, in the thermosphere and mesosphere can then be deduced.

Observations alone are not sufficient for understanding the physics of the atmosphere. Credit must also be given to the parallel progress in computer technology that has allowed the development and exploitation of complex models of the upper atmosphere. As a result of this progress a few years ago it became possible for the UCL three-dimensional, time-dependent global model of the thermosphere [Fuller-Rowell and Rees, 1980, Fuller-Rowell, 1981] to be combined with the University of Sheffield model of the ionosphere [Quegan et al., 1982, 1986] to become a powerful tool, prominent in modelling studies of the upper atmosphere [Fuller-Rowell et al., 1987, Rees et al., 1988]. The insights from such models have been an invaluable aid to explaining qualitatively how various ionospheric features are formed that had been observed for two decades without being understood. These features include localised ionization troughs and extended tongues of ionization which were found not to be anomalies but the natural consequence of the interplay between magnetic convection and corotation [Knudsen, 1974, Spiro et al., 1978].

Table 1.1

<table>
<thead>
<tr>
<th>Location</th>
<th>instrument wavelength periods of regular and reliable observation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kiruna (67.4N, 20.4E)</td>
<td>FPI 6300Å every winter from Nov.'81</td>
</tr>
<tr>
<td>Kiruna (67.4N, 20.4E)</td>
<td>FPI 5577Å every winter from Jan.'87</td>
</tr>
<tr>
<td>Kiruna (67.4N, 20.4E)</td>
<td>triple FPI 6300Å under development</td>
</tr>
<tr>
<td>Kiruna (67.4N, 20.4E)</td>
<td>DIS 6300Å under development</td>
</tr>
<tr>
<td>Kilpisjarvi (69.1N, 20.8E)</td>
<td>FPI 6300Å every winter from 1987</td>
</tr>
<tr>
<td>Ny Alesund (78.9N, 11.9E)</td>
<td>FPI 6300Å Jan.'83, Dec.'83 and Jan.'84</td>
</tr>
<tr>
<td>Longyearbyen (78.2N, 15.6E)</td>
<td>FPI 6300Å every winter from 1986</td>
</tr>
<tr>
<td>Bear Lake (41.9N, 111.4W)</td>
<td>FPI 8430Å from Sept.'89</td>
</tr>
</tbody>
</table>
1.2 Thesis Outline

The instrument used to obtain all the results presented in this thesis is the single etalon FPI, the details of which are given in Chapter 2 along with the method of analysis of the data. All of the observations were limited to the nighttime hours when the signal is sufficiently distinct from the background emissions. However, the increased finesse of a triple etalon extends this observation period well into the dawn and twilight hours, so that eventually a 24 hour coverage is anticipated. Unfortunately the triple etalon FPI is an expensive instrument and difficult to operate owing to its complexity, so the vast majority of results obtained so far come from the array of single etalon FPIs.

Single etalon FPIs are located at three sites in northern Scandinavia: at Kiruna, Sweden (67.8°N, 20.4°E), Kilpisjarvi, Finland (69.1°N, 20.8°E) and the island of Svalbard (78°N, 14°E). These three sites have provided an extensive database of the neutral winds of the upper thermosphere derived from the Doppler shifts of the red line emission of atomic oxygen (OI) at 6300Å. The emission line at 6300Å has long been favoured for monitoring the behaviour of the upper thermosphere. It is an emission with a high intensity in the visible region of the electromagnetic spectrum. The lifetime of the transition is long, approximately 100 seconds, which allows the oxygen atom to reach thermal equilibrium with the surrounding neutral gas. The intensity profile peaks at a height of around 240 km. Thus altogether the 6300Å emission gives a good measure of the behaviour of the thermospheric neutral gas.

The single etalon FPI measuring the 6300Å emission was first set up at Kiruna in 1980, but it was not run routinely until November 1981. Since then it has been run routinely for every winter until the present time. Consequently this database of FPI observations is probably one of the largest of its kind. It has been used to investigate the long-term behaviour of high-latitude thermospheric neutral winds at night in response to changes in the solar cycle, season and geomagnetic activity. These results are given in Chapters 3 and 4.

The stores of data from the instruments at Kilpisjarvi and Svalbard are much smaller owing to the difficulties of maintaining instruments at remote sites where one visit near the beginning of each winter is the only time available for trouble-shooting and, in the case of Svalbard, collecting the data. The winter of 1987 to 1988 was the sole winter where all three instruments were running together reliably. Chapter 5 presents composite wind pictures produced by putting together the wind vectors obtained simultaneously from all three auro-
ral sites during this winter. The wind pictures are used as case studies of high-latitude neutral wind behaviour which can be compared with the UCL-Sheffield three-dimensional, time-dependent global model of the thermosphere/ionosphere [Fuller-Rowell et al., 1987, Rees et al., 1988, Rees and Fuller-Rowell, 1989]. An individual night's behaviour can be dramatically different from the average behaviour shown in Chapters 3 and 4.

The fourth site is at Bear Lake, Utah in North America (41.9°N, 111.4°W). This is a collaborative project with Utah State University to monitor the mid-latitude mesosphere. The results from the first year of observations are shown in Chapter 6. The mesosphere has also been called the 'ignorosphere' owing to our ignorance of its behaviour. It has always been difficult to observe directly, being too low for satellites and too high for rockets and balloons. By incorporating a state-of-the-art Gallium Arsenide (GaAs) detector into the FPI the emission line of the hydroxyl molecule at 8430Å has been clearly observed by the ground-based detector at Utah. This emission peaks at around 85 km thus allowing remote observation of the mesosphere. Previous detectors have had poor quantum efficiencies in the infra-red region and so the GaAs detector has provided a real breakthrough for emission lines such as this, and by-the-by opened up the mesosphere to our view.

A general overview of the behaviour of the thermosphere will be given in the remainder of this chapter. More specific details will be dealt with in each of the other chapters, as required by the experimental analyses. Description of the mesosphere will be left to Chapter 6, which will be generally self-contained and separate from Chapters 3, 4 and 5.
1.3 The Earth’s Upper Atmosphere

The Earth’s atmosphere extends from the surface of the Earth to heights of several Earth radii. However, for millenia Man’s knowledge of the atmospheric wind systems relied on passive observations of natural phenomena and so could not extend higher than around 100 km. The winds from the highest measurable altitude of 80 – 110 km were first estimated from observations by eye of the distortion of persistent meteor trails. Since then the advent of radars, rockets and artificial satellites has revolutionized atmospheric physics. Today radars are used to track meteor trails accurately and chemical tracers are propelled into the sky by rocket as far as 200 – 300 km. Above this, satellites, continuously circling the Earth, produce ribbon images of the atmosphere for all heights and times.

There are several schemes for labelling different height regions according to their individual properties, and all of them split the atmosphere into many layers like a giant onion. The identification of each of the layers is determined by criteria such as the temperature profile, chemical composition, the height above which molecules can escape the Earth’s gravity and the ionized gas content. All of these are illustrated in Figure 1.1. In terms of the temperature distribution the atmosphere is divided into four regions: the troposphere, stratosphere, mesosphere and thermosphere. The height region containing the ionized gas is called the ionosphere which is subdivided by a further set of labels according to the height distribution of the electron density, as shown in Figure 1.2. Ledges in the electron density profile serve to distinguish the D-, E- and F-regions. This latter set of labels together with the temperature profile labels are the most useful for describing the environment of the observations made here.

The definition of the upper atmosphere is not a fixed one. The stratosphere straddles the loose boundary between the upper and lower atmospheres. Meteorologists banish the stratosphere to the upper atmosphere while atmospheric physicists relegate it to the lower atmosphere. For this thesis the upper atmosphere refers to the mesosphere and thermosphere alone. However, by putting the mesosphere and thermosphere together it should not be assumed that the two layers are at all similar. As can be seen from Figure 1.1 the separation of the homosphere and heterosphere lies between the mesosphere and thermosphere. Thus the mesosphere is within the region where there is complete mixing of all the chemical constituents of the atmosphere, while the thermosphere lies in the region where each molecular component takes a height distribution governed by
its mass and temperature. In addition the ionosphere is almost totally within the thermosphere, so altogether the mesosphere and thermosphere are radically different. Consequently in certain studies the mesosphere is distinguished as part of the middle atmosphere, together with the stratosphere and lower thermosphere.

1.4 The Physics of the Upper Atmosphere

The classification most commonly used for the atmosphere is based on the temperature distribution shown in Figure 1.1. Each stratum of the atmosphere is bounded by a temperature maximum or minimum. The lowest stratum is the troposphere, encompassing the region from the surface of the Earth to a height of 10—12 km. For this region the temperature falls off at a rate of $10^\circ$/km until it reaches a minimum value at a height called the tropopause. The next region is the stratosphere which experiences a temperature rise due to energy absorbed by an ozone layer at around 50 km, which height being at the temperature peak is therefore named the stratopause. The mesosphere is a region of temperature drop. The mesopause is at a height of 80—85 km and has the coldest temperature of the whole atmosphere at around 180 K. Above the mesopause the temperature rises until around 300 km when it reaches a constant value of 1000—2000 K. This uppermost region is the thermosphere.

The distribution of major gases in the thermosphere is shown in Figure 1.3. Below the turbopause, which is around 100 km above the Earth’s surface, the major gases are $\text{N}_2$ and $\text{O}_2$ in a proportion of 4:1. The gases are mixed by winds and turbulence so that the concentration ratio does not change with height. Above the turbopause the gases are no longer turbulently mixed. The vertical distribution of a gas above the turbopause becomes determined by the balance between the vertical pressure gradient and gravity. This is given by the hydrostatic equation 1.1, where $\frac{dP}{dh}$ is the incremental change in pressure $P$ for an incremental change in height $dh$ for particles of mass $m$ and concentration $n$, and $g$ is the acceleration due to gravity.

$$\frac{dP}{dh} = -nmg \quad (1.1)$$

Equation 1.2 is the ideal gas law where $k$ is the Boltzmann constant and $T$ is the temperature:

$$P = nkT \quad (1.2)$$
The variation of pressure with height is found to be an exponential function, as given in equation 1.3. This function is obtained by substituting equation 1.2 into equation 1.1 and integrating from any base height $h_0$, at which the pressure is $P_0$, to height $h$ and pressure $P$.

$$P = P_0 \exp(-z)$$ \hspace{1cm} (1.3)

Generally the pressure is given in terms of the reduced height of the gas, $z$, which is calculated using equation 1.4.

$$z = \int_{h_0}^{h} \frac{dh}{H} = \frac{(h - h_0)}{H}$$ \hspace{1cm} (1.4)

Thus the height distribution of each gas above the turbopause is determined by its scale height, $H$, which is calculated using equation 1.5.

$$H = \frac{kT}{mg}$$ \hspace{1cm} (1.5)

The scale height depends on the gas molecular mass, and consequently the chemical composition of the atmosphere changes with height and the light gases become more abundant at higher altitudes. This process is called diffusive separation. In contrast, the turbulent mixing of gases below the turbopause results in all the major gases having the same scale height, derived from a molecular mass of 29 which is the concentration weighted average mass at the turbopause.
1.5 The Thermosphere

The thermosphere can be treated as a fluid subject to the hydrodynamic equations of motion. At 300 km the interval between collisions among particles is around one second and the mean free path of the particles is around 1 km. However, above the exobase, at 600 km, the air particles move in ballistic orbits with few collisions, in which case the hydrodynamic fluid equations are no longer valid.

The global movement of thermospheric air masses is the result of a complex mixture of driving forces, most of which are thought to be known, if not fully understood. Heat from solar radiation and auroral sources generate an active redistribution of energy through the pressure gradients created. Equation 1.6 is the equation of motion for horizontal winds with velocity $U$.

$$\frac{dU}{dt} = F - 2\Omega \times U - \nu_{ni} (U - V) + \frac{\mu}{\rho} \nabla^2 U + g$$

Equation 1.6 can be broken down into several component forces. Here $F = -\frac{1}{\rho} \nabla P$ is the driving force due to a horizontal pressure gradient where $\rho$ is the mass density and $P$ is the pressure.

$2\Omega \times U$ is the Coriolis force where $\Omega$ is the angular velocity of the rotating Earth.

$\nu_{ni} (U - V)$ is the force due to ion drag caused by collisions between neutral particles and ions. It therefore depends on the frequency of the collisions, which is given by the ion-neutral collision frequency coefficient $\nu_{ni}$, and the relative velocity between neutrals and ions which move with velocity $V$.

$\frac{\mu}{\rho} \nabla^2 U$ is the force due to viscosity where $\mu$ is the coefficient of viscosity.

The thermosphere may be treated as a single fluid when applying the equations of motion. It is not necessary to take account of each individual species because any difference in the motions of the constituents is much less than the overall wind flow.

The direction of flow of winds is dependent on the ratio of the Coriolis force to ion drag [Rishbeth and Garriott, 1969, Rishbeth, 1988]. At lower altitudes, such as the lower ionosphere and troposphere, where the winds are strongly controlled by the Coriolis force and where ion drag is small, the winds circulate along the isobars around pressure ‘highs’ and ‘lows’, because the horizontal pressure gradient balances the Coriolis force. This is called ‘geostrophic flow’, where the magnitude of the wind velocity at a latitude $\phi$ is calculated by equation 1.7a.
In contrast, at the F-region peak where the frictional forces of viscosity and ion drag are large, the winds tend to flow along the pressure gradients. The ion drag term is important above 150 km where the ionospheric plasma is motion is largely determined by the geomagnetic field [Rishbeth and Garriott, 1969]. The magnitude of the wind velocity in this case is given by equation 1.7b, where $I$ is the magnetic dip angle.

$$U = \frac{F}{2\Omega \sin \phi}$$

(1.7a)

At altitudes where both the Coriolis force and ion drag are similarly significant the angle $\psi$ between the direction of wind flow and $F$ is given by equation 1.7c, and the magnitude of the wind is given by equation 1.7d.

$$\psi = \arctan \left( \frac{2\Omega}{\nu_{ni} \sin \phi} \right)$$

(1.7c)

$$U = \frac{F}{\sqrt{\nu_{ni}^2 \sin^2 I + 4\Omega^2 \sin^2 \phi}}$$

(1.7d)

Whatever the type of wind flow the winds are always subject to the continuity equation which states the principle of conservation of mass (equation 1.8), where the rate of change of the gas concentration, $n$, depends on the rate of production and loss of the gas, $P$ and $L$ respectively, and the transport of the gas out of the region of observation.

$$\frac{dn}{dt} = P - L - \nabla \cdot (nU)$$

(1.8)
1.6 Solar Heating

Solar extreme ultra-violet flux (EUV, $\lambda \leq 1000\,\text{Å}$) is the primary source of heating of the thermosphere [Torr and Torr, 1985, Hedin and Mayr, 1987]. EUV radiation is absorbed during the day mainly between 100 km and 200 km. This is indicated in Figure 1.4 which shows the altitudes where solar flux in the ultra-violet wavelength range is attenuated to $e^{-1}$ of its incident value for vertical incidence [Chamberlain, 1978]. This figure also shows the primary species responsible for absorption at different wavelengths. Although the heat input is very small compared with the ultra-violet (UV, $1000\,\text{Å} < \lambda \leq 2000\,\text{Å}$) and visible solar radiation absorbed at lower altitudes, the density of the thermosphere is so little that the temperatures generated are high. Various paths are taken in converting the EUV radiation energy into heating of the neutral atmosphere.

The main photoionization reactions in the middle and upper thermosphere (above 150 km) are the photoionization of atomic oxygen and molecular nitrogen:

\[
\begin{align*}
O + h\nu &\rightarrow O^+ + e^- \\
N_2 + h\nu &\rightarrow N_2^+ + e^-
\end{align*}
\]

The ejected photoelectron undergoes Coulomb collisions with other electrons and ions and inelastic collisions with neutrals. Direct collisions with neutrals cause heating with 5% efficiency [Killeen, 1987]. Collisions with other electrons produce a hot electron gas. The majority of the photoelectron energy is lost to space through airglow emissions from particles that have been excited by inelastic collisions.

The energy carried in the ions produced by photoionization takes a more complicated path through various exothermic ion-molecule reactions which ultimately end in the dissociation of $O_2$ into two oxygen atoms.

\[
O_2 \rightarrow O + O
\]

These diffuse down to the lower thermosphere (between 90 km and 150 km) and upper mesosphere and release their energy of dissociation in a three-body recombination process, where $M$ is a third body which carries away the excess energy. Thus some of the energy of photoionization in the middle and upper thermosphere is lost to lower altitudes.

\[
O + O_2 + M \rightarrow O_2 + M
\]
The absorption of ultra-violet flux in the Schumann-Runge continuum of molecular oxygen, which is the spectral region between 1300Å and 1750Å, is the dominant heat source for the lower thermosphere in the altitude region 90 – 150 km. Metastable atomic oxygen in the O(1D) state is produced by photon dissociation of molecular oxygen. The atomic oxygen releases its excess energy to the neutral atmosphere on being collisionally quenched with about a 33% local heating efficiency [Torr et al., 1980a,b].

The heat generated during the day is conducted away to the denser, cooler atmosphere below 100 km at night and is subsequently lost through infra-red radiation cooling to space. The great importance of radiational cooling was proved by Roble and Emery [1983] who calculated the global mean temperature profile from the AE satellite measurements of solar UV and EUV fluxes using the latest evaluations of the neutral gas heating efficiencies. They found that the temperature profiles agreed with empirical model data for solar minimum, but at solar maximum the temperatures were too high. Agreement could only be achieved by including a significant amount of radiational cooling by the 5.3 µm band due to enhanced levels of [NO]. Thus the vertical temperature distribution is mainly a balance between the absorption of solar energy and loss through thermal conduction.

The earliest theoretical models considered the effects of solar heating alone on the Earth’s atmosphere [Jacchia, 1964, 1965, Geisler, 1966, Dickinson et al., 1968]. The Jacchia model, shown in Figure 1.5, was derived from the first satellite drag results which showed a high pressure region near the sub-solar point due to solar heating which was causing the atmosphere to expand on the sunlit side of the Earth. This high pressure region is often referred to as the ‘daytime pressure bulge’. In the Jacchia model the maximum temperature is 1300 K, occurring at about 14:00 Local Time (LT) and the minimum temperature is 1000 K occurring at about 03:30LT. Both extremes of temperature lie on the equator. Solar heating would result in pressure gradients that would drive winds as a mass movement of air to restore equilibrium. At auroral latitudes, thermospheric winds, driven purely by the pressure gradients, were expected to flow away from the subsolar point over the poles to the nightside of the Earth. Figure 1.6 shows wind velocities calculated from the Jacchia temperature distribution by Kohl and King [1967]. The average wind speed was predicted to be 200 m/s, with a peak equatorward meridional component occurring at 02LT.

Solar driven models were satisfactory as a first order description of the
global thermospheric wind flow. At low- and mid-latitudes the largest input of energy to the thermosphere comes from solar heating so the winds calculated by Kohl and King are a good approximation of the experimental results. However, in order to simulate the daily variation in temperature and the magnitude of the wind velocities found at high latitudes it was necessary to include an additional heat source at high-latitudes which was subsequently found to have a far more complicated origin, namely the magnetosphere.

1.7 Solar Cycle Variations in Solar Flux

Solar UV and EUV fluxes vary with the solar cycle. The variation in UV is only a few percent but the EUV fluxes can change by a factor of 2-3 as is shown in Figure 1.7 [Torr et al., 1980, Hinteregger, 1977, 1981, Hinteregger et al., 1981]. Since the forces controlling the thermosphere are all carefully balanced it becomes necessary to know how changes in the intensity of the solar fluxes affect that balance.

Solar EUV radiation does not reach the earth’s surface because it is absorbed by the intervening layers of atmosphere, so it can only be measured by satellite or rocket. This results in a very small quantity of intermittent data which makes it a poor long-term index. For the analysis of the solar cycle dependence of the average thermospheric neutral winds smoothed monthly averages of the radio flux at a wavelength of 10.7 cm were used in proxy as an index of the EUV flux and hence of the solar cycle. The 10.7 cm radio flux is presented as the $F_{10.7}$ flux index with units of $10^{-22}$ W/m$^2$/Hz by the World Data Centre. The change in the $F_{10.7}$ flux over the period 1981 to 1990 is illustrated in Figure 1.8.

However, the correlation of the 10.7 cm flux with UV and EUV flux is not perfect since the radio fluxes come primarily from the corona of the sun while UV flux comes from the upper photosphere and the region spanning the chromosphere up to the solar corona is the source of solar EUV flux (Figure 1.9) [Akasofu and Chapman, 1972]. Therefore each flux band is generated by somewhat different physical processes. However, the differences are mainly in short term variations [Lean, 1988]. The 27 day variations of several UV emission lines compared with the 10.7 cm radio flux is shown in Figure 1.10, from Hall and Hinteregger [1970]. The running mean of the $F_{10.7}$ flux is better than the daily flux values and is an adequate representation of the EUV flux as discussed in the paper by Hedin [1984].

Other ground-based measurements have been put forward as better esti-
mates of the EUV flux than the 10.7 cm radio flux. For example, while the $F_{10.7}$ indices correlate very well with the emission from calcium plage in the active regions of the solar chromosphere Oster [1983a,b], the equivalent line-width of the He I line at 1083 nm was shown by Hinteregger [1981] to give a better estimate of solar UV flux. Despite this the $F_{10.7}$ index remains in strong use because the improvements in prediction given by the other indices are not sufficient to compensate for the paucity of their databases, for the $F_{10.7}$ indices provide the most complete long-term database of solar flux available. Results have been routinely collected since 1947 and it is consequently the most widely used index for correlation with thermospheric measurements.

For the analyses presented in Chapters 3 and 4 the exact correlation between the $F_{10.7}$ index and the EUV flux is not important. The data are sorted into two very broad levels of solar activity representing solar minimum and solar maximum. A boundary of $F_{10.7} = 110$ was chosen to divide the dataset conveniently into equal sized portions.

1.8 Ionospheric Convection at High-Latitudes

The global circulation of the atmosphere, due to the pressure gradients set up by solar heating, drives thermospheric winds anti-sunwards over the poles, from the dayside of the Earth to the nightside [Kohl and King, 1967]. In the absence of electric fields ions are constrained by the Earth’s magnetic field and cause a retarding drag on the flow of the neutral gas. However, at high latitudes an electric field is set up by the coupling between the magnetosphere and ionosphere so that the thermospheric winds at high latitudes can be either accelerated or retarded significantly by ionospheric convection within the auroral region and polar cap by momentum transferred from the convecting ions to the neutral gas through collisions, [Hays et al., 1979, Killeen et al., 1986, Rees et al., 1987a, Smith et al., 1985, 1986, 1988a].

The magnetospheric electric field is set up by the interaction between the magnetosphere and the solar wind which forms a magneto-hydrodynamic generator [Hones, 1986]. This electric field maps down into the ionosphere to drive the ions into a two-cell convection pattern [Fedder and Banks, 1972, Foster, 1984]. In the F-region the collision frequencies between the ions and neutrals are much smaller than the gyrofrequencies and so the plasma drift velocity, $v$, is determined by the electric field, $\mathbf{E}$, and the magnetic field, $\mathbf{B}$, according to equation 1.9 [Rishbeth and Garriott, 1969]:
\[ \mathbf{v} = \frac{\mathbf{E} \wedge \mathbf{B}}{|\mathbf{B}|^2} \] (1.9)

As can be seen from Figure 3.26, the average two-cell convection pattern is sun-aligned with the centre of the pattern offset to the nightside of the Earth by about 4°. This means that the ionospheric convection pattern moves in and out of sunlight as the Earth rotates since the geomagnetic poles in the northern and southern hemispheres are separated by about 9° and 15° latitude from the geographic poles.

The magnetosphere is a highly dynamic system and ionospheric convection responds rapidly to any changes in the processes governing magnetosphere-ionosphere coupling [de la Beaujardière et al., 1987, 1988, Etemadi et al., 1988]. The size and shape of the ion convection pattern is determined by the orientation of the Interplanetary Magnetic Field (IMF) vector [e.g. Heelis, 1984, McCormac et al., 1985, 1991, Rees et al., 1986, Foster et al., 1986b, Heppner and Maynard, 1987] and, as a consequence, the level of geomagnetic activity [e.g. Oliver et al., 1983, Lloyd, 1985, Rees et al., 1987b,c, Sica et al., 1986a]. Various empirical models are available which correlate ionospheric convection with these processes. For instance, illustrated in Figure 3.26 is a set of equipotential convection contours for four levels of particle precipitation derived from the Millstone Hill radar data by Foster et al. [1986]. With increasing particle precipitation the contours become closer and closer together, representing ever steeper gradients in the electric field. Such a model describes the gross behaviour of the ionosphere.

The convection pattern is composed of strong antisunward plasma flow across the polar cap and sunward plasma flow in the dusk and dawn flanks of the auroral oval. Through ion-neutral energy and momentum transfer the thermospheric neutral winds can be drawn into the ionospheric convection pattern. The extent to which this can happen depends partly on the position of the neutral gas parcel in the convection pattern. Momentum transfer is much stronger in the dusk sector than the dawn sector. Fuller-Rowell and Rees [1984] ascribe the difference in response between the dawn and dusk sectors to the ‘inertial resonance effect’. The natural motion of the winds in the Northern Hemisphere due to the earth’s rotation and the Coriolis force is for a clockwise rotation with a velocity proportional to colatitude. The driving force of ion drag in the dusk sector matches this motion and so a gas parcel in this region will be contained here for many hours building up speed and leading to large winds.
the dawn sector the winds are driven in an anticlockwise direction. The effect of the Coriolis force is to throw the gas parcel equatorward, out of the influence of the auroral oval, within a few tens of minutes. Hence, despite equal forcing, the asymmetry in the response of the high latitude neutral winds through the night.

1.9 High-Latitude Heating

High latitude heating sources were found to be essential to make theoretical models of the upper atmosphere match with empirical models, especially during geomagnetic disturbances. The simple model of mean global temperatures by Jacchia [1965] derived from satellite drag results (Figure 1.5) was superceded by models such as that by Roble et al. [1977], derived from satellite measurements of UV and EUV fluxes. The increased sophistication of the Roble et al. [1977] model includes both a seasonal and solar cycle variation as shown in Figure 4.4.

The source of high latitude heating is the solar wind which deposits energy via the magnetosphere into the auroral regions. High latitude heating is calculated to be a very significant proportion of the total heating rate of the thermosphere. Figure 1.11 from the paper by Roble and Emery [1983] shows their calculations of typical height profiles of heating rates which agree quite well with observed exospheric temperatures. The energy flux from the solar wind is less than that from UV radiation but the magnetosphere presents a large surface of interaction which increases its contribution proportionately [Prölls et al., 1988]. The solar wind energy is dissipated through the ionosphere in two forms: particle precipitation and Joule heating. With particle precipitation the injection of protons and electrons into the high-latitude thermosphere heats the atmosphere through frictional and collisional excitations. The total global thermospheric heating rate depends on the activity level, and is typically within the range of 0.05 - 0.6×10^{11} Watts. These estimates come from satellite measurements of average precipitating particle fluxes [Spiro et al., 1982, Hardy et al., 1984, 1985, 1991, Gussenhoven et al., 1984, Evans et al., 1987a, Fuller-Rowell and Evans, 1987]. Over 50% of the particle precipitation energy input results in neutral heating according to the calculations of M.H.Rees et al. [1983].

Joule heating arises through the electrical resistance to current flow within the auroral oval. Joule heating varies linearly with $K_p$. The average Joule heating over one hemisphere at equinox for $K_p = 1$ was about $0.25\times10^{11}$ and for $K_p = 4$ was about $0.85\times10^{11}$ as calculated from empirical models by Foster.
et al. [1983]. This is greater than heating by particle precipitation. However, since particle precipitation increases the conductivity within the auroral oval, and thereby the Joule heating, the two heating processes are not independent and therefore not easily distinguishable.

A statistical survey shows that both Joule and particle heating reach peak values in a region that is roughly oval-shaped (hence the name of 'auroral oval') and centred on each of the two magnetic poles as shown in Figure 1.12 [Feldstein and Starkov, 1967]. The oval is sun-aligned and displaced towards the nightside. It has a diameter of about 40° with a latitudinal extent of 5°–10°. The diameter and width of the oval ring vary according to the geomagnetic activity but the oval rarely extends below 60° invariant latitude.

There can also be significant heating within the inner boundary of the auroral oval, called the polar cap, depending on the conductivity. In the summer, when the polar cap is sunlit for 24 hours, there is evidence that the solar EUV ionization is sufficient to support significant Pedersen currents here [Fujii et al., 1981].

The instantaneous distribution of Joule heating and particle precipitation is very different from the statistical averages. It is a far more complex pattern, representing the curious development of a substorm. One of the most obvious evidences of the high latitude energy influx is through aurora. Akasofu published his extensive studies on auroral activity based on all-sky camera pictures in 1964 (Figure 1.13) [Akasofu, 1964, 1968]. These show how intricate the patterns can be and how quickly they change during the development of a magnetospheric substorm.
1.10 Large Vertical Winds

Vertical wind flow is small compared with horizontal flow, and the magnitude of the average vertical thermospheric wind has been generally within the limits of resolution of FPI observations. This has been the main reason why vertical motions were not indisputably verified until the early 1980s with the observations by Hernandez [1982b], although there were a few observations made in the previous decade by Rieger [1973] and Spencer et al. [1976] of very high vertical winds.

The vertical distribution of air pressure conforms to hydrostatic equilibrium so that the vertical pressure gradient is balanced by gravity (equation 1.1). Vertical flow is attributed to three possible reasons:

(i) thermal expansion or contraction of the atmosphere
(ii) convergent or divergent flow of the atmosphere within the horizontal plane which leads to a downward or upward motion, respectively, in accord with the continuity equation
(iii) escape of gas from the top of the atmosphere

Only hydrogen and helium are light enough gases to escape from the Earth’s atmosphere, but they are minor constituents and so (iii) is not usually considered a significant cause. Particular interest has been paid to (ii) since convergent or divergent flow of the atmosphere is caused by localised heating, which is a significant cause of thermal up-welling within the auroral oval [Rishbeth et al., 1987, Rees et al., 1984c].

Thermospheric winds measured from a single site are calculated using the assumption that any vertical wind component is insignificant compared with the horizontal component, and that any systematic upflow or downflow that might occur is not long-lived and is spatially localised. However, Fuller-Rowell [1987] showed that the whole region of the auroral oval correlates closely with significant vertical winds. The vertical winds are the result of a combination of diverging horizontal wind flow and a change in the height of the pressure surface. The influence of the auroral oval is therefore likely to dispute the validity of this assumption when applied to the vertical wind data from the three Scandinavian sites. Observations of large vertical winds of 20 – 30 m/s concurrent with converging or diverging horizontal wind fields have even been observed at a mid-latitude site by Biondi [1984], which would have implications for the data from Utah. In addition there is evidence for vertical winds induced
by gravity waves at both auroral sites [Rees et al., 1984c,d] and mid-latitude sites [Hernandez, 1982b].

1.11 Auroral and Airglow Emissions Observed With the UCL FPI

Airglow and aurora are separate phenomena, although both arise through photochemical processes in the upper atmosphere. Airglow is virtually unstructured and occurs at all times and at all latitudes. It is mainly activated by solar flux in the ultra-violet and X-ray spectrums and at night has a greater intensity than the total intensity of starlight [Hargreaves, 1979]. Meanwhile the aurora is highly structured, being one of the manifestations of the entry of energetic charged particles into the upper atmosphere at high-latitudes during solar and geomagnetic disturbances. Of the three auroral and airglow emissions monitored at present by the UCL instruments the results that will be presented in this thesis refer only to the atomic oxygen transition at a wavelength of 6300Å and the emission from the Meinel band at 8430Å.

1.11.1 Atomic Oxygen Emission at 6300Å

The red line emission of atomic oxygen at a wavelength of 6300Å has been studied for many years since it was first identified in the nightglow spectrum [Cabannes, 1935]. It is one of the most prominent features in the visible wavelength region of the auroral spectrum, and belongs to a triplet of forbidden line transitions between the ground configuration states $^3P - ^1D$. As such the oxygen red lines have been used to classify the aurora and estimate the peak emission height of the auroral and particle emission spectra [Rees and Roble, 1986].

The estimation of the peak emission height of the 6300Å emission profile is based on the chemical processes producing the excited parent metastable state O($^1D$), which has an excitation energy of only 1.96eV [e.g. Bauer, 1973]. There are several possible mechanisms which might produce O($^1D$) owing to the small excitation energy involved. The dominant source of O($^1D$) during daylight hours (zenith angles < 105°) is photodissociation of O$_2$ by the absorption of ultraviolet radiation in the Schumann-Runge continuum, which lies in the wavelength region 1300-1750Å [Dalgarno and Walker, 1964, Noxon and Johanson, 1972]:

$$O_2(X^3Σ^-_g) + hν → O_2(B^3Σ^-_u) → O(^3P) + O(^1D)$$

However, during the nighttime hours auroral electron precipitation is the only source of O($^1D$) apart from persistent nightglow [Rees and Roble, 1986]. The
airglow charge exchange between $\text{O}^+$ and $\text{O}_2$ followed by the dissociative recombination of $\text{O}_2^+$ accounts for most of the nighttime production rate of $\text{O}(^1\text{D})$ [Bates, 1946, Cogger et al., 1974, 1980], though this process is of little significance in the sunlit atmosphere [Hays et al., 1978]. Dissociative recombination of $\text{O}_2^+$ is the first of the seven mechanisms listed below which contribute to the auroral production of $\text{O}(^1\text{D})$ proposed by Rees and Roble [1986]:

(i) dissociative recombination of $\text{O}_2^+$ ions

$$\text{O}_2^+ + e^- \rightarrow \text{O} + \text{O}(^1\text{D})$$

(ii) electron impact by energetic auroral electrons

$$\text{O} + e^- \rightarrow \text{O}(^1\text{D}) + e^-$$

(iii) excitation by electrons in the high-energy tail of the ambient thermal population through electron-impact

$$\text{O} + e^- \rightarrow \text{O}(^1\text{D}) + e^-$$

(iv) electron-impact dissociation of $\text{O}_2$

$$\text{O}_2 + e^- \rightarrow \text{O} + \text{O}(^1\text{D})$$

(v) transition from higher electronic states

$$\text{O}(^1\text{S}) \rightarrow \text{O}(^1\text{D}) + h\nu(5577\text{Å})$$

(vi) $\text{N}(^2\text{D}) + \text{O}_2 \rightarrow \text{NO} + \text{O}(^1\text{D})$

(vii) $\text{N}^+ + \text{O}_2 \rightarrow \text{NO}^+ + \text{O}(^1\text{D})$

Figure 1.14 compares the volume emission rate height profiles for the 6300Å emission with and without auroral electron precipitation [Sica et al., 1986a]. The average peak emission height assumed in the interpretation of the data presented in this thesis is 240 km, which is the peak emission height for the airglow. As the figure shows, auroral precipitation can lower the height quite significantly. Figure 1.15 shows the individual contributions of the above seven processes to
the volume emission rate of the 6300 Å emission as calculated by Rees and Roble [1986] for auroral electron precipitation with a Maxwellian spectrum. Of these auroral processes dissociative recombination of O$_2^+$ ions has been calculated to be the major source of the 6300 Å emission above a height of around 280 km [Rees and Roble, 1986], while the reaction between N($^4$D) and O$_2$ has been proposed by [Rusch et al., 1978] to account for 80% of the observed emission below 200 km.

Due to the long lifetime of the intermediate O($^1$D) state, the energy can be lost through collisions with other atoms and molecules before the 6300 Å quantum of energy can be emitted, this is called quenching [Rusch et al., 1975]. Quenching increases at lower altitudes as the densities of molecular nitrogen and oxygen increase. The curves of Figure 1.15 have been calculated to allow for quenching of O($^1$D) by collisions with molecular nitrogen which is the major chemical species at these altitudes:

$$N_2 + O(^1D) \rightarrow N_2 + O(^4S)$$

1.11.2 Hydroxyl Radical Emission at 8430 Å

The hydroxyl radiation was first observed by Slipher [1929] and eventually identified to be the rotational-vibrational bands of the excited hydroxyl radical, OH*, by Meinel [1950] (and consequently named after him). The hydroxyl radical has up to nine vibrational levels and the combined emission from all the bands dominates the night airglow [Krasovskij and Šefov, 1965]. The atmospheric emissions in the infra-red spectral region between 5200 Å and 4.0 μm are all primarily due to hydroxyl bands, with some contribution from O$_2$ bands [Gush and Buijs, 1964, Vallance Jones, 1964]. Two mechanisms have been proposed for creating this excited molecular state: the O$_3$ – H process by Bates and Nicolet [1950]:

$$H + O_3 \rightarrow O_2 + OH^*$$

and the Krasovskij process [Krasovskij and Šefov, 1965]:

$$H + O_2 \rightarrow O + OH^*$$

However, the Krasovskij process was shown to be a minor reaction, unlikely to contribute more than 1% of the total hydroxyl emission [Hesstvedt, 1968].
Rocket and satellite borne instruments show that the intensity of emissions from the Meinel bands have maxima in the height region between 70 km and 95 km. Figure 6.11 shows the altitude distribution of the hydroxyl emission for three different ranges within the Meinel bands obtained from rocket flights [Packer, 1961, Tarasova, 1961, Tarasova and Slepova, 1964]. The emission concentrated on here has a wavelength of 8430 Å, coming from the excited \( \text{P}_1(3) \) state in the \((6,2)\) Meinel band. The emission peaks at around 88 km, and has a FWHM of about 6 km [Thomas and Young, 1981, Baker and Stair, 1988], which is just below the mesopause at 90 – 95 km [von Zahn and Kurzawa, 1989]. This makes it a suitable measure of the winds of the mesosphere.

Generally the hydroxyl \((6,2)\) emission band consists of weak emissions in the infra-red, with intensities of 400 – 600 Rayleighs. This has made them very difficult to observe with a standard FPI because photomultipliers and photon detectors, such as the S-25 detector, have had very poor Detective Quantum Efficiencies in the infra-red spectral region due to the low energies involved [Smith et al., 1988b]. Through the infra-red capabilities of the GaAs detector an FPI set up at a mid-latitude site has been able to produce a fine set of observations of the hydroxyl emission and consequently the winds of the mesosphere [Rees et al., 1990]. The results of this work are presented in Chapter 6.
Figure 1.1

[Diagram showing the upper atmosphere based on temperature, composition, escape, and ionization.]

Nomenclature of the upper atmosphere based on temperature, composition, escape, and ionization.

[Hargreaves, 1979]
Typical vertical profiles of the mid-latitude ionosphere. [Hargreaves, 1979]

Distribution of major gases in the thermosphere for an exospheric temperature of 800 K. [COSPAR International Reference Atmosphere, 1972]
The altitude at which the rate of absorption of solar radiation is at a maximum. The principle atmospheric constituents that absorb the radiation in the different wavelength bands are indicated. [Handbook of Geophysics and the Space Environment, 1985]
Figure 1.5

Global temperature distribution in the thermosphere at equinox in conditions of medium solar activity. No account is taken of high-latitude heat sources due to energetic particles. [Jacchia, 1965]

Figure 1.6

Winds at 300 km computed from the temperature distribution of Figure 1.5. [Kohl and King, 1967]

Figure 1.7:
IRRADIANCE RATIO: SOLAR MAXIMUM/SOLAR MINIMUM

[Diagram showing the ratio of solar spectral irradiance with wavelength (Å) on the x-axis and irradiance ratio on the y-axis, with bars indicating the ratio for different wavelength intervals.]
Figure 1.8
Figure 1.9

The structure of the sun, internal and atmospheric. [Akasofu and Chapman, 1972]

Figure 1.10

The 27-day variations in several UV emission lines observed from the OSO 3 satellite in May and June 1967. The 10.7 cm radio flux for the same period is also given for comparison [Hall and Hinteregger, 1970]
Global mean heating rate profiles [Roble and Emery, 1983]

Variation in the size of the auroral oval with activity as denoted by the magnetic index $Q$. Coordinate system is CG latitude and CG local time. [Feldstein and Starkov, 1967]
Figure 1.13

The substorm in the luminous aurora. (a) $T = 0$, (b) $T = 0 - 5$ minutes, (c) $T = 5 - 10$ minutes, (d) $T = 10 - 30$ minutes, (e) $T = 30$ min - 1 hour, (f) $T = 1 - 2$ hours. [Akasofu, 1968].
Comparison of model 6300Å emission profiles with and without auroral electron precipitation. [Sica et al., 1986]

Computed altitude profiles of the [OI] 6300Å volume emission rate produced by an incident Maxwellian electron spectrum with characteristic energy of 0.4 keV and total energy input rate of 2 ergs/cm²/s. The numerals identify the various processes that contribute to the total volume emission rate: 1. dissociative recombination of O²⁺ ions; 2. electron impact by energetic auroral electrons; 3. electron impact by the thermal population; 4. dissociation of O₂ molecules; 5. cascading from the O(5) level; and 6. the reaction between N(2D) atoms and O₂ molecules. The reaction between N⁺ ions and O₂ molecules is shown by a dashed line, 7., to indicate a tentative contribution. [Rees and Roble, 1986]
2.1 Introduction

Two criteria in particular determine the choice of instrument used for observations of the upper atmosphere. The first requires an instrument of high spectral resolution because the Doppler shifts associated with winds typical of the upper atmosphere are very small. For example, an average wind of 200 m/s in the upper thermosphere would produce a Doppler shift of 4.2 mA on the red-line emission at 6300 Å. The second criterion requires high sensitivity due to the weakness of emissions from the airglow and aurora of the upper atmosphere, which are typically of the order of $10 - 1000$ kiloRayleighs ($1$ Rayleigh $= 10^6$ photons/cm$^2$/s). For example, the intensity of the red line emission at 6300 Å ranges between $20kR$--$50kR$ [Shemansky and Vallance Jones, 1968], while the intensity of the green line emission at 5577 Å is around $100kR$ [Vallance Jones, 1964]. In comparison the intensity of full moonlight is of the order of $1000kR$ [Hargreaves, 1979]. Consequently daylight observations are extremely difficult to produce since these emissions are swamped by the background intensity from sunlight. The Fabry-Perot Interferometer (FPI) has been chosen for satisfying both criteria, but its particular advantage over conventional prisms or grating spectrometers is that for a given resolving power the FPI can transmit a much greater light flux than the others [Jacquinot, 1954].

The emissions observed tend to be those of forbidden transitions of neutral and ionized species. The lifetimes of the excited states are long but, owing to the rarity of the atmosphere at such heights, they can survive being de-excited by collisions with other particles through other, permitted, electronic states. The length of their lifetimes allows the excited states to reach statistical equilibrium with the surrounding atmosphere. Thus the radiation that is eventually emitted during de-excitation reflects both the kinetic temperature in the Doppler width and the bulk motion in the Doppler shift. As a result the principles of experimental technique and data analysis whereby the winds and temperatures can be derived from the emission line profiles are simple and direct, requiring very few a-priori assumptions. However, the self-sufficiency of the derivation has
the drawback that little else can be learned of the atmosphere apart from the
temperature and wind speed.

The technique of observing spectra using a FPI is well established, dating
back to the turn of this century [Fabry and Perot, 1899]. The basic instrument
has been refined and improved over the years with the addition of complex
reflective coatings on the etalon surfaces and various means of recording the
displayed fringe patterns. A simple description of the FPI is given here but
reference should be made to a text such as ‘Optics’ by Hecht and Zajac [1980]
or, for a full and detailed description, to the classic text ‘Principles of Optics’
by Born and Wolf [1987].

The first use of FPIs for observing the upper atmosphere was made by
Babcock in 1923, who attempted measurements of the Doppler width of the
OI \((1D - 1S)\) transition at night [Babcock, 1923]. The measurement of upper
atmospheric winds from Doppler shifts was then pioneered by Armstrong [1968]
over thirty years later, since when the use of FPIs has become well established.

The earliest observations were recorded on photographic plates with limi-
ted success owing to the low quantum efficiency and non-linear behaviour of
photographic emulsions. Jacquinot and Dufour [1948] introduced the use of
photomultipliers as detectors which greatly improved the efficiency of the light
detection. The original photomultiplier instruments used pinhole detectors to
detect the signal while the optical path length was changed by pressure scanning
or by varying the plate separation. Attempts to increase the etendue of the FPI
by multiple element masks (according to the original suggestions by Jacquinot
and Dufour [1948]) and image intensifiers have not been wholly successful due
to the huge effort needed to build and maintain such instruments and then anal-
lyse the data. The success of the UCL instrument has been to solve this problem
by the development of a two-dimensional Imaging Photon Detector with a high
optical performance, excellent long-term stability and extremely low noise.
2.2 The UCL Fabry-Perot Interferometer

Figure 2.1 is a schematic diagram of the UCL Fabry-Perot Interferometer. The FPI is composed of a scanning mirror, a Fabry-Perot etalon, a focusing telescope, an interference filter, the Imaging Photon Detector (IPD) and a calibration source. The components are mounted on a rigid four-bar optical bench which maintains the alignment and focus of the instrument. Outside there is a signal processing unit and thermal control unit, both for the IPD, and a personal computer of at least an XT specification running the electronic controller unit for the whole instrument. Normally the FPI is set up to observe through a hole in the roof of a building at the observing site, with a plexiglass dome to protect the instrument from the elements.

Light is deflected into the interferometer by a mirror set at a zenith angle of 140°, which has been changed to 135° for special observations during rocket releases. The field of view is 1° so only light from the sky travelling at a zenith angle of 60° ± 0.5° can enter the interferometer.

The mirror rotates in a clockwise direction through eight positions, surveying the North, North-East, East, Calibration Lamp, South, Zenith, West and North-West, pausing in each position to take an image. The mirror is driven around by a motor and opto-sensors pick up marks on the rim at the top of the FPI which indicate the stopping positions. The mirror can be commanded to halt for any length of time while the IPD is collecting photons. Different integration times are used for each FPI, determined by three main criteria which are:

(i) the local average intensity of the 6300Å emission
(ii) the sensitivity of the IPDs used
(iii) control of the quantity of data being collected

The FPI containing the most sensitive detector, apart from the new Gallium Arsenide (GaAs) detector, is housed in the Institute of Space Physics at Kiruna. The integration time is 60 seconds and it therefore produces a large volume of data which is regularly backed-up by our colleagues at the Institute.

The FPI at Kilpisjarvi is not as sensitive as the one in Kiruna and so the integration time is longer at 180 seconds, and the quantity of data collected is less.

The instrument on Svalbard is visited only once a year so it is necessary to keep the data to a minimum quantity so as not to overflow the hard disk.
before the operatives arrive. Thus the IPD used here has the lowest sensitivity and the integration time is long at 360 seconds. However, in order to maximise the spatial coverage the FPI observes only the four cardinal directions plus the zenith and calibration lamp.

Future plans include use of optical discs, with their massive storage capabilities, to allow the instruments to run at optimum sensitivity and minimum integration times without running out of space.

A particular strength of the UCL interferometer has been due to the development of a stable and rugged Fabry-Perot etalon and its surrounding mounts [Rees et al., 1981, Killeen et al., 1982b]. Stability is very important for a non-scanned system since calibration is achieved by cross-matching the observed emissions with emissions from standard sources such as the neon calibration lamp used with the UCL instruments.

The etalon used is composed of two fused silica etalon plates set at a fixed distance apart using three zerodur spacers. Zerodur is a polycrystalline ceramic with a very low coefficient of thermal expansion of \(0.2 - 0.4 \times 10^{-6} /K\). The spacers can be individually adjusted to optimize the parallelism of the etalons. Once set the spacers maintain the gap with a high degree of thermal stability [Killeen et al., 1982b]. The plate separation is either 10 mm or 14 mm and the diameter is 15 cm, thus giving a very large aperture. The etalon plates are polished flat to \(\lambda/200\) and are coated with a dielectric semireflecting layer. The reflectance is 0.81 at 6300Å.

The etalon is mounted in a sealed and evacuated container where the temperature is carefully controlled by the computer to be stable to ±0.1°. Thus the effects of changes in pressure and temperature on the refractive index of the air in the etalon gap, and consequently the optical path length, are minimised.

The IPD provides the UCL instrument with a powerful advantage over the small, on-axis sampling apertures used in conventional interferometers. The IPD is an electronic photon counting device which allows a complete two-dimensional image to be observed and stored. Thus all of the light impinging on the etalon is used and the complete image is an average of the emission spectrogram over the integration period. In comparison the pinhole detector can only give consecutive measurements of intensity along the fringe profile.

The IPD consists of a photocathode, with a diameter of either 18 mm or 25 mm, followed by a proximity focussed lens which projects the photo-electrons
onto a system of three consecutive \( z \)-configuration microchannel plate intensifiers. Eventually the electron cloud reaches a resistive anode, which is a sheet with four contacts around its edges. The \( x \) and \( y \) coordinates of the electron cloud may be calculated by comparing the ratios of the charge received by each contact. The error on the position is calculated to be about \( \pm 13 \mu m \), on using the estimated Gaussian point spread function of \( 25 \mu m \) quoted by Greenaway et al. [1982]. The detector output of the photon position is converted into a digital two-dimensional coordinate array of dimension \( 256 \times 256 \). Details of the IPD are given in Rees et al. [1980, 1981], McWhirter et al. [1982] and Greenaway et al., [1982].

The most commonly measured upper atmospheric emission has a wavelength of 6300Å. This is an emission in the visible red and therefore is within the limits of the Detective Quantum Efficiency (DQE) of the first detectors used, which had a S-20 specification. The longer the photon wavelength the lower the energy and hence the more difficult to observe using a photocathode. The limits of the emissions that can be observed were pushed into the near infra-red when devices with a full S-25 photocathode specification become available.

The latest improvement to the IPD has been the installation of a GaAs photocathode, which is a state-of-the-art device with three very important features:

(i) high DQE of 20%

(ii) sensitivity which extends beyond 950 nm (c.f. S-25 photocathodes which are limited to 800 nm)

(iii) the thermionic emission of the entire device of \( 256 \times 256 \) pixels at \(-25^\circ C\) is only 80 counts/s, which is almost negligible.

The production of the GaAs detector has opened up new vistas, at last allowing the atmospheric community the opportunity of observing not only the near infra-red, but also very faint emissions within this and previously available wavelength regions. Unfortunately there are very few GaAs detectors available as yet, but of the two that UCL possesses one has already been used very successfully at a mid-latitude site since August 1989. This instrument has been used to observe the infra-red OH emission from the Meinel band at 8430Å, and the 6300Å emission, which at this latitude is caused by airglow alone and consequently is extremely weak.
Further, the GaAs detector could close the gap in the quality of observations of weak infra-red emissions made by satellites and ground-based instruments. Ground-based observations have previously been limited by the near-IR background night-glow continuum. Satellites have been able to circumvent this problem by making observations of the limb of the Earth’s atmosphere which suffers less from this background emission, with the added bonus that the Van Rhijn effect enhances the spectral emission intensity. However, although satellite instruments can get better results, cost dictates that ground-based instruments will always play a significant role, especially as satellite instrument technology is always several years behind due to the timescales controlling their deployment.

2.3 Data Analysis

2.3.1 Etalon Theory

The Fabry-Perot etalon consists of two glass prisms set so that the two surfaces facing each other are perfectly flat and parallel, separated by a distance $d$. The two prisms are slightly wedge-shaped to reduce interference from any reflections off the outer surfaces, while the two inner surfaces are coated with a metal or dielectric film for high reflectance $R_f$. Thus as light enters the gap it is predominantly multiply reflected between the two inner surfaces, but with each reflection a small portion of light is transmitted out. The components of the wave that eventually emerge from the etalon are not only different in amplitude from one other but different also in phase, which is determined by the angle of incidence, $\theta$, and the size of the etalon gap.

Outside the etalon a lens gathers the wave components together onto the viewing screen. The path difference determines how the exiting wave components interfere with one another. For a wavelength $\lambda$ a path difference of $n\lambda$ will result in constructive interference, where $n$ is an integer number, while a path difference of $(n + \frac{1}{2})\lambda$ will result in destructive interference. Equation 2.1 is the equation for constructive interference, where $\theta_n$ is the angle of the $n$th order of constructive interference:

$$2d \cos \theta_n = n\lambda$$

The effect of constructive and destructive interference, when all the components leaving the etalon are allowed to interfere, is the concentric ring pattern shown in Figure 2.2. The formula for the total intensity $I_{tot}$ in terms of the peak
intensity $I_o$, depends on the reflectance and phase difference, $\delta$, and is given by equation 2.2, which is called the Airy function:

$$I_{tot} = \frac{I_o}{1 + \frac{4R_f \sin^2 \frac{\delta}{2}}{(1-R_f)^2}} \tag{2.2}$$

Figure 2.3 is a plot of intensity against phase difference, illustrating equation 2.2. The sharpness of the peaks is determined by the reflectance i.e. the higher the reflectance the greater the difference between the maximum and minimum values. Figure 2.3 includes the Airy functions for several different values of reflectance.

### 2.3.2 Collecting Data

The light rays impinging on the Fabry-Perot etalon come at angles within the 1° field of view. This allows the small angle approximation to be used in calculating the positions of the fringes. A direct proportionality between a peak at angle $\theta$ and the square of the radius of the peak, $R$, may be derived, resulting in the Airy function of equation 2.2 translating simply into a function dependent on $R^2$. Storing the full image would take up too much computer memory and is unnecessary for the FPI since the whole image is coming from the same patch of sky within a 1° field of view. Therefore once the two-dimensional image is collected it is reduced to a one-dimensional image in terms of the $R^2$ values and stored as a one-dimensional array of data. In contrast, the Doppler Imaging System (DIS) uses a wide angle lens to observe over a field of view of 100° so that the part of the sky that is within this field of view maps directly onto the image, therefore making each portion of the image unique [Batten, 1989, Rees et al., 1984a, 1989]. Yet the memory allocation required to store the 2-D images is in excess of the capabilities of the present generations of PCs and consequently, in order to minimise the quantity of data collected yet without losing spatial resolution, the DIS images are divided into 24 segments, each of which is reduced to a $R^2$ image of one-dimension.

The centre of the image is found very carefully during the calibration procedure which is carried out at least once a year when operatives go on field trips to visit all the sites. It is essential to get the best possible value for the centre since all the data will be reduced to radius-squared form using radii determined from this centre position. Fortunately the FPIs are very stable and there is little drifting of the centre throughout the observing period. The main cause of a change of centre is when the FPI is dismantled for any maintenance work.
Figure 2.4a shows an actual full two-dimensional image of the neon calibration lamp observed by the FPI at Kilpisjarvi on the 1st December 1990. The image is usually far from the ideal of Figure 2.2, with rings that may be more oval than circular. Distortion due to the etalon is negligibly small. What distortion there is is largely due to incompatibility between the gains in the \( x \) and \( y \) directions of the resistive anode, or any slight misalignment of the IPD or lenses.

Figure 2.5 shows an example of a real reduced to radius-squared image which should be compared with the Airy function of Figure 2.3. The image chosen is the reduced to radius-squared profile of Figure 2.4a. The degradation of the real data from the mathematical ideal is severe and a great deal of work has been undertaken to reconcile the two, for example the series of papers over the past 25 years by Hernandez [1966, 1970, 1974, 1978].

2.3.3 Correcting Data

Before the peaks of the reduced to radius-squared images are fitted the data must be corrected using a Flat-Field image and Thermionic Emission image, examples of which are shown in Figures 2.4b and c respectively, collected on the 1st December 1990 at Kilpisjarvi. A Flat-Field is an image collected from a diffuse white light source in order to monitor the response of every pixel of the detector. Each pixel of the eventual data image is subsequently normalised using the Flat-Field to remove the instrument-dependent factor of the fringe profile. The Thermionic Emission image is used to measure the ‘dark counts’. To obtain a Thermionic Emission image the whole instrument must be totally screened from light while the IPD takes a reading. What ‘photons’ that are collected are due to the natural electronic noise of the instrument. Cooling the photocathode, which is the main source of noise, reduces the thermionic emission. In earlier instruments the ‘dark counts’ could be quite significant and there would often be a few ‘hot spots’ where there was a large charge concentration due to dust particles or irregularities in the surface of the resistive anode. Data from ‘hot spot’ regions had to be rejected. Nowadays, long experience and improved technology have produced IPDs that are quiet and uniform, producing excellent images. In addition the environment of the IPD is very carefully controlled in that the entire interferometer is encased in a thermally controlled, light-tight enclosure, where the temperature is maintained to within about \( 3^\circ - 5^\circ \)C of a fixed temperature. This temperature is usually chosen to be around \( 5^\circ \) above room temperature,
which in most cases is a setting of around 25°C. However, the GaAs detector needs to be cooled to −20°C or −30°C to reduce the thermionic emission to a negligible level. A Peltier cooler is used in this case, with circulating water to remove the heat from the hot junction.

In spite of these improvements, the electronics in the IPD will invariably deteriorate over the years, in which case the calibrating measurements become more and more important as can be imagined on seeing Figure 2.4d, which is the Flat Field image of an IPD near the end of its lifetime, measured on the 30th January 1989 at Longyearbyen.

2.3.4 Fitting Peaks

Once the data have been corrected the analysis process can begin. A function is fitted to the peaks to obtain values of the radius, line intensity, line width and background intensity. There are several methods of analysing the data in use such as that used by Killeen and Hays [1984] for a 12 segment, multi-ring anode Image Plane Detector. The function used to fit the peaks here is described in Lloyd [1985]. It is based on a deconstruction of a FPI intensity profile by Hernandez [1966] into several components. The image recorded by the IPD is a distortion of the original profile by the finite bandpass width of the instrument. It can therefore be considered to be a convolution of several functions. The emission line being observed has a Voigt profile $B$ which is a convolution of a Gaussian profile, the width of which reflects the thermal temperature of the emitting gas, and a Lorentz profile, which represents the effects of pressure broadening on the gas. Biondi and Fiebelman [1968] have confirmed that the 6300Å emission is predominantly Gaussian.

On passing through the FPI the emission line profile is convolved with the instrument profile. The instrument profile is a convolution of several components broken down by Hernandez into a succession of Gaussian and Tophat profiles. These are:

$$A = \text{the Airy function of the ideal etalon}$$
$$D_f = \text{departure from flatness of the plates}$$
$$D_g = \text{microscopic flatness imperfections}$$
$$F = \text{effect of the exploring diaphragm}$$

Thus the final observed profile Y is given by the following convolution:

54
\[ Y = B \ast A \ast D_f \ast D_g \ast F \]  

(2.3)

Each of these functions serve to broaden the original source profile. Lloyd [1985] bypassed consideration of each individual function by creating one function which is a convolution of a Tophat and Gaussian function to encompass the components \( B \ast D_f \ast D_g \ast F \). By a process of iteration this function can successfully fit the top half of the FPI data peaks, over a range of 40-50 bins, to an accuracy of up to 0.01 of a bin which is equivalent to an error of less than 1 m/s. This fitting routine is fast and efficient but cannot cope with finding Full-Width Half Maximum values from which temperatures are derived. This has not been a handicap because we have not been able to afford the frequency-stabilised lasers necessary to produce the \( \delta \) function profile that could give a direct measure of the instrument function profile.

There is another fitting routine developed by Batten [1989] which deals with the data in Fourier space, and therefore has the potential to be modified to search out temperatures. A smooth fit may be made over the whole data profile, as opposed to just one peak, using a cosine bell (Hanning) filter [Koopmans, 1974] to cut out high frequency components. The accuracy of the peak positions is of a similar order of magnitude to that of the Lloyd routine but the data processing takes twice as long. In general the Lloyd routine is favoured for analysing most of the data where the data peaks are fairly simple in form, but the Batten routine is used to analyse the OH emission data at 8430Å from Utah owing to the presence of additional peaks due to contaminating emissions from other sources and also the double peak of the calibration lamp. The two routines cannot be alternated due to a consistent difference in their calculations of the peak positions when using the same input parameters necessary for their working. (These input parameters are the Free Spectral Range and the Offset, which will be discussed later in this chapter.) However, if one or the other fitting routine is used, with its own version of the input parameters, then there is little difference between the final derived wind speeds since these values are dependent on the relative peak positions rather than the absolute positions.
2.3.5 Calculating Wind Speeds

Figure 2.6 shows the plots of peak position against Universal Time for each of the viewing directions. A red Doppler shift moves the peaks closer in to the centre of the image while a blue shift moves the peaks outwards. Thus on assuming a constant overhead wind an observation to the North, say, might show a red or blue Doppler shift, revealing that the wind component along the line-of-sight shows an air mass that is either retreating from or advancing towards the observer, respectively. Meanwhile the observation from the South would show a line-of-sight wind component representing an air mass moving in the opposite direction. Consequently when the peak positions from these two directions are plotted on the same graph, as in Figure 2.6, the peaks move symmetrically in opposite directions. Any asymmetry in the positions would imply that the wind field is not perfectly uniform, which is not really surprising considering that the radius of the observing circle is nearly 400 km.

The wind speed $v$ giving rise to a Doppler shift $\Delta \lambda$ on the emission line at wavelength $\lambda$ may be calculated from equation 2.4, where $c$ is the speed of light:

$$v = \frac{c\Delta \lambda}{\lambda}$$  

(2.4)

Yet first it is necessary to know the value of $\Delta \lambda$, which is represented by a shift of the peak from a zero Doppler shift or baseline position. This is a difficult and sometimes contentious procedure. Ideally each instrument would be calibrated with a laboratory source of the emission line being monitored. There is as yet no suitable source of atomic oxygen emitting at 6300Å owing to the difficulties of maintaining a source that originates from a forbidden transition with a very long lifetime of around 110 seconds [Garstang, 1951]. When a laboratory sample source is not available it is important that the alternative calibration source chosen can be transmitted through the same filter, which means that it must be of a similar wavelength. This ensures that the spectral characteristics of the instrument are the same for both emissions. The calibration source used for the 6300Å line is a neon radio-frequency excited spectral lamp which emits at 6304Å. This is a very cheap and reliable source, having a long life-time and being easily installed in the instrument. A neon calibration lamp has also been used as a temporary measure to calibrate the OH emission at 8430Å until a He-Ne laser is installed and working. The neon lamp gives a double peak in this wavelength region which consequently has a larger error associated with the
peak position than the OH peak position, but apart from that is reliable as a steady calibration source.

The calibration lamp peaks are used to indicate the stability of the instrument. The peaks are narrower than the emission line owing to the lower temperatures of their source, and therefore may be fitted more accurately. The centre coordinates of the fringes are measured from calibration lamp fringes during the initial set-up procedure at the beginning of each winter campaign. During the observation period the calibration lamp is referred to once every cycle. The data are also used as a diagnostic which helps the planning of the preventative maintenance undertaken during the yearly field trips. Drifts in temperature can be easily differentiated from drifts in the pressure within the etalon due to leakage. Temperature drifts manifest themselves as fluctuations in the calibration lamp peak position about a mean position determined by the ambient temperature, while a leak causes the calibration lamp peak to move gradually and inexorably outwards. Except in a few cases the calibration lamp can be used to compensate successfully for any such failures in the instrumentation. These will be discussed under section 2.4, which is labelled ‘Problems’, at the end of this chapter.

The zero Doppler shift or baseline is determined using the assumption that the average vertical component of the thermospheric winds is a small fraction of the horizontal component. Generally this is true since the vertical wind components are small, averaging at less than 10 m/s, while the horizontal wind components average out at 200-300 m/s. However, larger vertical winds do occur, particularly at high latitudes, for example average downward winds of 30-50 m/s for periods of 4-6 hours as well as short bursts of magnitude 100-150 m/s have been reported over Northern Scandinavia by Rees et al. [1984a] and similar observations over Mawson, Antarctica (67.6°S, 62.9°E) by Wardill and Jacka [1986].

The baseline peak positions are calculated using both the calibration lamp peak positions and the zenith peak positions. The computer routine that determines the best possible method of calculating the baseline passes through an order of criteria determined by what data are available. The vertical winds are too variable to use on their own to determine the baseline, therefore the optimum method uses the calibration lamp peak positions plus a constant offset value to create a baseline. The expected value of the calibration lamp peak position at the time when the emission peak position was observed is calculated
by linear interpolation from the observed calibration lamp peak positions.

The best method for calculating the offset for a night takes an average of the separation of the zenith peak positions and the calibration lamp peak positions through the night. If the calibration lamp is available but there is too little zenith data to get a good average for the offset then a long-term average offset from several days of data is used to obtain the baseline. If the calibration lamp fails but zenith data is available, a parabola is fitted to the zenith peaks to create the baseline since the fluctuations of the zenith peaks are too much for them to be used alone as a baseline. The fluctuations are caused by the natural variability of the vertical winds. In addition the line-of-sight observation of the zenith cuts a shorter path through the emission layer than the observation of the azimuth at a zenith angle of 60°. Thus the intensity of the emission observed in the zenith is less which results in a poorer signal to noise ratio and consequently a larger error on the fit.

All the peak positions are compared with the baseline to calculate the line-of-sight winds. The factor which converts Doppler shifts of the peaks in radius-squared space to wind speeds in metres per second is derived from the Free Spectral Range (FSR). The FSR, $\delta f_{sr}$, is the change in wavelength needed to generate a set of peaks that will overlap the original set of peaks due to an emission at wavelength $\lambda$, but offset by one order. The wind speed $v$ is then calculated from the ratio of the change in $r^2$ due to the Doppler shift with the equivalent free spectral range change in $R^2$, using equation 2.5:

$$v = \frac{(r'^2 - r^2) \ c\lambda}{(R_n^2 - R_{n+1}^2) \ 2d}$$

(2.5)

The FSR is not significantly affected by Doppler shifts of a magnitude typical of thermospheric wind data. For instance the FSR for a wavelength of $\lambda = 6300\,\text{Å}$ and etalon gap of $d = 15\,\text{mm}$ is $142\,\text{mA}$. The Doppler shift caused by an average thermospheric wind speed of $v = 200\,\text{m/s}$ is $\Delta \lambda = 0.0042\,\text{mA}$. This change in wavelength would modify the FSR by $7 \times 10^{-5}\%$, which is negligible compared with the peak fitting errors. So, within the limits of our observations, an average FSR from the calibration lamp data will be the same as that from any of the other directions. It has not been possible to display two full peaks for both the 6300Å and calibration lamp emissions. The years when the calibration lamp, rather than the 6300Å emission, has two full peaks displayed produce a more
consistent measure of the FSR since the calibration peaks are narrower and therefore are amenable to a more accurate fit of the data.

The assumption that the vertical wind component is negligible is carried through to the derivation of the horizontal wind components. Thus the line-of-sight winds from the observations made at zenith angle of 60° are converted to horizontal winds by dividing by \( \sin 60° \). The consequence of this assumption is generally fairly small and is discussed under the ‘Problems’ section at the end of this chapter. Figure 5.2 is a sample illustration of the line-of-sight horizontal wind components obtained from the three Scandinavian sites.

At present line-of-sight observations of thermospheric winds are obtained from a FPI at a single site. To get true wind vectors there would need to be more than one observing site for the purposes of triangulation. However, it is possible to get a fairly good representation of the wind field for most of the sites in the following manner where a horizontal wind vector, \( F(\theta) \), is calculated by fitting a simple sinusoidal curve to the components of the horizontal wind observed in several evenly spaced directions which have bearings which are an integer multiple of an angle \( \phi \). The method used is that of the Fourier Series with only the first two terms utilised. This assumes that the wind field is uniform over the whole of the field of view but to assume otherwise would introduce a subjective bias. Thus the function used is:

\[
F(\theta) = a_1 \cos \theta + b_1 \sin \theta
\]

Each of the parameters \( a_1 \) and \( b_1 \) would be calculated from the integration of the data points as directed in any mathematical text. In this context there are at most six data points and so a summation procedure has been substituted.

The function \( F(\theta) \) represents a wind vector with magnitude \( W \) and bearing \( \phi \), so that:

\[
F(\theta) = W \cos(\theta - \phi)
\]
This equation may be expanded to:

\[ F(\theta) = W \cos \theta \cos \phi + W \sin \theta \sin \phi \]  

(2.8)

which, when compared with equation (2.6) gives:

\[ a_1 = W \cos \phi \]
\[ b_1 = W \sin \phi \]

Thus the magnitude and bearing of the wind are given by equations 2.9 and 2.10 respectively:

\[ W = \sqrt{a_1^2 + b_1^2} \]  

(2.9)
\[ \phi = \tan^{-1} \left( \frac{b_1}{a_1} \right) \]  

(2.10)

The wind magnitude and bearing may be calculated from the six line-of-sight observations available at Kiruna, Kilpisjarvi and Utah. However, the IPD at Longyearbyen is not very sensitive and the integration time is set to be long, in which case the four cardinal directions alone are observed in order to maintain maximum coverage. On moonlit nights the FPI does not observe in the South, where the moon is, in order to protect the IPD from overloading, though the increase in the background count would make this a very poor quality observation anyway. The method shown above of calculating a wind vector is not appropriate for less than five data points, thus a modified method is used. The average value of the wind components from observations to the North and South is used as the meridional component and the average of the observations to the East and West is used as the zonal component. If one of these is not available then the total value of the observation in the opposite direction is used as the component value. From this meridional and zonal average, magnitudes for the components in the NE, NW, SE, and SW are generated. Finally the Fourier Series method is used to fit the function \( W \cos(\theta - \phi) \) to this new set of data points and then calculate the overall horizontal wind vector. This procedure serves to smooth somewhat the final calculated vectors since one cycle of observations can take from 10 to 30 minutes to complete depending on the site.
2.4 Problems Encountered During Data Analysis

This following section outlines several problems that had to be addressed over the years during which the FPIs at various sites have made observations of the thermosphere and mesosphere.

2.4.1 Instrument Temperature Drift

The FPI is mounted in a container that is thermally controlled so that the temperature is maintained within 3-5° C and the etalon itself has its own control so it is stable to better than 0.1° C [Rees et al., 1982]. Normally the system works well and any small temperature drifts are easily compensated for using the calibration lamp data. Occasionally things do go wrong. Unfortunately when trouble occurs with instruments at remote sites nothing can be done until the next field trip a year later. For instance, the data from Kilpisjarvi during the winter of 1986-87 was made useless by a fault in the temperature control. The temperature control overcompensated in its reaction to temperature changes which led to sinesoidal-type oscillations superimposed on the data being monitored. The period of the temperature oscillation was around the same length of time as the period of rotation of the mirror which made it impossible to extract the oscillation from the data. This is shown by Figure 2.7 which has the peak positions from all directions plotted along the same universal time (UT) axis.

Another instance of poor temperature control occurs with the FPI at Utah. The cause is the drastic plummeting of temperature at nightfall which is typical of a mid-continent site. The FPI is in a brand new observatory which does not appear to maintain the temperature of the room containing the instrument. The instrument temperature control cannot cope with the huge temperature changes as, for example, can be seen from the data peaks collected on the 22nd October 1989 (Figure 2.8). In such cases the calibration lamp can be used fairly satisfactorily to correct for the systematic temperature variation. The details of the correction used are given in subsection 2.4.4 which deals with the specific problems encountered with the FPI at Utah.
2.4.2 Changes in the Refractive Index of the Etalon Gap Due to Pressure Leaks

The gap between the etalon plates is evacuated with a vacuum pump when the FPI is first set up at the beginning of a campaign. However, it was discovered from the behaviour of the peak positions that the etalons in the instruments at Kilpisjarvi, Kiruna and Utah have leaks, the magnitude of the fault being large, medium and small respectively. The leaks resulted in a steady increase of the peak positions of the emission lines with time due to the change in pressure within the etalon gap as air leaked in. The rate of drift of the peaks is given in Table 2.1.

Equation 2.11 is the general equation giving the condition for maximum intensity in the Airy function. Equation 2.11 differs from equation 2.1 by the introduction of the refractive index, \( \mu \), of the medium within the etalon gap. Equation 2.1 is, in fact, the special case for an evacuated etalon where \( \mu = 1 \). It is assumed in all the calculations to obtain Doppler wind speeds that the etalon is evacuated and so only equation 2.1 is used. However, a change in the pressure between the two etalon plates will cause a change in \( \mu \). It is therefore important to know by how much \( \mu \) changes over the period of observation. If there is a significant change in the optical path length then any calculation of Doppler shifts will need to include a measurement of the refractive index.

\[
2\mu d \cos \theta_n = n\lambda \tag{2.11}
\]

The variation of the radius-squared value of the peak position, \( r^2 \), with time, \( t \), is linear, as can be seen in Figure 2.9b, and therefore can be represented by equation 2.12.

\[
r^2 = Mt + C \tag{2.12}
\]

Values for the constants \( M \) and \( C \) were then found from a least squares fit to the average daily variation of \( r^2 \). The small angle approximation is used to replace \( \theta \) in equation 2.11 by the radius of the peak and the focal length of the converging lens when manipulating the equations for the FPI. Eventually, a function giving the time dependence of the refractive index in terms of known and calculated quantities may be derived, using \( F(t) \) for the FSR. The change in \( \mu \) from time \( t = 0 \) to time \( t \) is given by equation 2.13.
\[
\mu(t) - \mu(0) = \frac{(r^2(t) - r^2(0))}{(\frac{2d}{\lambda} - \frac{\lambda}{C})} \\
= \frac{Mt}{(\frac{2d}{\lambda} - C)}
\] 

(2.13)

It is reasonable to assume that at the beginning of the season when the FPI is first set up there is a good vacuum in the etalon gap so that \(\mu(0) = 1\). Table 2.1 shows what the refractive index would be expected to be after 365 days. This is a worst possible case since no instrument is left to languish without attention for much more than a year before being overhauled and recalibrated. As can be seen from the results the change in \(\mu\) is extremely small in all cases. So far, because we are monitoring relative changes in peak positions and not absolute changes, such errors in calculating the optical path length due to the leak are much less than the errors due to peak fitting. Consequently the approximation of equation 2.1 still applies quite adequately.

**Table 2.1**

**The Rate of Drift of Peak Positions**  
**Due to Etalon Pressure Leak**

<table>
<thead>
<tr>
<th>Location</th>
<th>Period of Data</th>
<th>rate of drift</th>
<th>(\mu(365))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kilpisjarvi</td>
<td>8/12/88 (\rightarrow) 11/1/89</td>
<td>1 bin every 7.4 hours</td>
<td>0.9998</td>
</tr>
<tr>
<td>Kiruna</td>
<td>7/11/86 (\rightarrow) 19/4/87</td>
<td>1 bin every 4.3 days</td>
<td>1.00001</td>
</tr>
<tr>
<td>Utah</td>
<td>30/8/89 (\rightarrow) 25/12/89</td>
<td>1 bin every 5.9 days</td>
<td>1.000007</td>
</tr>
</tbody>
</table>
2.4.3 Unstable Free Spectral Ranges

It is possible to derive a simple inverse relation between the refractive index and the FSR in terms of the etalon gap and focal length of the converging lens, \( L \). This is given in equation 2.14. In which case the rate of change of the refractive index due to a pressure leak could have been calculated more directly from the FSR than from \( r^2 \).

\[
\mu = \frac{\lambda L^2}{dF}
\]  

(2.14)

However, it was found that the FSR calculated from real data varied too much to give a realistic idea of how \( \mu \) varied due to air leaking into the etalon gap. Thus the linear variation in \( r^2 \) gave a far more reliable measure of the rate of change in \( \mu \), as has been described in the previous section. The cause of the departure from the theoretical prediction of the relation between \( \mu \) and \( F \) is instrumental. There is an inevitable distortion of the image by the electronics of the IPD and any distortion becomes magnified towards the edges. Figure 2.10 shows a grid pattern observed with the IPD to show the resolution and distortion. The image quality is good near the centre but at the edges the straight lines are distorted to appear curved. In addition the IPD is not so sensitive at the edges, as can be seen from a flat field image such as Figure 2.4b. Despite uniform illumination of the detector the IPD registers a maximum intensity near the centre which fades off towards the edges. As a result the second peak is broader and less intense than the first peak in \( r^2 \) space, resulting in larger errors when fitting the peak.

A further problem with the second peak is that in \( r \) space the second peak is narrower than the first. This means that when the 2D image, which is in \( r \) space, is reduced to \( r^2 \) form any distortion from the ideal circular fringes will make a greater difference to the reduced second peak than the first because the distortion of the peak position around the circle will therefore be a larger fraction of its width.

The variation observed in the FSR which is caused purely by instrumental distortion is demonstrated very well by the 6300Å observations at Kiruna over the winter of 1987 to 1988. Figures 2.9a and b show the variation of the FSR and peak positions, respectively, over this period. On the 8th November the instrument was recalibrated and a new centre coordinate was found, hence the jump in the value of the FSR compared with the value used in September 1987. Between November 1987 and the end of February 1988 the FSR fluctuated by around two bins about a steady mean value, during which time both the first
and second peaks were in positions less than bin 230. The FSR computed from the difference in the two peak positions varied randomly about a mean value due to the distortion by the IPD electronics, but the spread in values was about 1%, so the distortion was not great.

Then after the 25th February the FSR changed quite dramatically (Figure 2.9b). At this time the second peak was very close to the edge of the detector and the FSR value jumped up by five bins and then dropped down by about seven bins until the end of the campaign in mid-April. The severity of the change in the FSR when the second peak was close to the edge of the detector showed that the detector distortion in this region was no longer trivial. It could be deduced that for this particular IPD the value calculated for the FSR could no longer be relied upon when the second peak was close to or beyond bin number 230.

2.4.4 Problems Particular to the Observation of the 8430Å Emission

A laser had been purchased for calibrating the OH data. Unfortunately it broke down almost as soon as it was run, but by good fortune the neon lamp used to calibrate the 6300Å data had an emission at a wavelength that passed through the filter for the OH line. However, this neon emission produced two broad, closely overlapping peaks as can be seen in Figure 2.11. The peak fitting routine used for the 6300Å data was not able to provide an adequate measure of the peak position and so a Fourier Transform method of peak fitting devised for the Doppler Imaging System by Batten [1989] was used for both the calibration and data peaks. This Fourier Transform method is superior in versatility to the other routine but, due to its mathematical complexity, is about three times slower and more demanding in its requirements for initial parameters. So far it has been used only for the analysis of the OH data and not for the data from the six other FPI instruments which observe the 6300Å and 5577Å emissions.

The overall method of analysis of the OH data is the same as for the 6300Å data. However, the lack of observations of the zenith and the broad double peak of the neon calibration lamp demanded various modifications of the procedure used, which are described below.
2.4.4.1 Determination of the 8430Å Free Spectral Range

The OH emission at 8430Å is extremely weak. The intensity is of the order of 100 R, therefore the size of the etalon gap was chosen to overlap the emission at 8430Å with its doublet pair in order to maximise the intensity. This resulted in a FSR that was too large to allow the satisfactory display of two full peaks of the 8430Å emission or its calibrating emission. The FSR is important to the conversion from a Doppler shift in terms of bin position to the equivalent wind speed in metres per second. In order to determine the value of the FSR it was necessary to refer to the FSR of the 6300Å emission observed on the 29th August 1989, the sole night when two clear peaks were observed with the GaAs detector. From this value of the FSR the value of the FSR could be calculated for the 8430Å emission. However, each time the FPI was recalibrated and the centre coordinates were changed it was necessary to calculate the new FSR using a series of interconnected equations which ultimately relate the new FSR to the 6300Å FSR. There were four changes in the centre coordinates used for the winter of 1989-1990. They were small adjustments of about ±2 — ±3 pixels for a total image of 256 x 256 pixels. An error in the initial value of the 6300Å FSR would introduce a systematic error in the factor used to convert Doppler shifts which are measured in terms of the radius-squared parameter to the equivalent wind speeds. This is likely to introduce a systematic error of about 2% for a FSR of around 135 bins. The equations used to calculate the FSR values are as follows:

The diameter of the detector is 20 mm and the focal length of the lens used to focus the image from the etalon is 150 mm. It is therefore valid to use the small angle approximation for cos θ in equation 2.1 and so relate the wavelength of the emission λ directly to the value of the radius r of the peak position, as shown in equation 2.15.

\[ n\lambda = 2d(1 - \frac{r_n^2}{2L^2}) \] (2.15)

The free spectral range, FSR, is the distance between the peak positions of successive orders, n and n + 1 so that \( FSR = r_n^2 - r_{n+1}^2 \). Equation 2.15 can be applied to two different wavelengths \( \lambda_1 \) and \( \lambda_2 \) to give equation 2.16:

\[ \frac{\lambda_1}{\lambda_2} = (\frac{d_1/L_1^2}{d_2/L_2^2})\frac{FSR_1}{FSR_2} \] (2.16)
If a calibration source with a wavelength $\lambda_{\text{cal}}$ is brought in and the distance between the data source peak position and calibration peak position $R_m$ is given by $\text{OFFSET} = r_n^2 - R_m^2$ then further manipulation of the equation 2.15 gives equation 2.17:

$$\frac{\lambda_1}{\lambda_2} = \frac{(m\lambda_{\text{cal1}} - n\lambda_1)}{(M\lambda_{\text{cal2}} - N\lambda_2)} \frac{FSR_1}{FSR_2}$$  \hspace{1cm} (2.17)

More simply though, from equation 2.16 it can be seen that, provided there is no change in the focussing parameter, i.e. $d/L^2$ is constant, the FSR may be determined from a simple comparison of the ratios of the wavelengths of the emission line and the calibration lamp:

$$\frac{\lambda_1}{\lambda_2} = \frac{FSR_1}{FSR_2}$$  \hspace{1cm} (2.18)

The choice of the equation used depends on what parameters remain constant and what have changed. Such variables as the focal length of the focussing lens or the actual etalon used are rarely altered once the FPI is set up at a site. Generally, the only instrument parameters to be altered during a campaign are the centre coordinates of the full Fabry-Perot ring pattern or the gain on the $x$ and $y$ potentials of the IPD, which can be adjusted to enlarge or contract the image. This effectively changes the value of $d/L^2$, in which case the free spectral range, $FSR_1$, must be found using equation 2.17.

The centremost order of the Airy function is determined solely by the size of the etalon gap i.e. the maximum number of whole wavelengths that can fit in. The etalon gap sizes used are 10 mm, 14 mm and 20 mm and the typical wavelength observed is the 6300Å emission. Therefore the value of the order of the central peak $n$ is several tens of thousands. Since $d$ is known to, at most, four significant figures it is not possible to pinpoint the absolute value of $n$. The Benoit method [Born and Wolf, 1987] could be used, but it requires at least two other sources of accurately known wavelengths to calibrate the measurement. Consequently the only practical use of equation 2.17 is when $\lambda_1 = \lambda_2$, in which case the unknown FSR may be found from the ratios of the offsets, as shown in equation 2.19:

$$\frac{FSR_1}{FSR_2} = \frac{OFFSET_1}{OFFSET_2}$$  \hspace{1cm} (2.19)
Thus although a FSR could be measured only from data collected on the 29th August 1989 for the 6300Å emission when two full peaks were displayed, it was possible to calculate the FSR for the 8430Å data collected subsequently, where one peak alone was displayed, through the following procedure:

1) First the average FSR was measured for the 6300Å data observed on the 29th August 1989.

2) The centre coordinates for all subsequent nights were too different from that of the 29th August to use equation 2.18 to calculate the FSR. However, equation 2.19 could be applied to the data collected between the 1st to the 5th October 1989 since the wavelengths are the same i.e. 6300Å.

3) The centre coordinates for the 8430Å observations taken on the 14th October 1989 were similar to the centre coordinates of the 6300Å measurements taken on the night of the 1st October 1989. It was therefore possible to assume that the focussing ratio $d/L^2$ remained constant and use equation 2.18 to calculate the FSR.

4) The average offsets for the remaining periods of 8430Å data varied quite a lot, thereby indicating that it was not appropriate to assume that $d/L^2$ was constant between them all in order to use equation 2.18. Therefore, for the data affected by the centre coordinate changes on the 30th and 31st August and the 3rd September the FSRs were found by using the FSR calculated for the 14th October and then applying equation 2.19. The final values calculated for the FSRs are given in Table 2.2, and also the equivalent wind speeds for a peak shift of one bin.

### Table 2.2

Calculated Free Spectral Ranges for the 8430Å data.

<table>
<thead>
<tr>
<th>date</th>
<th>$\lambda$/Å</th>
<th>$x,y$</th>
<th>OFFSET /bins</th>
<th>FSR /bins</th>
<th>1 bin = speed (m/s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>29/8/89</td>
<td>6300</td>
<td>123,133</td>
<td>$-35.80 \pm 0.2$</td>
<td>131.9 ± 4.</td>
<td>34.9 ± 1.</td>
</tr>
<tr>
<td>30/8/30</td>
<td>8430</td>
<td>128,134</td>
<td>60.48 ± 0.5</td>
<td>196.3 ± 8.</td>
<td>31.4 ± 1.</td>
</tr>
<tr>
<td>31/8/89 → 2/9/89</td>
<td>8430</td>
<td>128,135</td>
<td>60.73 ± 0.3</td>
<td>197.1 ± 8.</td>
<td>31.3 ± 1.</td>
</tr>
<tr>
<td>3/9/89 → 26/9/89</td>
<td>8430</td>
<td>129,137</td>
<td>57.72 ± 0.1</td>
<td>187.3 ± 7.</td>
<td>32.9 ± 1.</td>
</tr>
<tr>
<td>1/10/89 → 5/10/89</td>
<td>6300</td>
<td>131,131</td>
<td>$-37.30 \pm 0.5$</td>
<td>137.4 ± 5.</td>
<td>33.4 ± 1.</td>
</tr>
<tr>
<td>14/10/89 → 8/11/89</td>
<td>8430</td>
<td>129,133</td>
<td>56.66 ± 0.1</td>
<td>183.9 ± 7.</td>
<td>33.5 ± 1.</td>
</tr>
</tbody>
</table>
2.4.4.2 Determination of the 8430Å Baseline

As with the 6300Å emission there is no laboratory source of the 8430Å emission available to give the peak position for a zero Doppler shift, from which the Doppler shifts may be calculated. From theoretical considerations the average vertical winds are small compared with the horizontal winds, therefore the usual procedure used for the 6300Å observations has been to calculate average baseline peak positions from the peaks observed in the zenith and the calibration lamp peaks. Unfortunately, owing to physical constraints on the deployment of the FPI at Utah it has not been possible to make zenith observations. Therefore a process of averaging the 8430Å data from all the horizontal directions has been used to create the best estimate of the baseline, under the assumption that there is no divergence of the wind field.

It is necessary to use the calibration lamp data to provide a measure of the instrument stability, yet a large error ensues when determining the position of such a poorly defined peak as that of the double peak emission from the calibration lamp at Utah. The standard method used to find a baseline for the 6300Å data propagates the large calibration errors through to the derivation of the Doppler shifts of the 8430Å peaks. The errors on the peak positions of the 8430Å data are the equivalent of 1—2 m/s while the errors on the calibration peaks are around 10—15 m/s. Thus the excellent quality of the 8430Å data would be spoilt by the neon calibration lamp. The standard method is acceptable for the analysis of the 6300Å data because the calibration lamp emission at 6305Å has a narrower FWHM value than the 6300Å peak, but for the analysis of the 8430Å data two alternative methods have been devised in an attempt to maintain the high quality of these observations:

(i) The first method uses a constant peak position for the baseline. This peak position is the sum of the average calibration lamp peak position for the whole night and an offset. The offset is calculated by finding the difference between the calibration peak and the average peak position from all the azimuthal directions for every 15 minutes and then calculating a long-term average over several days. This method reduces the effect of the large error of the calibration peak position on the calculation of the baseline. For instance, if the calibration peak error is \( \epsilon_{cal} \) and there are \( n \) observations made throughout the night, the error of the average calibration peak position is then only \( \epsilon_{cal}/n \).

(ii) The second method finds the average peak positions of the zonal and
meridional observations at 15 minute intervals. If there is no gradient in the wind field then the average zonal and meridional peak positions should each give the position of zero Doppler shift since, for instance, a red shift observed in the south should appear simultaneously as a blue shift in the north.

Unfortunately method (i) relies on the temperature stability of the instrument and assumes that there are no pressure leaks so that the calibration peaks are distributed normally around a true average peak position. (The stability of the detector itself is the result of careful and deliberate design and construction, proved through many years of observation and therefore is not called into doubt [Rees et al., 1981, 1982].) This method could be applied satisfactorily to the data collected between the 30th August 1989 and the 1st September 1989, inclusive, when the FPI was at the Centre for Atmospheric and Space Sciences, and the calibration peak positions showed that the instrument was in a stable, controlled environment. However, it was not appropriate to use a constant value baseline in circumstances where there was a significantly large variation or a drift in the calibration peak positions throughout the night, as was the case when the instrument was subsequently moved to the Hardware Ranch Observatory near Bear Lake (41.9°N, 111.4°W) on the 2nd September 1989. From that date onwards there was a dramatic swing in the calibration lamp peak positions every evening. For example, Figure 2.8 shows that on the 16th October 1989 there was a steep drop in the calibration lamp peak position between 01 UT and 03 UT. This may be attributed to the change in the ambient temperature as the temperature outside the hut dropped from day-time to night-time values. At Utah, which is in the centre of a continent, this temperature drop can be as much as 30°C which is beyond the temperature stabilizing capabilities of the controls within the hut.

As a result method (ii) was devised to account for the large temperature swings of the data collected at Bear Lake. In this method the calibration lamp data was not used and so the large error in the peak position was not incorporated into the determination of the baseline. However, the observations in the seven directions were not taken simultaneously since the integration time was 240 seconds, so that one complete cycle through the seven observing directions, plus the calibration lamp, would take at least 32 minutes (allowing time for the rotation of the mirror). As a consequence, even if a constant wind field existed overhead, it would be unlikely to remain at the same value over the whole of
such a long observing cycle. Thus it should not be surprising that observations in opposite directions do not show equivalent Doppler shifts. In addition there did appear to be a significant gradient in the wind field at certain times during the night. Therefore, although the data were not marred by the large calibration lamp error, there was an error due to an ignorance of the true baseline. This error could not be measured but could only be minimised according to the rule of the continuity equation 1.7b, by making the divergence of the horizontal flow of wind as small as possible.

2.4.4.3 Contamination of the 8430Å Emission

The bandwidth of the pre-filter used in 1989 was quite broad at 3 nm. Normally a 1 nm bandwidth filter has been used with the other FPIs but a broad filter was used initially in order to transmit the emission from the calibration lamp. Unfortunately the filter was also found to transmit a thermospheric emission resulting in occasions of severe contamination of the 8430Å emission.

The wind data from 18th September 1989 and 20th October 1989 showed unusually large Doppler shifts in the North and North-East directions. The Doppler shifts in the other directions were much smaller. Normally the maximum wind speed is of the order of 50 m/s, but in these cases the maximum wind was about 300 m/s. On examining the raw fringes it became obvious that the apparently large Doppler shifts were caused by the appearance of an extra emission peak. The contaminating peak was very broad and encroached into the 8430Å peak thus distorting the peak and causing an erroneous peak position to be found.

This particular contamination also occurred on the 18th September and the 15th October. Figure 2.12 illustrates the form of the contamination on the 18th September in a series of consecutive fringes collected between 02:26UT to 04:04UT. The contamination was extremely strong to the north and north-east throughout the period shown. There was a slight contamination of the peak observed in the north-west and east. There was a radical difference in the intensities observed round the viewing circle in that the intensities to the north and north-east were around 3 times the magnitude of the intensities in the other directions. The localisation of the contamination and the increased intensity of the hydroxyl emission to the northern part of the sky implied an auroral source. This was confirmed by the Kp indices for the time period shown which ranged between 7- and 8° due to a severe magnetic storm. The 27 day interval between
the storms implicated a localised energy source on the sun as the cause, because the period of rotation of the sun is 27 days.

The contaminant has been identified as the 8446Å emission of OI4 from the upper thermosphere. The contamination by the 8446Å emission was most apparent during extremely high geomagnetic activity, for example both the nights of the 18th September 1989 and the 20th October 1989 achieved maximum $K_p$ values of around 8. Unfortunately the 8446Å peak was too close to the 8430Å peak to be easily distinguished, otherwise the observations could have been used to do a simultaneous measure of the thermosphere under active conditions! At present the peak fitting routine is programmed to reject apparently contaminated data where the peak position is more than a certain distance from the baseline under the assumption that winds greater than 100 m/s do not occur normally in the upper mesosphere. Future plans involve a more sophisticated routine that can separate the two peaks and possibly find the correct peak positions for each. The underlying presence of the 8446Å emission at all times, as well as during magnetic storms, has not yet been considered, and the consequent distortion of the 8430Å peak is not easily quantifiable. However, this complication will be removed when the laser is installed and a narrow filter can then be used.

There are also other, unidentified, blips that occur at random and at irregular positions on the detector. There is not enough memory on the PCs to keep full images of the data and so it is not possible to tell whether the blips are localised in space, with the source perhaps being lightning bolts, or a smear caused by stars scudding across the field of view during the 240 second integration period (the field of view of the FPI is 1° and so within this integration period a star could travel right across it).

A major problem caused by contamination is that an anomalous rapid change in the wind speed and direction could be interpreted as being evidence of a gravity wave passing through the field of view. The peak fitting routine is unable to weed out all contaminated peaks because the contaminating emissions can sometimes distort the 8430Å peak without being obviously distinguishable. However, if a surge in the wind has been identified it has been imperative to check carefully that there has been no contamination of the original peak before attributing the surge to a gravity wave. Such a check can only be done by plotting every peak and looking for suspicious asymmetries in the shape of the peak. This procedure is expensive in time and graphic materials and therefore
only appropriate for individual case studies.

2.4.5 The Effect of Cloud on the Doppler Shifts

Early evening and late morning were frequently cloudy at Kiruna during the field trip in November 1988. The presence of cloud causes scattering of the light emitted so that Doppler shifts from any direction are scattered and mixed together with the Doppler shifts from all other directions. As a result the light collected by the FPI has lost its directional information and is an average of all the Doppler shifts in the sky. Consequently, if there is uniform emission over the whole sky and a uniform wind field, and the light is scattered isotropically, then the observed Doppler shift in all directions averages out to zero.

However, when aurora is present there are definite Doppler shifts apparent even during cloudy weather. Kiruna is at the outer edge of the auroral oval which means that aurora generally appears only in the northern part of the sky. Thus there is no longer a uniform illumination and the averaging out process due to cloud scatter will show a Doppler shift dominated by the Doppler shift in the North. This average Doppler shift will be observed in all directions giving the appearance of a convergent or divergent wind field. The mathematics of the effect of a large intensity gradient over the field of view combined with cloud cover is dealt with in two papers by Abreu [1983, 1985].

An example of a night which starts off with total cloud cover and ends up clear comes from Kiruna on the 29th November 1986, shown in Figure 2.13a. From the beginning of the observing period, just after 14UT, until about 19UT there are no Doppler shifts, and consequently zero winds, observed in any direction. After this time significant winds appear in all directions. The all-sky camera indicates cloud cover until 17UT.

An example of the effects when cloud covers only part of the sky is again taken from Kiruna, on the 13th November 1986, shown in Figure 2.13b. The Kiruna all-sky camera shows cloud for most of the night yet the FPI observed Doppler shifts in all directions up to around 21UT. After this time the peak positions observed to the West, East, North-East and Zenith converge to show a zero Doppler shift position but observations to the North and North-West still show large shifts. Two possible reasons exist for this apparently non-uniform wind field:

(i) There is a gradient in the wind field which is real, with a large wind in the North dominating before 21UT.
(ii) Cloud cover is limited to the southern sky so that there is a resultant zero Doppler shift due to scatter only in that part of the sky.

Corroboration from a minute investigation of the Kiruna all-sky camera film, or data from the EISCAT radar, which is not affected by cloud scatter, would resolve this problem. However, such an investigation can only be done for special cases since it is such a labour intensive procedure and there is such a large outpouring of data each winter from several instruments.

2.4.6 The Effect of an Error in the Zenith Angle
The FPI makes observations at a zenith angle of 60°. It is assumed that the neutral wind is horizontal so that the wind can be calculated from the line-of-sight wind component $V$ using $V/\sin(60)$. Therefore an error of 1° in the zenith angle would give an error of 1% in the horizontal wind. The derived winds for all the observing directions would consequently suffer from a systematic increase or decrease in magnitude, but there would be no gradient in the wind field introduced by this error.
2.4.7 The Effect of an Error in the Orientation of the FPI

The FPI is set up and orientated as carefully as possible, but it would be instructive to know what the effect of a tilt by an angle $\alpha$ from the vertical would have on the calculation of the winds. For example, if we consider a simple case where the table is tilted along the East-West axis so that to the East the actual zenith angle for the observation is $(\theta + \alpha)$ while to the West it is $(\theta - \alpha)$, then the calculation of the zonal winds to the east, $V_e$, and to the west, $V_w$, from the line-of-sight wind, $V$, gives values according to equations 2.20 and 2.21.

$$V_e = \frac{V \sin(\theta + \alpha)}{\sin \theta} \tag{2.20}$$

$$V_w = \frac{V \sin(\theta - \alpha)}{\sin \theta} \tag{2.21}$$

It is unlikely that the FPI could be tilted by more than about $5^\circ$ without it being noticed. Thus, on calculating the extreme case using a value of $\alpha = 5^\circ$ it is found that $V_e$ and $V_w$ then differ by 10%. Therefore a tilt in the instrument axis away from vertical will cause a systematic error through the whole data set by introducing a false gradient of up to 10% in the wind field.
2.4.7.1 What can be Deduced about the Instrument Inclination from the Data Collected so Far?

The radius of the observation site is very large, at around 400 km. It is therefore to be expected that there may be gradients due to both large and small scale structure in the thermosphere as well as to temporal changes which are beyond the resolution of the ten minute scan time of the FPI. However, it is possible, to a certain extent, to distinguish actual structure in the thermospheric wind field from the systematic errors introduced by the inclination of the FPI by looking at long term averages of the Kiruna data:

i) There is no consistent difference between the observations to the North and South or to the East and West for all the different averages. This shows that even if there is a tilt to the instrument it is not large enough to cause a significant systematic gradient.

ii) With the zonal wind components the gradient changes from positive to negative around midnight which is consistent with the site passing from the dusk sector to the dawn sector of the ion convection pattern and thus moving from the influence of a clockwise to an anticlockwise circulation pattern.

iii) The vertical winds will provide a good indication of the orientation of the instrument. It is assumed that the vertical winds average out to zero over a long period of time. This assumption is generally valid both theoretically and from the observations of other groups and is used together with a calibration lamp data to determine a baseline for the 6300Å Doppler shifts.

The average separation of the zenith peaks and calibration lamp peaks is added to the calibration lamp peaks to establish the baseline for each night. The neon calibration lamp provides a constant wavelength monitor of the instrumental behaviour. The calibration lamp peaks are totally independent of the orientation of the instrument, so that if the instrument orientation is correct the winds observed in the zenith position will vary around 0 m/s. However, if there is a significant tilt to the axis the observations in the zenith would contain a contribution from the horizontal wind component. This would introduce a bias in the zenith peak positions which would then affect the baseline derived from it. A constant wind gradient would then falsely appear from the subsequent analysis of the data. However, any horizontal component intruding into the vertical observations would reveal its presence since the zenith peaks would
subsequently vary in their peak positions in unison with the horizontal winds, but to a lesser extent depending on how much the FPI is tilted from the vertical. Such a variation in the zenith peak positions has not been apparent in the data, for example there were very large Doppler shifts in the azimuth directions from the data collected at Kiruna on the night of the 30th September 1988, yet the zenith peaks remained steady.

It should be noted that large variations in the data peak positions can occur for low emission intensities, but this is due to the errors involved in fitting poor quality data. Calibration lamp peaks maintain a very steady peak position because the fringe profiles have a high signal to noise ratio and can therefore be fitted more accurately than the data peaks, which are less well defined. The zenith peak positions have a poorer signal to noise ratio because the line-of-sight observation cuts a shallower path through the emission layer than the observations of the azimuth directions. Thus the zenith peak positions are the most affected by low intensities, when they can be very variable and inconsistent.

iv) The field of view of the FPI is one degree. If the secondary mirror used to reflect light from the zenith into the FPI is independent of the main body of the instrument and is positioned correctly then any large tilt in the axis would throw the zenith observation out of view. Therefore observations can only be made from a badly aligned instrument if the secondary mirror is not independent but is adjusted to accommodate the primary mirror.

2.4.8 Converging and Diverging Wind Fields

Generally the divergence of the wind field is close to zero. There are often significant gradients in the observed wind vectors but few circumstances of large divergence. Abreu et al. [1983] warned that a large difference in the intensity of emission over the field of view can create the appearance of convergence or divergence by light from the high intensity direction being scattered by the lower atmosphere to arrive at the point of observation as though from another direction. Abreu et al. point out that scattering contamination causes a broadening of the line profile. But the increase in the deduced temperatures are only of the order of a few tens of degrees. True convergence or divergence is accompanied by temperature changes of a few hundreds of degrees, appropriate to the field of view lying on a boundary between two flow patterns of different origins. Comparison of flow patterns with adjacent observation sites may also serve to indicate the reality of the apparent divergence.
2.4.9 Large Vertical Winds

Thermospheric winds here are calculated using the assumption that any vertical wind component is insignificant compared with the horizontal component, and that any systematic upflow or downflow that might occur is not long-lived and is spatially localised. However, Fuller-Rowell [1987] has shown theoretically that the whole region of the auroral oval correlates closely with significant vertical winds of up to 50 m/s, which are the result of a combination of diverging horizontal wind flow and a change in the height of the pressure surface. Vertical winds that are greater than 20-30 m/s in magnitude have sometimes shown up in the data for the early evening or early morning hours. At these times of the day large vertical winds are more likely to be the fault of a poor signal to noise ratio due to a high background count from the evening or morning sun rather than any interesting atmospheric phenomenon.

It is important to know what effect a sizable vertical wind component would have on the calculations of the horizontal winds. For simplicity only a two-dimensional analysis will be presented here using the zonal component as an example rather than dealing with the whole three-dimensional wind vector. Thus, an FPI looking at a horizontal zonal wind component, \( V_h \), and vertical wind component, \( V_z \), from a zenith angle of \( \theta \), where \( \theta = 60^\circ \), will see a line-of-sight wind of \( U_e \), given by equation 2.22 to the East and \( U_w \) to the West (equation 2.23). (Reference should be made to Figure 2.15.)

\[
U_e = V_z \cos \theta + V_h \sin \theta \quad (2.22)
\]
\[
U_w = V_z \cos \theta - V_h \sin \theta \quad (2.23)
\]

In the standard method of processing the data we assume that the zenith wind component is zero and therefore calculate the horizontal zonal component to be:

to the East:
\[ V_e = \frac{U_e}{\sin \theta} \]
to the West:
\[ V_w = \frac{U_w}{\sin \theta} \]

Yet the actual horizontal component should be \( V_h \) in both directions, but the presence of a vertical wind results in an apparent gradient in the wind field which is given by equation 2.24:

78
\[ V_e + V_w = \frac{(V_z \cos \theta + V_h \sin \theta) + (V_z \cos \theta - V_h \sin \theta)}{\sin \theta} = \frac{2V_z}{\tan \theta} \] (2.24)

If, for instance, \( V_z = 50 \text{ m/s} \) the apparent gradient is \( 58 \text{ m/s} \). For an average horizontal wind of \( V_h = 200 \text{ m/s} \) this represents an alteration of 15\% to the real horizontal wind speed. Tristatic measurements are necessary to provide the information to measure a wind vector completely. Without such a facility it is necessary to neglect the vertical wind component in order to calculate the overall wind field pattern throughout the night with the minimum assumptions.
Figure 2.1

Schematic diagram of the Fabry-Perot Interferometer [Lloyd, 1985]
Figure 2.2

Fabry-Perot fringes [Born and Wolf, 1987]

Figure 2.3

Airy function [Born and Wolf, 1987]
Figure 2.4
An example of the reduced to radius-squared image of the calibration lamp obtained from the FPI at Kilpisjärvi.

Figure 2.5

BIN NUMBER, RADII-SQUARED

INTEGRITY

Direction: CAL LAMP
Integration time = 600 seconds
Time: 14 27 22 on 1/15/90

KILPISJÄRVI 6300A
Figure 2.6
Figure 2.7

85
Figure 2.8
Figure 2.9
Figure 2.10
Emission produced by the neon lamp used to calibrate the S430A data.

Figure 2.11

BIN NUMBER, RADIAL SQUARED

INTENSITY

Direction: NORTH
Integration time = 240 seconds
Time: 5:31:55 on 16/9/89

UTAH 8430A
Figure 2.13
Figure 2.14

Figure demonstrating the error produced during the calculation of the horizontal zonal wind when the instrument vertical axis is tilted at an angle $\alpha$ from the vertical. $\theta$ is the zenith angle for the line-of-sight observation.

Figure 2.15

Figure demonstrating the error produced by a significant vertical wind component, $V_z$, during the calculation of the horizontal zonal wind, $V_e$ which assumes a zero vertical wind. $V_h$ is the actual horizontal zonal wind and $\theta$ is the zenith angle for the line-of-sight observation.
3.1 Introduction

There is a wealth of investigation into the seasonal and solar cycle variation of the thermosphere. Empirical observations collected from sites all over the world include those of Jacchia and Slowey [1973], Babcock and Evans [1979], Wickwar et al. [1984] and Biondi et al. [1990]. Semi-empirical and theoretical models include Roble et al. [1977], Hedin [1984] and Fuller-Rowell et al. [1988]. Consequently the main trends in the behaviour of the thermosphere are thought to be well known and quite well understood. However, databases from past analyses of the thermosphere using FPI and radar results have been fairly small, often using fewer than 50 sample periods of observations. This has served as a statistical limit to the number of categories into which a database can be divided. Thus seasonal analyses of the thermosphere have generally concentrated on the contrast between summer and winter behaviour alone. This has been thought to be adequate since the two solstitial seasons are expected to represent the two extremes of the seasonal variation, in tune with the annual variation of the solar heating of each hemisphere of the earth. As a consequence there has been much less investigation into the behaviour of the thermosphere during the equinox seasons. Those papers that deal with seasonal variations have usually contained an implicit assumption that the spring and autumn thermosphere are fundamentally the same. For example in papers such as Roble et al. [1977] for model results and Alcaydé et al. [1974] for empirical results, the behaviour of autumn alone, or a conglomeration of the two, is presented as sufficient representation of both equinoxes. Even work on semi-annual variations do not point out the asymmetries that are often apparent [e.g. Roosen, 1966, Holzer and Slavin, 1982, de la Beaujardière et al., 1991]. In most cases this is because the database is small and so any asymmetry is probably ignored as data noise.

Thermospheric wind data have been collected systematically every winter for over nine years from a site at Kiruna, Sweden (67.8°N, 20.4°E). During moderate levels of geomagnetic activity the site is situated at the outermost edge of the auroral oval. The contraction and expansion of the auroral oval with
geomagnetic activity allows observations of a wide range of thermospheric behaviour, from winds typical of mid-latitude sites to those of polar cap sites. The database contains 1242 nights of data, which makes it possibly the largest database of thermospheric winds available from a single site. The database is large enough to divide it into three seasons: autumn, winter and spring; and two levels of solar activity, without losing statistical reliability.

A long-term, large-scale averaging of data as presented here should smooth over all the more minor influences on the neutrals to leave only the results of the competition between solar heating and ionospheric convection. The results show a marked seasonal and solar cycle variation. In particular, the investigation of the equinoctial behaviour of high-latitude thermospheric winds and a comparison with winter solstice winds has shown up some extremely interesting results that are quite at variance with the gross behaviour predicted by models such as Roble et al. [1977]. Of especial interest is the large asymmetry in the high-latitude thermospheric winds at spring and autumn equinox, which has not been predicted by models.

3.2 Description of the Database

The complete database of thermospheric winds observed at Kiruna ranges from November 1981 to April 1990, covering the period between mid-September and mid-April for each year. The year was divided into quarters for a seasonal analysis of the data. Any data from the period between $\pm 45$ days of the 21st September, inclusive, were classified as autumn winds, $\pm 45$ days of the 21st December as winter winds and $\pm 45$ days of the 21st March as spring winds. By such a division the centroid of each season would be as close as possible to the times of autumn and spring equinox and winter solstice. Unfortunately, the period of observation being from mid-September to Mid-April did not allow the autumn and spring databases to be centred on the 21st September and 21st March, which are the dates of the autumn and spring equinoxes. On using the complete database of 1242 nights the average dates for the autumn and spring databases were found to occur on the 9th October and 4th March. This makes the average winds at both the equinoxes more 'winter-like' rather than typical of equinoctial behaviour. However, since both sets of data are evenly disposed about the winter solstice on the 21st December, it is reasonable to make direct comparisons between them.

In the next phase of this study, the database was divided in terms of both
season and solar activity. Figure 1.3 shows the variation of $F_{10.7}$ solar flux from 1981 to 1992. It can be seen that the FPI database covers a large part of the eleven year solar cycle. A solar flux value of $F_{10.7} = 110$ was chosen to divide the database into two similar sized portions: data collected in the period between September 1983 and April 1988 represents solar minimum while data collected before and after this period represents solar maximum. The two portions of data were then subdivided into autumn, winter and spring, as described above.

Overcast, snowy or rainy periods identified by the Kiruna all-sky camera have been removed due to the problem of light scattering by cloud, as described in Chapter 2. This amounted to about 60% of the data collected being rejected. It should be noted that there is an initial quality control on the data. Data are rejected when the signal to noise ratio of the observed emission peaks is poor and so the database is already selective for good observing conditions before periods of bad weather are removed. Therefore the 60% of data rejected owing to bad weather is not representative of the actual percentage of bad weather.

The period during which data are collected is limited by the maximum photon count allowed to enter the FPI without harming the IPD, consequently observations are made only during the nighttime hours. There is an almanac within the program used to collect the data which maximises the length of time that the FPI is running. The almanac is consulted for the time when the sun is 8° below the horizon at dusk and dawn when the FPI is switched on and off respectively. As time progresses from autumn into winter and the nights lengthen, the FPI is switched on at earlier and earlier times and switched off at increasingly later times. After the winter solstice the nights get shorter once more and the observing period is accordingly curtailed. Thus data is guaranteed for the core period between about 18UT and 03UT for all three seasons, with a maximum observing period for the average winds in winter of seventeen hours, between 14UT and 07UT. The average wind speeds at the beginning and end of the observing periods are calculated from the smallest number of samples, which come from the longest nights for that season. Thus the average winds for the first and last hour or two are biased by a few days from the overall average date for each season towards mid-winter. Although the average date of the winter winds in the first and last hour of the observation period are no more than 5 days away from the 21st December, the average dates of the equinox winds in the first and last hour are biased by around 20 days towards mid-winter from the 4th March and 9th October. The importance of this bias depends on how rapidly
the seasonal variation occurs. However, the general analysis and interpretation of the data presented here will depend on the average winds nearer to the centre of the observation periods.

In general the FPI at Kiruna would be started up in mid-September by the staff at the Institute of Space Physics, who are familiar with the general maintenance of the instrument. Yet in some years the FPI was not started until November, when the UCL team arrived, due to problems with the instrument or when a new instrument needed to be installed. Then at the close of each observing campaign, the FPI is switched off in mid-April when the sun is below the horizon for too short a time for collection of a useful quantity of data. The winter period, therefore, has the largest database because it is the most likely of the three seasons to contain a full complement of data for every year, with three full months of observation, and with each night covering the longest span of time.

As explained in Chapter 2 the orientation of the instrument at Kiruna has been changed twice during the period of this database. In order to combine data from these different orientations it was necessary to convert everything to a set coordinate system. It was decided to use geographic coordinates since seasonal variations are dependent on the inclination of the geographic pole with respect to the sun. Thus all references to line-of-sight winds observed to the north and south, or to the east and west apply to true geographic meridional or zonal winds.
3.3 Seasonal Variation in Thermospheric Winds

The seasonally averaged wind components for autumn, winter and spring are displayed in Figures 3.1, 3.2 and 3.3 respectively. These plots display the horizontal wind components in terms of geographic coordinates for each of the seven observing directions. The data is plotted using the format of south and east being the positive directions for the observations from the four cardinal directions. As for the other directions, down, north-east and south-east are positive for the observations from the zenith, north-east and north-west respectively. The vertical lines on each data point are not error bars, but mark the standard deviation of the observed wind speeds. The errors on the individual wind speeds are, on average, about 20 m/s, yet the standard deviation of the averaged winds can be over 100 m/s. The large size of the standard deviations show that even sorting the winds with respect to both season and solar cycle fails to account for the full behaviour of the high-latitude thermosphere.

Table 3.1 displays some average values of parameters associated with the three categories of data to aid in judging the quality of the data. The spread in the average values of the \( F_{10.7} \) and \( K_p \) indices allow for much overlap between the data categories and thus for fair direct comparisons between the seasons. The \( K_p \) averages were obtained by converting each \( K_p \) value into the nearest equivalent \( A_p \) value using Table 3.3 [Bartels, 1932], then calculating the average \( A_p \) before converting back to the nearest \( K_p \) equivalent. This procedure was mathematically necessary because the \( K_p \) index is a semi-logarithmic scale, which therefore cannot be averaged directly, while the \( A_p \) index, being linear, can.

First the seasonal behaviour for each season is described and then comparisons are made between them:

a) autumn

The meridional wind observed to the north and south of Kiruna remains equatorward throughout the night, reaching a maximum value of 109 m/s between 01UT and 02UT. The meridional wind reduces to zero, indicating a change to a poleward flow, at the very beginning and end of the observing period, at around 17UT and 04:15UT. This is very different from the prediction by Kohl and King [1967] derived from a simple model of the thermospheric wind driven purely by pressure gradients, as described in Chapter 1. In their model, which does not include the high-latitude ionospheric convection, the high-
latitude meridional wind for nighttime conditions has a diurnal variation with maximum equatorward flow at just after 01UT and maximum poleward flow 12 hours later. Thus the meridional wind turns from northward to southward at 19UT and back again at 07UT. The average speed is 140 m/s and there is little change in the wind vector magnitude throughout the night. Although the time when the meridional wind has its maximum southward component agrees with the observed data, the times of direction change and the length of time for an equatorward flow, as well as the difference in the magnitude of the winds, show how significantly ionospheric convection modifies the winds.

The zonal wind component sweeps down from an initial speed eastwards of around 50 m/s, changes direction at 16:30UT and increases to a maximum westward value of 77 m/s to the east of Kiruna and 59 m/s to the west at around 17UT. After 16:30UT the zonal wind is strongly westward, indicative of the neutral flow being entrained into the dusk cell of the auroral oval. To the east of Kiruna the zonal wind remains westward until 20:15UT, but to the west of Kiruna it remains westward for an hour and a half more: until 21:45UT.

By midnight a large gradient in magnitude has appeared in the zonal winds. To the east the zonal wind reaches a maximum value of 106 m/s just after midnight while the average zonal wind observed to the west peaks at 50 m/s. In the midnight sector pure solar heating would cause equatorward winds. The large size of the zonal component indicates that in this case ion drag draws the neutral atmosphere into the eastward circulation of the dawn cell. The zonal wind to the east then drops in magnitude and at 03:30UT becomes westward, while to the west it changes direction reluctantly at 03:45UT, changing back and forth for over half an hour.

b) winter

The average winds observed during the winter months of November to January follow similar general trends to those of autumn, differing in degree only. The nights are longer and the period of observation ranges from 14UT to 07UT. After 20:30UT the zonal wind components have a standard deviation that averages out at around ±70 m/s. This is 25% smaller than the standard deviations for the equinoctial average winds at equivalent times. The smaller standard deviations may result from a much larger number of data samples in the winter database compared with the number in the databases for the two equinoxes, as can be seen in Table 3.1. Though it is also possible that the average winter
winds are genuinely less variable than the winds for the two equinoxes in the period after Kiruna has passed through the Harang discontinuity, because before 20:30UT the standard deviation of the winter winds is about 30% larger than the standard deviations of the equinox winds before this time.

For all three seasons the vertical winds remain very low throughout the night averaging out at around 0 m/s, but with a standard deviation of about ±40 m/s. In comparison the horizontal winds are much more variable, with an average standard deviation ranging from ±70 m/s to 100 m/s. However, it is interesting to note that the average vertical winds in winter appear to be more variable after midnight when the standard deviation jumps to around ±70 m/s. Such a change in the standard deviation is not evident with the other two seasons. There also appears to be a small but systematic change from a slightly downward zenith wind before 18UT, to a slightly upward wind afterwards. Then at 05UT the zenith wind suddenly becomes downwards again. This pattern is exhibited in the equinoctial vertical winds also, although the time when the vertical winds change direction is later, at around 21UT. The procedure for analysing the data relies on the assumption that the average zenith wind over the whole of the night is zero, from which a zero Doppler shift baseline is created (discussed in Chapter 2). It is possible that this assumption is only valid over a whole 24 hour period, during quiet to moderate geomagnetic conditions, when there are no significant thermal upwellings. But until there is strong evidence that the average nighttime vertical winds are anything but zero it is appropriate to use zero when analysing the data. In the meantime, the average vertical winds displayed in Figure 3.1, 3.2 and 3.3 may indicate a diurnal variation with a small amplitude which should be investigated.

The maximum speed of the winter meridional wind is 100 m/s equatorwards. There is little difference in the magnitude of the meridional wind observed to the north and the south. However, the meridional wind observed to the north of Kiruna changes direction at 16:45UT and 05:15UT, while to the south it changes direction at 17UT and 05:30UT. This is an apparent skewing of the wind field over the field of view so that the meridional component south of Kiruna lags the north by 15 minutes, though they both have an equatorward component for 12 hours and 30 minutes.

c) spring

The most noteworthy aspect of the average thermospheric winds in spring
is their size. The meridional wind observed to the north of Kiruna attains a maximum southward speed of 164 m/s just after 01UT, which is 50% larger than in autumn. The meridional wind observed to the north of Kiruna turns southward at 17UT and returns to northward at 05:30UT.

As with the other two seasons, the large standard deviations reveal the evening zonal winds to be the most variable, and the large gradient in the winds observed to the east and west of Kiruna is also apparent. The zonal winds at 18UT have a westward speed of 85 m/s both to the east and west of Kiruna. However, the zonal wind is observed to turn eastward at 20:15UT to the east of Kiruna while to the west it turns at 21:30UT. After these times the maximum eastward zonal wind speed is 50 m/s to the west of Kiruna, and 91 m/s to the east. The zonal winds observed to the east and west subsequently drop in size and briefly turn westward at 03:30UT for a couple of hours, rising to a maximum of 45 m/s, before returning to an eastward direction at 05:30UT and 05:45UT respectively.

Table 3.2 summarises the average meridional and zonal winds observed to the north and west of Kiruna at four separate times: 18UT, 21UT, 00UT and 03UT. The values given here are the weighted mean of the average winds for five 15 minute intervals, centred on each time, where the average winds are weighted by their sample size. This gives a smoothed picture of the wind variations since it incorporates how the winds change over the one and a quarter hour period, and so will allow a simpler comparison of the seasonal differences. The north and west observations are used to illustrate the smoothed seasonal variations because they have the largest databases. The databases of observations to the south and east contain only 70% of the number of observations in the other directions because the FPI does not observe to the south for 4 days on either side of a full moon. Since the winter of 1982-1983 the FPI at Kiruna has been oriented so that instrument North points at 12° west of geographic north. The process of interpolating the data along the observed line-of-sight directions to a geographic coordinate system has used data from the two observed directions on either side of the required direction, provided that they are within 90° bearing from the required direction. As a result the absence of observations to the instrument south has meant that both geographic south and geographic east have a reduced database.

At 18UT the autumn and spring meridional wind components are of a
similar magnitude, with spring larger by 15%, but both are at least 30% larger than the winter meridional wind. By 21UT the winter meridional wind has become larger than the autumn, until around midnight, when there is little difference in their magnitudes. Around 01UT the autumn wind surges above the winter values, although this is not shown in these smoothed values, and then drops so that at 03UT the two are similar in size again. By comparison the spring meridional wind rises massively through the night so that at midnight it is up to 68% larger than the autumn and winter meridional wind components. However, the average vectors, calculated from the meridional and zonal wind components, do not differ quite as much. The spring wind vector at midnight is found to be only 63% larger than the autumn vector and 44% larger than the winter vector.

There is a seasonal difference in the length of time that the meridional winds remain flowing in a southward direction, and the times of turning from north to south and back again. Although the meridional wind remains southward for the same length of time in winter and spring (12 hours 30 minutes), in winter the meridional winds to the north of Kiruna turn southward 15 minutes sooner than in spring. In contrast the autumn meridional winds to the north of Kiruna turn southward at the same time as the spring meridional winds, but remain southward for only 11 hours 15 minutes.

Comparing the zonal winds it was found that the most dramatic show of the competition between solar heating and ion drag is in the dusk cell of the auroral oval. Between 14UT and 18UT, for all three seasons, the standard deviation of the zonal winds is up to three times higher than average. This is due to the large spread in the possible values of the early evening zonal winds, when Kiruna is in the dusk sector of the ionospheric convection pattern. The FPI at Kiruna would see eastwards winds in the early evening if the force due to the pressure gradients is larger than that due to ion drag, but if ion drag is dominant then the winds would be westwards. In the region of the Harang discontinuity solar heating and ion drag both drive the neutral gas in a similar direction and so their separate effects are not so easily distinguishable.

Ion drag in the dawn cell is not so effective as in the dusk cell due to the ‘inertial resonance effect’ described in Chapter 1 [Fuller-Rowell and Rees, 1984]. Therefore by around 20UT the variability of the zonal winds reduces to around ±100 m/s. Although the standard deviation of the winter zonal winds then remains at this magnitude for the rest of the night, the zonal winds of
the two equinox periods show an increase in the standard deviation around midnight, which then falls off by dawn. In contrast, the nighttime behaviour of the meridional winds rarely varies from a general anti-sunward flow and yet the standard deviation is still quite large at around ±70 m/s.

The smoothed average zonal wind observed to the west of Kiruna at 18UT is similar in size for autumn and winter at almost 68 m/s, but 31% more westward in spring. Although this westward/sunward flow at this time reveals that ion drag is stronger than the pressure gradients for all three seasons, the strength of the ion drag would appear to be less in winter because it is contending with pressure gradients that are smaller due to a reduced period of daylight heating, with the sun lower on the horizon.

By midnight the zonal wind magnitude is greatest in winter, while spring and autumn are similar in size and 30% smaller than winter. The vectors indicate that the winter winds are the most strongly entrained into the eastward flow of the dawn cell with a bearing of 147°, followed by the autumn with a bearing of 158° and finally the spring winds. The spring wind vector at midnight has a bearing of 166° and so is predominantly southward and therefore still antisunward.

After 03UT, as can be seen in Figures 3.1, 3.2 and 3.3, the equinox wind is slightly westward for a short period. In contrast the winter wind is strongly westward for the last 3 – 4 hours before the end of the observation period. The trend would imply a return to an eastward flow at around 07:30UT compared with about 05:30UT for the equinox winds. The westward, anti-sunward wind indicates a weakening of the ion drag so that the driving force of solar heating becomes dominant. This is compatible with Kiruna moving outside the regime of the dawn cell.

Thus, in general, the autumn thermospheric winds above Kiruna appear to be closer in behaviour to the winter winds rather than to spring. The spring winds exhibit the largest magnitudes, especially in the midnight sector. A seasonal variation in thermal forcing might explain the increased size of the spring winds over those of winter, when the period of sunlight is much shorter, but the autumn winds do not comply with this explanation. When considering the effect of ionospheric convection it might be expected that the average driving force due to ion drag in winter might be higher due to the increased electron densities known as the Winter Anomaly [Yonezawa, 1959, Croom et al., 1960]
Farmer et al., 1990], but this does not seem to be the case.

Table 3.1

AVERAGE PARAMETERS ASSOCIATED WITH SEASONALLY AVERAGED THERMOSPHERIC WINDS

<table>
<thead>
<tr>
<th>season</th>
<th>average date</th>
<th>$N$ per 15min. $F_{10.7}$</th>
<th>$K_P$</th>
</tr>
</thead>
<tbody>
<tr>
<td>autumn</td>
<td>9th Oct</td>
<td>40±21</td>
<td>138±46</td>
</tr>
<tr>
<td>winter</td>
<td>21st Dec</td>
<td>115±42</td>
<td>136±48</td>
</tr>
<tr>
<td>spring</td>
<td>4th Mar</td>
<td>83±43</td>
<td>144±49</td>
</tr>
</tbody>
</table>

Table 3.2

SEASONALLY AVERAGED THERMOSPHERIC WINDS

averaged meridional winds observed to the north of Kiruna
(m/s, positive South)

<table>
<thead>
<tr>
<th>season</th>
<th>18:00</th>
<th>21:00</th>
<th>00:00</th>
<th>03:00</th>
<th>UT</th>
</tr>
</thead>
<tbody>
<tr>
<td>autumn</td>
<td>35.5</td>
<td>78.8</td>
<td>93.1</td>
<td>50.5</td>
<td></td>
</tr>
<tr>
<td>winter</td>
<td>20.3</td>
<td>84.6</td>
<td>95.0</td>
<td>52.5</td>
<td></td>
</tr>
<tr>
<td>spring</td>
<td>42.2</td>
<td>133.2</td>
<td>158.0</td>
<td>93.2</td>
<td></td>
</tr>
</tbody>
</table>

averaged zonal winds observed to the west of Kiruna
(m/s, positive East)

<table>
<thead>
<tr>
<th>season</th>
<th>18:00</th>
<th>21:00</th>
<th>00:00</th>
<th>03:00</th>
<th>UT</th>
</tr>
</thead>
<tbody>
<tr>
<td>autumn</td>
<td>-67.4</td>
<td>-27.8</td>
<td>37.2</td>
<td>13.9</td>
<td></td>
</tr>
<tr>
<td>winter</td>
<td>-67.6</td>
<td>-22.8</td>
<td>61.1</td>
<td>30.3</td>
<td></td>
</tr>
<tr>
<td>spring</td>
<td>-88.1</td>
<td>-18.7</td>
<td>38.5</td>
<td>14.9</td>
<td></td>
</tr>
</tbody>
</table>
Table 3.3

Conversion of $K_p$ into $Ap$ indices

[Bartels, 1932]

<table>
<thead>
<tr>
<th>$K_p$</th>
<th>$0^\circ$</th>
<th>$0^+\hspace{1cm}$</th>
<th>$1^-\hspace{1cm}$</th>
<th>$1^0\hspace{1cm}$</th>
<th>$1^+\hspace{1cm}$</th>
<th>$2^-\hspace{1cm}$</th>
<th>$2^0\hspace{1cm}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Ap$</td>
<td>0</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>$K_p$</td>
<td>$2^+\hspace{1cm}$</td>
<td>$3^-\hspace{1cm}$</td>
<td>$3^0\hspace{1cm}$</td>
<td>$3^+\hspace{1cm}$</td>
<td>$4^-\hspace{1cm}$</td>
<td>$4^0\hspace{1cm}$</td>
<td>$4^+\hspace{1cm}$</td>
</tr>
<tr>
<td>$Ap$</td>
<td>9</td>
<td>12</td>
<td>15</td>
<td>18</td>
<td>22</td>
<td>27</td>
<td>32</td>
</tr>
<tr>
<td>$K_p$</td>
<td>$5^-\hspace{1cm}$</td>
<td>$5^0\hspace{1cm}$</td>
<td>$5^+\hspace{1cm}$</td>
<td>$6^-\hspace{1cm}$</td>
<td>$6^0\hspace{1cm}$</td>
<td>$6^+\hspace{1cm}$</td>
<td>$7^-\hspace{1cm}$</td>
</tr>
<tr>
<td>$Ap$</td>
<td>39</td>
<td>48</td>
<td>56</td>
<td>67</td>
<td>80</td>
<td>94</td>
<td>111</td>
</tr>
<tr>
<td>$K_p$</td>
<td>$7^0\hspace{1cm}$</td>
<td>$7^+\hspace{1cm}$</td>
<td>$7^-\hspace{1cm}$</td>
<td>$7^0\hspace{1cm}$</td>
<td>$8^+\hspace{1cm}$</td>
<td>$9^-\hspace{1cm}$</td>
<td>$9^0\hspace{1cm}$</td>
</tr>
<tr>
<td>$Ap$</td>
<td>132</td>
<td>154</td>
<td>179</td>
<td>207</td>
<td>236</td>
<td>300</td>
<td>400</td>
</tr>
</tbody>
</table>
3.4 Seasonal and Solar Cycle Variation in Thermospheric Winds

On separating the data further in terms of solar activity it was found that the seasons did not maintain the same order in terms of the response of the neutral winds to ionospheric and thermal forcing. Figure 3.4 contains plots of the average meridional winds observed to the north of Kiruna at solar maximum for autumn, winter and spring, and the zonal winds observed to the west of Kiruna. Figure 3.5 contains the equivalent plots for solar minimum. Table 3.4 summarizes the averages of parameters associated with each season for each level of solar activity. There is little difference in the average values of solar flux for the three seasons at solar maximum, nor is there much difference at solar minimum. The levels of geomagnetic activity are moderate in all cases. Tables 3.5 and 3.6 summarize the average wind speeds by giving the smoothed meridional and zonal winds for solar maximum and solar minimum, respectively, averaged over one hour and 15 minutes in the same procedure as for Table 3.2 for the whole database.

Comparing each season for a solar cycle dependence shows that increased solar flux correlates with an increase in the wind vector magnitude e.g around midnight the increase is up to 70%.

For both solar maximum and solar minimum the spring meridional winds are much larger in the midnight sector than the winds for the other two seasons. At solar maximum the spring meridional wind at midnight is 68% larger than the autumn wind and 52% larger than the winter wind. At solar minimum the spring meridional wind at midnight is only 37% larger than the autumn wind though still 52% larger than the winter wind.

The period of time that the meridional winds remain southward has a solar cycle dependence as well as a seasonal dependence. For example, a comparison of the winter meridional winds to the north of Kiruna shows that the meridional wind is southward for 1 hour 15 minutes longer at solar minimum than at solar maximum. At solar minimum the winter meridional wind turns southward at 16:15UT and returns northward at 05:15UT, while at solar maximum the winter meridional wind turns southward at 17UT and returns northward at 04:45UT.

The zonal winds before 21UT are more westward, and therefore sunward, at solar maximum than at solar minimum, which indicates increased ion drag in the dusk cell of the ionospheric convection pattern. This is true for all three seasons, though to different extents: the largest solar cycle dependent difference in the evening zonal winds occurs in the spring and the smallest in the autumn,
as can be seen by comparing the speeds in Tables 3.5 and 3.6. Similarly the zonal winds between 21UT and 03UT are more eastward at solar maximum than at solar minimum, indicating the increased strength of the dawn cell in diverting the neutral winds. However, in this period the autumn zonal winds show the largest increase at solar maximum and the spring zonal winds the smallest.

It is also observed that the evening zonal winds turn eastward at an earlier time with increased solar activity. The biggest difference occurs in the spring zonal wind which at solar maximum turns westward nearly 2 hours before it does so at solar minimum. The zonal winds then turn westward in the early hours of the morning at a later time at solar maximum than at solar minimum. This means that overall the nighttime zonal winds remain flowing eastward for a longer time at solar maximum than at solar minimum. Again the biggest solar cycle difference occurs in the spring zonal winds which turn westward an hour and a quarter later at solar maximum than at solar minimum. Thus at solar minimum the nighttime zonal wind is eastward for 6 hours in autumn, 5 hours 30 minutes in winter and 4 hours in spring, but at solar maximum it is eastward for 5 hours 30 minutes in autumn, 6 hours 30 minutes in winter and 7 hours in spring. The apparent deviation of the autumn winds from this general solar cycle dependence is caused by the zonal winds observed to the west of Kiruna at solar minimum. These winds are generally small, indicating a weak dawn cell, and in fact changes direction briefly to become westward 4 hours after first turning eastwards, but then quickly turn eastwards again. However, when the autumn zonal winds observed to the east of Kiruna are compared it is found that increased solar activity does indeed increase the length of time that the zonal wind remains eastward: from 5 hours 15 minutes at solar minimum to 7 hours at solar maximum, and therefore the general observation holds true.

The relative seasonal differences at solar maximum and solar minimum are also of interest. On a gross level the season which has the most westward zonal wind before 21UT has the least eastward zonal wind between 22UT and 03UT, and vice versa. The seasons which have the most and least westward evening zonal wind are different at solar maximum and solar minimum. At solar maximum the spring evening zonal wind is the most westward and the autumn zonal wind is the least, while at solar minimum the autumn evening zonal wind is marginally more westward than the spring, while the winter zonal wind is the least westward by about 50%. As a result, the average bearings of the winds at 18UT show that for solar minimum the autumn zonal winds
have a bearing of 244° (calculated from Table 3.6) and are therefore the most entrained into the dusk convection cell, followed by spring (bearing = 237°) and lastly winter (bearing = 232°). In contrast, at solar maximum it is the winter winds that are the most entrained into the dusk cell, with a bearing of 259°, followed by spring (bearing = 248°), while the autumn winds are the least entrained, with a bearing of 240°.

The bearings of the neutral winds at 00UT, when Kiruna is in the vicinity of the dawn cell, contrast with the results for the dusk cell. At solar minimum the winter winds are the most entrained into the eastward ionospheric convection (bearing = 146°), followed by spring (bearing = 166°) and then autumn (bearing = 171°). At solar maximum the autumn winds are the most entrained (bearing = 148°), then winter (bearing = 149°) and finally spring (bearing = 166°). The difference between the dawn and dusk cell is pronounced.
### Table 3.4

**AVERAGE PARAMETERS ASSOCIATED WITH SEASONALLY AVERAGED THERMOSPHERIC WINDS FOR SOLAR MAXIMUM $F_{10.7} > 110$**

<table>
<thead>
<tr>
<th>season</th>
<th>average date</th>
<th>$N$ per 15min. interval</th>
<th>$F_{10.7}$ $Kp$</th>
</tr>
</thead>
<tbody>
<tr>
<td>autumn</td>
<td>10th Oct</td>
<td>20±6</td>
<td>183±16 3°</td>
</tr>
<tr>
<td>winter</td>
<td>23rd Dec</td>
<td>57±18</td>
<td>179±22 4−</td>
</tr>
<tr>
<td>spring</td>
<td>5th Mar</td>
<td>51±27</td>
<td>178±26 4°</td>
</tr>
</tbody>
</table>

### AVERAGE PARAMETERS ASSOCIATED WITH SEASONALLY AVERAGED THERMOSPHERIC WINDS FOR SOLAR MINIMUM $F_{10.7} < 110$

<table>
<thead>
<tr>
<th>season</th>
<th>average date</th>
<th>$N$ per 15min. interval</th>
<th>$F_{10.7}$ $Kp$</th>
</tr>
</thead>
<tbody>
<tr>
<td>autumn</td>
<td>8th Oct</td>
<td>22±13</td>
<td>97±19 4−</td>
</tr>
<tr>
<td>winter</td>
<td>18th Dec</td>
<td>59±24</td>
<td>92±19 3+</td>
</tr>
<tr>
<td>spring</td>
<td>3rd Mar</td>
<td>34±16</td>
<td>89±17 3°</td>
</tr>
</tbody>
</table>
Table 3.5

SEASONALLY AVERAGED THERMOSPHERIC WINDS
FOR SOLAR MAXIMUM $F_{10.7} > 110$

<table>
<thead>
<tr>
<th>season</th>
<th>18:00</th>
<th>21:00</th>
<th>00:00</th>
<th>03:00</th>
<th>UT</th>
</tr>
</thead>
<tbody>
<tr>
<td>autumn</td>
<td>42.1</td>
<td>101.8</td>
<td>112.0</td>
<td>46.9</td>
<td></td>
</tr>
<tr>
<td>winter</td>
<td>24.5</td>
<td>99.0</td>
<td>120.8</td>
<td>52.4</td>
<td></td>
</tr>
<tr>
<td>spring</td>
<td>48.8</td>
<td>154.2</td>
<td>183.3</td>
<td>104.7</td>
<td></td>
</tr>
</tbody>
</table>

averaged meridional winds observed to the north of Kiruna

<table>
<thead>
<tr>
<th>season</th>
<th>18:00</th>
<th>21:00</th>
<th>00:00</th>
<th>03:00</th>
<th>UT</th>
</tr>
</thead>
<tbody>
<tr>
<td>autumn</td>
<td>-73.8</td>
<td>-25.0</td>
<td>71.2</td>
<td>14.3</td>
<td></td>
</tr>
<tr>
<td>winter</td>
<td>-120.7</td>
<td>-25.5</td>
<td>74.3</td>
<td>50.3</td>
<td></td>
</tr>
<tr>
<td>spring</td>
<td>-119.1</td>
<td>-10.0</td>
<td>44.4</td>
<td>27.3</td>
<td></td>
</tr>
</tbody>
</table>

averaged zonal winds observed to the west of Kiruna

Table 3.6

SEASONALLY AVERAGED THERMOSPHERIC WINDS
FOR SOLAR MINIMUM $F_{10.7} < 110$

<table>
<thead>
<tr>
<th>season</th>
<th>18:00</th>
<th>21:00</th>
<th>00:00</th>
<th>03:00</th>
<th>UT</th>
</tr>
</thead>
<tbody>
<tr>
<td>autumn</td>
<td>29.5</td>
<td>65.3</td>
<td>79.2</td>
<td>56.9</td>
<td></td>
</tr>
<tr>
<td>winter</td>
<td>16.4</td>
<td>72.5</td>
<td>74.7</td>
<td>52.7</td>
<td></td>
</tr>
<tr>
<td>spring</td>
<td>34.5</td>
<td>103.9</td>
<td>113.7</td>
<td>74.9</td>
<td></td>
</tr>
</tbody>
</table>

averaged meridional winds observed to the north of Kiruna

<table>
<thead>
<tr>
<th>season</th>
<th>18:00</th>
<th>21:00</th>
<th>00:00</th>
<th>03:00</th>
<th>UT</th>
</tr>
</thead>
<tbody>
<tr>
<td>autumn</td>
<td>-61.6</td>
<td>-29.6</td>
<td>12.1</td>
<td>13.2</td>
<td></td>
</tr>
<tr>
<td>winter</td>
<td>-21.2</td>
<td>-20.5</td>
<td>51.0</td>
<td>10.1</td>
<td></td>
</tr>
<tr>
<td>spring</td>
<td>-52.5</td>
<td>-31.2</td>
<td>27.7</td>
<td>-7.0</td>
<td></td>
</tr>
</tbody>
</table>
3.5 Height of the 6300Å Emission Peak

The peak intensity of the 6300Å airglow emission has been found to lie at an altitude of between 200 km and 300 km by models based on the observations from sounding rockets and satellites Parkinson and Zipf, 1971, Hays et al., 1978, Sharp et al., 1979, Rees and Roble, 1986]. This is described in more detail in Chapter 1. The intensity peaks at just below the F-region electron density peak [Herrero and Meriwether, 1980].

It is known that the height of the F-region peak varies with latitude, and also with season and solar cycle, as is illustrated in Figure 3.6 [Wright, 1962]. There will be a similar latitudinal, seasonal and solar cycle variation in the peak emission height of the 6300Å emission peak since dissociative recombination of O$_2^+$, which is dependent on the electron density, is the major source of nighttime O($^3D$) atoms [Bates, 1946]. However, this variation in height is not considered and experimenters using FPI observations of the 6300Å emission have usually assumed that the peak emission height is around 240 km [Lloyd, 1985, Burnside and Teply, 1989], as is the assumption used here.

It was realised that it might be possible to calculate the height of the 6300Å emission peak from the effect of a prominent feature of the auroral oval on the averaged thermospheric winds observed at Kiruna. This feature is the region separating the dusk and dawn cells of the ionospheric convection pattern, known as the Harang discontinuity [Harang, 1946, Foster et al., 1981, Vickrey et al., 1981]. As the night progresses, the Earth rotates under the auroral oval, which is fixed with respect to the geomagnetic pole, and orientated by the sun. The site at Kiruna thus moves from the region of westward ion flow in the dusk cell through to the dawn cell where the ion flow is eastward. Through ion-neutral collisions momentum is transferred to the neutral gas, causing it to flow with the ions. Therefore, as Kiruna passes under the Harang discontinuity the thermospheric zonal winds will be seen to turn from a westward flow to an eastward flow. On average Kiruna passes under the Harang discontinuity at around 21UT.

If it is assumed that the height of the peak intensity of the 6300Å emission is 240 km, and the zenith angle of the line-of-sight observation is 60°, then the radius of the field of view of the FPI is 416 km. Therefore, at the latitude of Kiruna, the time taken for the earth to rotate so that the portion of the thermosphere observed to the east of Kiruna becomes the portion observed to the west of Kiruna is 1 hour 20 minutes. The average autumn thermospheric winds, from the whole dataset, show that to the east of Kiruna the zonal wind
changes from westward to eastward at 20:15UT, while to the west this change occurs an hour and a half later at 21:45UT. Therefore, the delay in the time of observation of the Harang discontinuity between the east and west of Kiruna is compatible with the time taken for the earth to rotate into position.

Using this argument it should be possible to use the time lag to calculate the distance between the two points observed to the east and west of Kiruna, and therefore the height of the 6300Å emission peak. From the results of this chapter, the seasonal and solar cycle variation of the 6300Å emission peak may be investigated. Table 3.7 shows the results of this analysis.

At solar maximum the lag between the times of reversal of the zonal winds predict heights for the 6300Å emission peak that are within the expected range of 200 – 300 km. However, the most surprising outcome of these calculations is that the 6300Å emission peak appears to be 150 – 200 km higher at solar minimum than at solar maximum. This is not a credible result since the F-region peak increases in height with increasing solar activity, and the 6300Å peak emission height should increase with it.

The accuracy in the measurement of the time lag is not in doubt because, although there is a large standard deviation in the average wind speeds, the time at which the zonal wind speeds reverse is well defined for all the divisions of the database. The standard deviations merely indicate that there is a range of a few hours over which the zonal wind can change from a westward to eastward direction, and therefore there will be a significant spread in the calculated emission heights.

A possible flaw in this line of argument may be because the auroral oval is highly dynamic, and the location of the Harang discontinuity will accordingly vary considerably. In fact, the Harang discontinuity is a feature of statistical models and may not appear in any distinct form on an individual night. But since it does appear in statistical averages of ionospheric convection, it might be expected to appear in the average behaviour of the thermospheric winds presented here.

One reason for the unlikely emission heights at solar minimum may be that when the ion densities are lower, the convection electric field has a reduced strength for diverting the neutral gas through ion-neutral collisions. The pressure gradients at solar minimum will also be reduced but the strengths of these two main driving forces may not vary proportionately from solar minimum to solar maximum. Thus it may be that at solar minimum the influence of the
pressure gradients on the neutral winds clouds the influence of ionospheric convection so that the Harang discontinuity cannot be clearly and indisputably identified by the reversal of the evening zonal winds.

The most obvious answer to this conundrum is as follows: both points observed to the east and west of Kiruna trace the same paths in terms of geographic latitude. However, the ionospheric convection pattern is fixed with respect to the geomagnetic pole which precesses around the geographic pole. On assuming that the peak emission is at a height of 240 km then the difference in geographic longitude of the points observed to the east and west of Kiruna is 6.8°, i.e. the geographic coordinates of the point to the east is approximately (67.8°N, 23.8°E), and to the west the geographic coordinates are (67.8°N, 17.0°E). However, in a geomagnetic coordinate system the coordinates are (64.3°N, 121.3°E) and (65.3°N, 115.7°E) respectively. Therefore the path traced through the ionospheric convection pattern by the point observed to the east of Kiruna is not the same as that traced by the point observed to the west of Kiruna. Consequently the time when each point of observation passes through the Harang discontinuity cannot be compared and used to determine the height of the peak emission. This explanation could also account for why the nightly variation of the horizontal thermospheric wind observed to the west is not a time-lagged copy of the winds observed to the east.

The conclusion of this little investigation is that the passage of the high-latitude thermosphere through features of the auroral oval, and the consequent effect on the winds, cannot be used to deduce the height of the 6300Å emission peak reliably. Yet the results have indicated that there is likely to be a significant variation in height with season and solar cycle which should be considered in interpreting the data, although a quantitative measure cannot yet be made.
Table 3.7

Estimation of the Height of the 6300Å Emission Peak from the Time Lag in the Observation of the Harang Discontinuity to the East and West of Kiruna

<table>
<thead>
<tr>
<th>Season</th>
<th>All Data</th>
<th>Solar Maximum</th>
<th>Solar Minimum</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>time lag</td>
<td>equivalent time lag</td>
<td>equivalent time lag</td>
</tr>
<tr>
<td></td>
<td>/hours</td>
<td>/km/hours</td>
<td>/km/hours</td>
</tr>
<tr>
<td>autumn</td>
<td>1.5</td>
<td>271</td>
<td>1.5</td>
</tr>
<tr>
<td>winter</td>
<td>1.25</td>
<td>225</td>
<td>1.0</td>
</tr>
<tr>
<td>spring</td>
<td>1.25</td>
<td>225</td>
<td>1.25</td>
</tr>
</tbody>
</table>
3.6 Seasonal and Solar Cycle Variation in Ion Velocities

The seasonal and solar cycle variation in the ion drag that is strongly apparent in the neutral winds was subsequently confirmed by the observations of ion velocities from the EISCAT radar at Tromsø, Norway (69.6°N, 19.2°E). Figures 3.7 and 3.8 show the seasonal variation of the ion velocities for solar maximum and solar minimum respectively [Farmer and Jarvis, 1991]. The four quarters of the year are centred on the equinoxes and solstices ±45 days, as with the Kiruna thermospheric winds. The average plasma velocities were calculated from tristatic observations at a height of 280 km, using the Common Program runs CP-1 and CP-2 during the years 1984 to 1990. Data from 1984 to 1987 represent solar minimum and data from 1988 to 1990 represent solar maximum. Average velocities for each season and level of solar activity were calculated for each hour from between 70 and 500 data points. Each run is capable of contributing around 17 data points per hour. There is a considerable standard deviation on each individual velocity vector of a similar order of magnitude to the standard deviation of the average neutral winds.

The general behaviour of the plasma velocities is the same for all four seasons and both levels of solar activity: from around 20UT to 08UT the velocities are predominantly eastwards and from around 08UT to 20UT they are westwards. The meridional component is very small compared with the size of the zonal component. Tables 3.8 and 3.9 show the average values of the ion velocities over one hour for solar maximum and solar minimum, respectively, for comparison with Tables 3.5 and 3.6.

The effect of increased solar flux is to increase the magnitude of the ion velocities by a large factor which is dependent on the season. This introduces both a seasonal and solar cycle variation in the size of the overall average ion velocity. At solar maximum the largest ion velocities overall undoubtedly occur during the spring, followed by summer, autumn and finally winter. However, at solar minimum both autumn and winter have similarly large overall velocities, followed by summer and finally spring. Consequently there is a considerable equinoctial asymmetry so that at solar maximum the spring ion velocities are up to three times larger than in autumn, while at solar minimum the spring ion velocities are very small and several times smaller than the autumn ion velocities. This is in contrast with the average neutral winds where the wind velocities around midnight are larger in spring than in autumn at both levels of solar activity. It is presumed that the difference between the seasonal variation
of the ion velocities at solar maximum and solar minimum relates to the extent of the auroral oval: at solar maximum Tromsø is within the vicinity of the auroral oval but at solar minimum the site is outside, in the region where increased ion densities result in smaller neutral winds, as is the case at mid- and low-latitude sites [Kohl and King, 1967].

There does not seem to be any simple correlation between the bearings of the neutral winds and the ion velocities at 18UT and 00UT, when Kiruna is in the region of the dusk and dawn convection cells respectively. The correlation between the bearings of the ion velocities and neutral winds is not improved by making comparisons with the average ion velocities at 17UT and 23UT in order to allow for the 1-2 hours delay in the response of the neutrals to changes in the ion velocities. At 17UT the most strongly westward flow i.e. bearing closest to 270°, at solar minimum occurs during spring (bearing=279°), followed by winter (bearing=253°) and then autumn (bearing=223°), while at solar maximum the most westward flow occurs during autumn (bearing=257°), followed by spring (bearing=251°), and finally winter (bearing=244°). It appears that the ion velocities become less westward and more southward with increased solar flux in the dusk cell.

At 23UT the most eastward ion velocities (i.e. bearing closest to 90°) for solar minimum occur during spring (bearing=87°), then autumn (bearing=98°) and winter (bearing=146°). For solar maximum the most eastward velocities occur during autumn (bearing=83°), then spring (bearing=104°) and winter (bearing=133°). The two spans in the values of the average bearings at solar minimum and solar maximum overlap, and the bearings for the individual seasons do not change in any consistent direction, so there does not appear to be any simple solar cycle influence in the dawn cell as with the dusk cell.
### Table 3.8

**AVERAGE ION VELOCITIES FOR SOLAR MAXIMUM 1988-1990**

<table>
<thead>
<tr>
<th>Season</th>
<th>18:00</th>
<th>00:00</th>
<th>06:00</th>
<th>12:00</th>
<th>UT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Autumn</td>
<td>60.4</td>
<td>-1.9</td>
<td>-30.5</td>
<td>-7.3</td>
<td></td>
</tr>
<tr>
<td>Winter</td>
<td>59.7</td>
<td>54.3</td>
<td>27.2</td>
<td>-15.7</td>
<td></td>
</tr>
<tr>
<td>Spring</td>
<td>81.8</td>
<td>0.4</td>
<td>-48.9</td>
<td>-20.4</td>
<td></td>
</tr>
<tr>
<td>Summer</td>
<td>80.0</td>
<td>-15.8</td>
<td>-63.5</td>
<td>-35.5</td>
<td></td>
</tr>
</tbody>
</table>

**average meridional ion velocities**

( m/s, positive South )

<table>
<thead>
<tr>
<th>Season</th>
<th>18:00</th>
<th>00:00</th>
<th>06:00</th>
<th>12:00</th>
<th>UT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Autumn</td>
<td>-127.</td>
<td>127.</td>
<td>19.0</td>
<td>-72.8</td>
<td></td>
</tr>
<tr>
<td>Winter</td>
<td>-71.8</td>
<td>71.4</td>
<td>-36.2</td>
<td>-139.</td>
<td></td>
</tr>
<tr>
<td>Spring</td>
<td>-55.5</td>
<td>269.</td>
<td>72.0</td>
<td>-177.</td>
<td></td>
</tr>
<tr>
<td>Summer</td>
<td>-87.4</td>
<td>151.</td>
<td>56.9</td>
<td>-184.</td>
<td></td>
</tr>
</tbody>
</table>

**average zonal ion velocities**

( m/s, positive East )

### Table 3.9

**AVERAGE ION VELOCITIES FOR SOLAR MINIMUM 1984-1987**

<table>
<thead>
<tr>
<th>Season</th>
<th>18:00</th>
<th>00:00</th>
<th>06:00</th>
<th>12:00</th>
<th>UT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Autumn</td>
<td>66.1</td>
<td>13.9</td>
<td>-53.3</td>
<td>-12.4</td>
<td></td>
</tr>
<tr>
<td>Winter</td>
<td>78.6</td>
<td>-4.6</td>
<td>-90.5</td>
<td>-21.0</td>
<td></td>
</tr>
<tr>
<td>Spring</td>
<td>-0.1</td>
<td>19.9</td>
<td>5.7</td>
<td>-25.3</td>
<td></td>
</tr>
<tr>
<td>Summer</td>
<td>23.1</td>
<td>-2.2</td>
<td>-22.9</td>
<td>2.2</td>
<td></td>
</tr>
</tbody>
</table>

**average meridional ion velocities**

( m/s, positive South )

<table>
<thead>
<tr>
<th>Season</th>
<th>18:00</th>
<th>00:00</th>
<th>06:00</th>
<th>12:00</th>
<th>UT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Autumn</td>
<td>-115.</td>
<td>130.</td>
<td>44.9</td>
<td>-62.2</td>
<td></td>
</tr>
<tr>
<td>Winter</td>
<td>-165.</td>
<td>165.</td>
<td>7.8</td>
<td>-59.7</td>
<td></td>
</tr>
<tr>
<td>Spring</td>
<td>-92.7</td>
<td>35.5</td>
<td>7.8</td>
<td>-20.1</td>
<td></td>
</tr>
<tr>
<td>Summer</td>
<td>-94.6</td>
<td>83.8</td>
<td>20.9</td>
<td>-33.5</td>
<td></td>
</tr>
</tbody>
</table>

**average zonal ion velocities**

( m/s, positive East )
3.7 Considering the Effect of a Long-Term Phase Lag in the Response of the Thermosphere to Solar Heating

The two main driving forces on thermospheric winds are the pressure gradients and ion drag. Both forces are dependent on solar flux levels but ion drag is also dependent on the chemical composition of the thermosphere, which determines the plasma density. The seasonal variation in the solar flux is symmetric about the winter solstice when the flux levels are at a minimum value, which means that the levels of solar flux at the spring and autumn equinoxes are the same. Thus a possible reason for an asymmetry between the spring and autumn thermospheric winds may lie in a phase lag between the seasonal variation in the solar flux and the seasonal variation in the chemical composition of the thermosphere.

Figure 3.9 illustrates the variation in the position of the solar terminator throughout the year. A view onto the northern hemisphere at a latitude corresponding to Kiruna i.e. 68°N shows the position of the solar terminator for each month, underneath is the corresponding plot of the daily variation of the solar zenith angle, \( \chi \). The response of the thermosphere at a particular level to solar ionization is determined by the solar zenith angle and the optical thickness at that altitude. Although the Kiruna FPI database covers only the nighttime thermosphere, when there is no direct solar insolation, the pressure gradients set up by solar heating on the dayside of the Earth and the general levels of plasma density generated by photoionization during the day are important to the nightside behaviour of the thermosphere. It was therefore decided to look into the possible effects of the variation of the average flux over a full 24 hours.

At the altitude of the E-region the atmosphere is optically thick to the main ionizing wavelengths, while at the height of the F2-peak the thermosphere is optically thin \([\text{Rishbeth and Garriott}, 1969]\). Since the height of the maximum emission of 6300Å lies around 50 km below the F-region peak the optical depth is expected to be somewhere between the two extremes; for example, Abreu et al. [1983] used a value of 0.27 for the optical depth in their work on the effects of atmospheric scattering on ground-based measurements of thermospheric winds. Consequently two simple models, given in equations 3.1 and 3.2, have been worked out to calculate the daily variation of photoionization for two possible extremes of optical depth. These are illustrated in Figures 3.10a and 3.10b. Equation 3.1 and Figure 3.10a show the expected photoionization for an optically thin atmosphere and equation 3.2 and Figure 3.10b for an optically thick atmosphere; each shows the expected variation through 24 hours for summer
and winter solstice and for the equinoxes. The total daily photoionization is assumed to be proportional to the integral under the curve.

Equations 3.1 and 3.2 are given in terms of the solar zenith angle, $\chi$, for a site at latitude $\theta$ and at a time represented by its longitude $\Gamma$. $\Gamma_p$ and $\Gamma_q$ are the longitudes where $\chi = \pm 90^\circ$, effectively the times of sunset and sunrise respectively. $\gamma_o (= 23.45^\circ)$ is the angle of tilt of the Earth’s rotational axis to the plane of the ecliptic and $(\omega t)$ represents the frequency, $\omega$, and time, $t$, of the Earth’s annual orbit around the sun, where $t$ is measured from summer solstice on June 21st. A simple circular orbit is assumed for the Earth.

For an optically thin atmosphere the optical depth $\tau = 0$, which means that there is no attenuation of flux as it passes through the atmosphere. Thus for an optically thin atmosphere the ionization builds up at a maximum rate (i.e. the same rate as for an overhead sun) as soon as the sun appears over the geometric horizon. Equally, when the sun drops down below the horizon and the photoionization reduces to nothing, the ionization falls off rapidly. This is represented by a tophat function as in equation 3.1, illustrated in Figure 3.10a, which is actually a measure of the length of the daylight hours.

a) optically thin atmosphere $\tau = 0$

$$\text{photoionization} = \begin{cases} 1, & \text{for } -90^\circ < \chi < 90^\circ \\ 0, & \text{for } \chi > 90^\circ \text{ and } \chi < -90^\circ \end{cases}$$

therefore total daily photoionization $= \int_{\Gamma_q}^{\Gamma_p} d\Gamma$

$$= \Gamma_p - \Gamma_q \quad (3.1)$$

For an optically thick atmosphere ($\tau \gg 0$) the photoionization is proportional to the solar flux incident on unit area. This varies with the cosine of the zenith angle during the daytime, reaching a maximum at local noon. In this model it is assumed that at night there is no solar flux and hence zero photoionization, so the function chosen to represent an optically thick atmosphere is given by equation 3.2 and illustrated in Figure 3.10b.
b) optically thick atmosphere $\tau >> 0$

\[
\text{photoionization} = \begin{cases} 
\cos \chi, & \text{for } -90^\circ < \chi < 90^\circ \\
0, & \text{for } \chi > 90^\circ \text{ and } \chi < -90^\circ 
\end{cases}
\]

Therefore total daily photoionization = \int_{\Gamma_p}^{\Gamma_q} \cos \chi \, d\Gamma

\[
= 2 \cos \theta \sin \left( \frac{\Gamma_p - \Gamma_q}{2} \right) \left( l \cos \left( \frac{\Gamma_p + \Gamma_q}{2} \right) + m \sin \left( \frac{\Gamma_p + \Gamma_q}{2} \right) \right) + n \sin \theta \left( \Gamma_p - \Gamma_q \right)
\]

where: $l = \cos \gamma \cos(\omega t)$, $m = \sin(\omega t)$, $n = \sin \gamma \cos(\omega t)$.

These two functions do not take into account the scattering and refraction of light that illuminates the thermosphere for some time before dawn and after sunset. Since Kiruna is on the Arctic Circle there is no solar flux at winter solstice without atmospheric scatter. Scatter and refraction may be allowed for by modifying equations 3.1 and 3.2 by incorporating the Chapman grazing incidence function, but this would be a sophistication unnecessary for the level of this investigation. Instead an approximate allowance may be made for the appearance of a small quantity of solar flux during the winter solstice by using a value of $\theta$ for a latitude that is a few degrees lower than Kiruna.

Next, the dependence of ion drag on the chemical composition of the thermosphere was investigated. The strength of the ion drag, irrespective of whether it is accelerating or retarding the neutrals, depends on the relative densities of ions and neutrals. The plasma density in the thermosphere is a function of production, loss and transport. In the lower thermosphere a chemical balance between production and loss alone can be assumed. With increasing altitude the lifetime of the ions becomes longer and the transport of plasma, either by diffusion along the field line or perpendicular to it by an electric field, plays a more important role. This is certainly the case at the height of the $F_2$ peak, particularly in winter and at the equinox, when the electric field has a dominant influence on the resulting structure at high latitudes [Fuller-Rowell et al., 1988]. However, in order to quantify the seasonal change in ion drag in the upper thermosphere in a simple manner it is assumed that the thermosphere at the $6300\text{Å}$ airglow peak height is close to chemical equilibrium. Even if transport
is partially effective at this altitude the mean levels of ionization will still be representative of values predicted by chemical balance, since the electric field will, to first order, merely redistribute the plasma.

The O\(^+\) ion is the major ion constituent and is usually used to represent the mean plasma density. The main chemical reactions involved in the production and loss of O\(^+\) ions are as follows:

\[
\text{O} + \text{h}\nu \rightarrow \text{O}^+ + e^- \quad \text{(production)}
\]
\[
\text{O}^+ + \text{N}_2 \rightarrow \text{NO}^+ + \text{N} \quad \text{(loss)}
\]
\[
\text{O}^+ + \text{O}_2 \rightarrow \text{O}_2^+ + \text{O} \quad \text{(loss)}
\]

If chemical equilibrium is assumed then the overall loss rate of O\(^+\) through recombination is proportional to the solar flux and the ratios: [O]/[O\(_2\)] and [O]/[N\(_2\)]. Alcaydé et al. [1974] calculated long-term variations of thermospheric temperature and composition using a molecular weighted sum of [O\(_2\)] and [N\(_2\)], but for this analysis of the gross seasonal variation of thermospheric winds above Kiruna it is sufficient to assume that the recombination of O\(^+\) is dependent only on the variation of [N\(_2\)].

The 24-hour average densities of O and N\(_2\) were obtained for every month from the MSIS-86 thermospheric model [Hedin, 1987]. The model parameters were chosen to simulate the average conditions prevalent during the Kiruna observations, thus model data were obtained for a latitude of 68°N, \(K_p = 3°\) and solar flux levels of \(F_{10.7} = 179\) for solar maximum and \(F_{10.7} = 92\) for solar minimum. These values were extracted for a constant pressure level corresponding to pressure level 11 in the UCL-Sheffield 3DTD model which is closest to the height of the peak emission of 6300Å [Fuller-Rowell et al., 1991]. Monitoring the atmosphere on constant pressure level surfaces is more appropriate than at constant heights since the seasonal changes in temperature which affect the height of the peak emission are accommodated [Garriott and Rishbeth, 1963].

The global circulation of the atmosphere causes the ratio between molecular and atomic constituents to change with the seasons. The major ion constituent of the thermosphere, O\(^+\), depends on the ratio of [O]/[N\(_2\)], which is a maximum in winter. So, despite the solar flux, and consequently photoionization, being at a minimum level for the Northern hemisphere in winter, the chemical composition set up by the global circulation of the atmosphere causes plasma densities to
be at a maximum value [Rishbeth and Setty, 1961]. This is a long observed phenomenon, known as the Winter Anomaly [Yonezawa, 1959, Croom et al., 1960, Foster, 1984, Farmer et al., 1990].

Figure 3.11 shows the seasonal variation in [O]/[N₂] compared with the seasonal variation in the total daily flux incident on Kiruna. The seasonal variation in flux is symmetric about winter solstice, but the asymmetry in [O]/[N₂] is such that the ratios in late February and early October, corresponding to the average times of the spring and autumn neutral winds, are not equal: [O]/[N₂] in spring is larger than in autumn. So the buildup of [O]/[N₂] from the minimum value at summer solstice would appear to be faster than might be expected, and the decay, after the winter solstice maximum, is slower.

On further investigation it was found that at a constant pressure level the asymmetry in [O]/[N₂] is largely caused by the variation in [N₂]. Figure 3.12 shows the seasonal variation in [N₂] and [O] at solar maximum and solar minimum. There is a pronounced asymmetry in the seasonal variation of [N₂] but only a minor asymmetry in [O]. As an aside, it is interesting to note that the seasonal variation at a constant height reverses this pattern, giving a strongly asymmetric seasonal variation of [O] and almost symmetric seasonal variation of [N₂].

The combined effect of the seasonal variation in the solar flux and the seasonal variation in the ion drag due to chemical composition changes may now be estimated by looking at the variation of the product of solar flux and [O]/[N₂]. Figure 3.13a shows the variations of the product of solar flux and [O]/[N₂] calculated for an optically thin atmosphere under solar maximum conditions using equation 3.1, while Figure 3.13b shows the equivalent function calculated using equation 3.2 for an optically thick atmosphere. Both the optically thick and thin atmospheres produce asymmetric functions that are a minimum at winter solstice. On comparing the values at around the 9th October and 4th March (the average dates of the autumn and spring databases for the neutral winds) it is apparent that the autumn values are generally slightly lower than spring in both cases. This would be in tune with the larger wind speeds observed in spring. However, this is found to be a relatively insignificant difference to account for the large equinoctial asymmetry observed in the neutral winds when calibrated against the difference between the values on the 4th March and the 21st December. Table 3.2 shows that at both solar minimum and solar maximum the winter
meridional winds are similar in magnitude to the autumn winds, while for both levels of solar activity the spring meridional winds are about a third larger than those in autumn and winter at equivalent times during the night. Yet looking at the functions shown in Figures 3.13a and b, the difference between the product of solar flux and [O]/[N₂] in late December and early March for either of the optical thicknesses is not equalled by the difference between the values in early October and early March. In other words, calibrating the spring-autumn difference in the product of solar flux and [O]/[N₂] against the spring-winter difference shows that there is not a significant enough spring-autumn difference to justify the huge difference observed between the meridional winds over Kiruna.

It should be noted that the MSIS model used in this analysis is a harmonic analysis of data from satellites and radar. The database is dominated by data from mid- and low-latitude sites. On average the thermosphere varies smoothly at these latitudes and is therefore amenable to harmonic analysis. At high latitudes the structure of the electric and magnetic fields renders the thermosphere far more complex. This consideration, together with the relative lack of empirical data from these latitudes, causes a harmonic analysis to be a less precise method of modelling the high-latitude thermosphere. The value of the model here has been to provide a gross measure of the thermospheric behaviour and so it is not appropriate to undertake a close scrutiny of the results. It is sufficient to say that there does appear to be a phase difference in the seasonal variation of the chemical composition of the thermosphere with respect to the seasonal changes in solar insolation. However, this does not provide evidence to account for the similarity between the autumn and winter thermospheric winds and for the large asymmetry observed in the Kiruna between the equinoctial thermospheric winds.
3.8 Considering the Influence of Solar Wind-Magnetospheric Coupling

Since the quantity and levels of solar radiation are the same in spring and autumn, and the seasonal variation in the plasma density cannot produce a satisfactory asymmetry between the two equinoxes (although a small asymmetry does exist) it was decided to investigate the seasonal variation in plasma drift to see if there is an asymmetry in the momentum transferred to the neutrals from the ions. For momentum transfer depends not only on the number density of the ions but also their velocities, which at high-latitudes are predominantly determined by the convection electric field.

The combination of the results of the seasonal variation in thermospheric winds and in ion velocities, in particular the observation that ion drag appears to be stronger in spring and autumn than in winter, suggested an investigation into an alternative mechanism, connected with the semi-annual variation of geomagnetic activity, which consists of maximum geomagnetic activity in March and September and minimum activity in June and December.

3.8.1 The Semi-Annual Variation in Geomagnetic Activity

The semi-annual variation of geomagnetic activity was first recognised at the turn of the century by Cortie [1912], who documented a semi-annual variation in the number of geomagnetic storms measured over a 23 year period at one station. Bartels [1932] also showed a semi-annual variation in the geomagnetic index $U$, which is a measure of the component of the geomagnetic field parallel to the Earth’s magnetic axis. The many attempts to explain the semi-annual variation fall into two schools of thought, led by Cortie and Bartels, called the ‘axial’ and the ‘equinoctial’ hypotheses respectively, which are described briefly below:

The axial hypothesis, according to Cortie [1912], attributed the geomagnetic storms to solar windstreams emanating radially from or near sunspots. The first sunspots of a new cycle, which is defined as beginning at solar minimum, appear at heliographic latitudes $20^\circ - 25^\circ$ in both the northern and southern hemispheres. Sunspots then appear at lower and lower latitudes so that the mean concentration moves progressively towards the solar equator until the end of the cycle. This is illustrated in Figure 3.14 which is a ‘butterfly’ diagram of the distribution of sunspots throughout a solar cycle, observed at the Mount Wilson and Las Campanas Observatories [Handbook of Geophysics and the
Space Environment', 1985]. The average location of sunspots over a complete solar cycle is therefore at solar latitudes greater than about 7°. The plane of the ecliptic is inclined at 7.2° to the sun's equator so that the Earth is within range of the most active regions of the sun twice a year when it reaches its maximum heliographic latitudes. Peak geomagnetic activity is consequently expected to occur on the 5th March and 5th September when the Earth reaches the maximum south and north heliographic latitudes respectively.

In contrast, the equinoctial hypothesis of Bartels [1932] proposed that the inclination of the Earth's magnetic field to the solar wind determines how effectively solar particles are captured by the geomagnetic field. From his results he presumed that the most effective inclinations occur at the equinoxes so that peaks in geomagnetic activity were predicted to occur on the 21st March and 21st September.

At the time when these hypotheses were made there were no space observations of the Interplanetary Magnetic Field (IMF) and there were no theories about solar wind-magnetosphere coupling. Therefore no physical mechanisms could be invoked. The only way to decide which of the two hypotheses was correct was to determine accurately the times of the peaks in geomagnetic activity to compare with the predictions. However, there was a wide dispersion in the measured times of peak geomagnetic activity. For instance, in Figure 3.15 there is a marked semi-annual variation in a twenty year average of the Ap index, collected over the years 1937-1956 [McIntosh, 1959]. An earlier analysis made of the annual variation of the U index by Chapman and Bartels [1940] over 59 years (1872-1930) is illustrated in Figure 3.16. The data was collected over almost three solar cycles and separated into quiet, intermediate and active years. The semi-annual variation is well defined for all three divisions but the times of the maximum values of U are quite different. Not only that, but although both the McIntosh and the Chapman and Bartels data show an asymmetry between spring and autumn activity levels, while the Ap variation has spring activity larger than autumn, the variation in U is the other way around. Russell and McPherron [1973] suggested that the lack of agreement in the phases and magnitudes of these measured semi-annual variations are the consequence of a real variability in the solar wind caused by blast waves, shocks and high velocity streams, superimposed on a situation where the Earth is more capable of extracting energy from the solar wind during the equinoctial months than during the solstitial months. But it might be expected that such short term variability
would be smoothed out over such a large database.

Once the solar wind and the IMF had been measured by satellites it became apparent that there was a strong correlation between the orientation of the IMF and the geomagnetic activity on Earth. The component of the IMF which is parallel to the Earth's geomagnetic axis, \( B_z \), was found to be the most important of the three components with regard to the input of solar wind energy to the magnetosphere and ionosphere. There was observed to be a large amount of geomagnetic activity with much particle precipitation for an IMF with \( B_z < 0 \), but with \( B_z > 0 \) there was relatively little activity noted in the auroral stations [Araki et al., 1984, de la Beaujardière et al., 1987, Makita et al., 1988, Hruska and Hruska, 1989, Brautigam et al., 1991]. This was explained by Dungey [1961] who postulated that merging of the IMF with the magnetosphere occurred in a quasi-steady process over the whole front surface of the magnetosphere which faces the sun. The direction of the \( z \)-component determines whether the magnetosphere is 'open' or 'closed' to penetration by the IMF. Figures 3.17a and b depict the geomagnetic field in polar section together with the IMF for negative and positive values of \( B_z \) respectively. With positive \( B_z \), as shown in Figure 3.17a, the IMF has a \( z \)-component parallel to the geomagnetic field, which configuration is not conducive to merging, hence a 'closed' magnetosphere. Figure 3.17b shows that for negative \( B_z \) the IMF has a \( z \)-component that is anti-parallel to the geomagnetic field and that can therefore merge. This results in an 'open' magnetosphere. The 'open' magnetosphere created by a negative component (\( B_z < 0 \)) allows solar plasma to be injected into the magnetosphere at the poles along the magnetic field lines, via the plasma sheet of the magnetotail. This led to the 'solar rectifier' model of Arnoldy [1971], where the interaction between the solar wind and magnetosphere is zero for \( B_z > 0 \) and is linearly proportional to the magnitude of the field for \( B_z < 0 \).

Further satellite observations and other analyses of the interaction of the solar wind with the magnetosphere have indicated that the merging region is localised, unlike the Dungey model. There is a large body of opinion based on satellite observations of flux transfer events that merging of the IMF with the geomagnetic field is localised to a region near the subsolar point [e.g. Russell and Elphic, 1978, 1979, Berchem and Russell, 1984]. However, there is an alternative proposition that merging occurs near the polar cusps [Crooker, 1979, Murayama et al., 1980].

Murayama et al. found that the \( AL \) index is strongly dependent on the in-
clination of the Earth’s dipole to the ecliptic plane. Consequently they proposed that merging must in fact occur near the cusp because changes in the inclination would not affect the hydrodynamic and electromagnetic conditions of the magnetosphere otherwise. In another paper by the same authors, Hakamada et al. [1980] looked at the effect of the component along the line of the Earth’s orbit, $B_y$, on the $AL$ index. They could only reconcile the discrepancies that they found with the predictions of the classical theory of merging by assuming that the region of merging has an elongated shape. Geometrically this is compatible with merging near the poles rather than near the dayside nose of the magnetosphere where the magnetopause is nearly spherical.

If merging occurs near the cusp then the attitude of the polar cusp to the solar wind becomes important. This would then bring in a source of asymmetry that may explain the equinoctial asymmetry in the thermospheric winds, because all of the conditions controlling the spring and autumn thermosphere are equivalent except for the attitude of the Earth to the IMF. The underlying explanation of both the semi-annual variation and the equinoctial asymmetry of the thermospheric winds may be that not only is the sun-Earth inclination at equinox more favourable for the coupling between the geomagnetic field and the IMF, but that the spring orientation is superior to the autumn.
3.8.2 Seasonal Variation in the Orientation of the Earth to the IMF

There are two coordinate systems to be considered here: the first has orthogonal axes \(i_o, j_o, k_o\) defined by \(i_o\) which connects the centre of the Earth to the sun and \(k_o\) which is orthogonal to the plane of the ecliptic, pointing to the North ecliptic. This is commonly known as the Geocentric Solar-Ecliptic (GSE) coordinate system. However, most experimental work dealing with solar-magnetospheric interactions uses a coordinate system called the Geocentric Solar Magnetospheric (GSM) coordinate system with axes \(i_m, j_m, k_m\). The GSM system differs from the GSE system in that although the \(i_m\) axis always points towards the sun as does the \(i_o\) axis, the \(k_m\) axis is determined by the projection of the Earth's geomagnetic axis onto the plane orthogonal to the \(i_m\) axis. Thus, since the Earth's rotational axis is tilted at a constant angle of \(\gamma_o = 23.5^\circ\) to the plane of the ecliptic, there is a seasonal variation in the orientation of \(k_m\). There is also a diurnal variation where the \(k_m\) axis bobs back and forth about the \(i_m\) axis in order to match the diurnal rotation of the geomagnetic pole, and consequently the magnetosphere, about the geographic pole. These two coordinate systems are mathematically defined in the paper by Russell [1971].

Satellite observations of the IMF show a very structured field. Figure 3.18a shows the average IMF magnitude to be around 5 nT. Figures 3.18b and c show the frequency distributions of the direction of the IMF. As can be seen from the figures the IMF appears to lie within the solar equatorial plane but at a skew angle of 45° from the sun-Earth axis. This led to the Parker spiral or 'garden-hose effect' model of the IMF shown in Figure 3.19, which may be explained in the following manner: the arrows indicate the radial path of the solar wind, away from the sun, while the spiral arms split the equatorial plane into sectors having opposite magnetic polarities. Although the solar wind moves out radially, carrying with it the solar magnetic field, the rotating sun generates the field at successive solar longitudes, hence the spiral shape of the interplanetary field.
It is apparent that, to a zeroth order approximation, the orientation of the IMF vector $\mathbf{B}$ at the Earth in the GSE coordinate system is either $\mathbf{B} = (i_0 - j_0)$ or $\mathbf{B} = (-i_0 + j_0)$, which means that the average magnitude of the components $(B_{ox}, B_{oy}, B_{oz})$ are:

\[
\begin{align*}
\langle B_{ox} \rangle &= \pm \sqrt{\frac{1}{2}} |\mathbf{B}| \\
\langle B_{oy} \rangle &= \mp \sqrt{\frac{1}{2}} |\mathbf{B}| \\
\langle B_{oz} \rangle &= 0
\end{align*}
\] (3.3) (3.4) (3.5)

i.e. $B_{ox}B_{oy} = -|\mathbf{B}|$. The $x$-component determines the sign of the sector structure. The regions of the solar wind where $B_{ox} > 0$ are commonly referred to as ‘towards’ sectors and regions where $B_{ox} < 0$ are referred to as ‘away’ sectors, according to whether the IMF points towards or away from the sun. The number of sectors varies with the level of solar activity, averaging out at four. Figure 3.20 is a schematic illustration of the solar current sheet which separates interplanetary space into regions of magnetic fields which are pointing either ‘towards’ or ‘away’ from the sun. A cross-section through the equatorial plane of Figure 3.20 will give the Parker spiral of Figure 3.19.

$B_{mx}$ has not been found to have any significant effects on the magnetosphere at all, therefore this analysis concentrates only on the effect that $B_{my}$ and $B_{mx}$ components may have in creating an asymmetry in the spring and autumn equinoxes.

Though the components of the IMF vector $\mathbf{B}$ in the GSE coordinate system are, on average, independent of the time of year, the $B_{my}$ and $B_{mx}$ components in the GSM system vary seasonally due to the different aspects presented by the Earth’s geomagnetic axis to the IMF throughout the year. This is illustrated in Figure 3.19, which is a slice through the ecliptic plane showing the position of the Earth’s North geographic pole relative to the sun for the two solstices and two equinox seasons, as well as the average orientation of the IMF vector. There is also a diurnal variation in the components $B_{my}$ and $B_{mx}$ due to the rotation of the geomagnetic axis about the geographic axis. Russell and McPherron [1973] proposed that the semi-annual variation in geomagnetic activity could be attributed to the seasonal variation in the orientation of the Earth’s magnetosphere to the IMF. They were able to simulate the semi-annual variation by using the solar rectifier model to represent the average quantity of
solar wind-magnetospheric coupling. The coupling function was calculated in terms of a half-wave rectified average of the $z-$ component of the IMF in the GSM coordinate system, $< B'_{mz} >$, using the following criteria:

$$\begin{align*}
B'_{mz} &= \left\{ \begin{array}{ll}
B_{mz}, & \text{for } B_{mz} < 0 \\
0, & \text{for } B_{mz} \geq 0
\end{array} \right.
\end{align*}$$

(3.6)

Russell and McPherron used the approximation to a spiral for the IMF vector with an average vector magnitude of $5nT$ and an equal proportion of 'towards' and 'away' sectors. The value of the solar wind-magnetosphere coupling is proportional to $< B'_{mz} >$, which is the mean of $B'_{mz}$ for the 'towards' and 'away' sectors. The resulting contour plot of $< B'_{mz} >$, shown in Figure 3.21, reveals both a semiannual and diurnal variation in the solar wind-magnetosphere coupling. The coupling strength varies with how negative the value of $< B'_{mz} >$ is. The most negative value of $< B'_{mz} >$ occurs near the 5th April and the 5th October. In addition, the diurnal rotation of the geomagnetic pole about the geographic pole causes the most negative $< B'_{mz} >$ to occur near 22UT during spring, and near 10UT during autumn.

The predicted diurnal variation in $< B'_{mz} >$ may be understood by referring to Figures 3.22a and b, which look at the orientation of the Earth in autumn and spring in the $j_\odot, k_\odot$ plane. In this plane the distribution of the IMF lies mainly along the $j_\odot$ axis. At autumn equinox the rotational axis of the Earth is tilted in one direction at an angle $\gamma_\odot$ to the $k_\odot$ axis, while at spring equinox the Earth is tilted in the opposite direction by the same amount. Consequently the average IMF components at the equinoxes over a full 24 hours, $< B_{mz} >$, $< B_{my} >$, $< B_{mx} >$, are found to be:

**autumn equinox**:

$$\begin{align*}
< B_{mz} > &= \pm \sqrt{\frac{1}{2}} |B| \\
< B_{my} > &= \mp \sqrt{\frac{1}{2}} \cos \gamma_\odot |B| \\
< B_{mx} > &= \pm \sqrt{\frac{1}{2}} \sin \gamma_\odot |B|
\end{align*}$$

(3.7)
spring equinox:

\[
<B_{mz}> = \pm \sqrt{\frac{1}{2}} |B| \tag{3.8}
\]

\[
<B_{my}> = \mp \sqrt{\frac{1}{2}} \cos \gamma_0 |B|
\]

\[
<B_{mz}> = \mp \sqrt{\frac{1}{2}} \sin \gamma_0 |B|
\]

It is apparent from Figures 3.22a and b that in autumn it is the projection of the ‘towards’ sectors onto the \(k_m\) axis which produces \(B_{mz} < 0\), while in spring it is the projection of the ‘away’ sectors. The diurnal rotation of the geomagnetic pole about the geographic pole causes the \(k_m\) axis to oscillate back and forth in the \(j_\alpha, k_\alpha\) plane with an amplitude of about 9°, which is the separation in latitude of the geomagnetic pole from the geographic pole. Since the geographic longitude of the geomagnetic pole is about 78°W, in autumn the magnetic dipole will lie along \(k_m\) at 10UT when the axis is tilted at its largest angle from the \(k_\alpha\) axis, and consequently the projection of \(B\) onto \(k_m\) will be a maximum. However, at 10UT in spring the \(k_m\) axis will be tilted at its smallest angle from the \(k_\alpha\) axis and consequently the projection of \(B\) onto \(k_m\) will be at its minimum value. In the same manner, at 22UT the projection of \(B\) onto \(k_m\) will be at its maximum value for spring and its minimum value for autumn. Thus, although there is no difference in \(<B'_{mz}>\) between spring and autumn over a full 24 hours, there is a difference over a 12 hour period.

Figure 3.23 shows the actual seasonal and diurnal variation in the average value of \(<B'_{mz}>\) calculated from 24 years of IMF data collected by satellite between 1963 and 1986 and stored in the World Data Centre at the Rutherford-Appleton Laboratory [Wild, private communication, 1991]. In Figure 3.24 the data are divided into six 4-hour periods, beginning with the period 00UT \(<\) time \(<\) 04UT. The seasonal and diurnal variation in \(<B'_{mz}>\) appears as predicted by the Russell and McPherron model, although the real data differs from the model in several details:

(i) There is indeed a semi-annual variation in \(<B'_{mz}>\), but the value of \(<B'_{mz}>\) is around 7% more negative in spring than in autumn.

(ii) There is a double minimum in spring centred on February and April, with the February minimum slightly lower than the April minimum. The Russell and McPherron [1973] model predicts that the minimum should occur on the 5th April.
(iii) With respect to the diurnal variation of $< B'_{mz} >$, the most negative value of $< B'_{mz} >$ occurs in the period $20\text{UT} \leq \text{time} < 24\text{UT}$ for spring, as predicted by Russell and McPherron [1973], but for autumn the most negative value of $< B'_{mz} >$ occurs over a wider time period, appearing in the two periods $08\text{UT} \leq \text{time} < 12\text{UT}$, and $12\text{UT} \leq \text{time} < 16\text{UT}$.

(iv) In the diurnal variation the most negative value of $< B'_{mz} >$ is around 25% more negative in spring than in autumn.

It should be noted that the solar rectifier model is a very good first approximation but has been superseded by more sophisticated theoretical models of the dependence of the coupling strength on the geometry of the IMF and the geomagnetic field. The coupling functions proposed are generally proportional to $\sin^N(\theta/2)$, where $N$ is an integer between 0 and 4, while $\theta = \arctan(B_{my}/B_{mz})$ [e.g. Gonzalez and Mozer, 1974, Akasofu and Ahn, 1980, Reiff and Luhmann, 1986, Fedder et al., 1991]. All these functions show the large amount of coupling that occurs for $B_{mz} < 0$, but of these the best approximations to experimental observations maintain a small amount of coupling for weakly positive $B_{mz}$. Each of the models produce a semiannual and diurnal variation of solar wind-magnetospheric coupling, though with differing amplitudes. The function with $N = 2$ has a negligible amplitude. The functions with $N = 3$ or $N = 4$ predict maximum coupling at $22\text{UT}$ for spring and $10\text{UT}$ for autumn, consistent with the solar rectifier model, but the function with $N = 1$ is 12 hours out of phase. Although Fedder et al. [1991] have chosen this last function: $\sin(\theta/2)$, as the best fit to their numerical calculation of the reconnection voltage, which they tentatively link to the cross polar cap voltage, the phase is not compatible with the argument proposed in the following subsection as an explanation of the large equinoctial asymmetry observed in the thermospheric winds above Kiruna.
3.8.3 Possible Explanation of the Equinoctial Asymmetry Observed in High-Latitude Thermospheric Winds

The consistent asymmetry between the spring and autumn meridional winds in the midnight sector of the auroral oval can now be easily explained in terms of the diurnal variation of the half-wave rectified average of the $z-$component of the IMF, $\langle B'_{mz} \rangle$. As outlined in section 3.7 there is a strong correlation between the magnitude of the southward component of the IMF and geomagnetic activity due to the efficiency of the coupling between the IMF and geomagnetic field. With increased coupling the Cross Polar Cap Potential (CPCP) increases. Fedder et al. [1991] have calculated values of the CPCP from simulations using a self-consistent time-dependent model of the solar wind-magnetosphere-ionosphere system using five different values of the IMF angle $\theta$. They propose that the best fit to their solutions is a $\sin(\theta/2)$ curve. Figure 3.25 plots the five normalised values of the CPCP and the curves for $\sin(\theta/2)$ and $\sin^2(\theta/2)$ for reference.

The consequence of a seasonal and diurnal variation in $\langle B'_{mz} \rangle$ is to introduce a seasonal and diurnal variation in the ionospheric convection pattern. Since the electric field intensifies within a few minutes of a decrease in $B_{mz}$ [de la Beaujardière et al., 1987, 1988, Etemadi et al., 1988] the diurnal variation in $\langle B'_{mz} \rangle$ will be followed very closely by the ionospheric convection pattern. Figure 3.26 plots four levels of ionospheric convection patterns according to Foster et al. [1986a]. The levels plotted are 1, 4, 7 and 9, which correspond to an average $Kp$ of $1^-$, $2^-$, $3^+$ and $5^-$ respectively. The mechanism proposed to produce the equinoctial asymmetry, and in particular a difference of up to 70% in the meridional winds around midnight, follows from the expected diurnal variation in the high-latitude electric field.

First plasma from the dayside of the Earth is convected over the polar cap to the nightside. Momentum from the fast moving anti-sunward flow of ions over the pole is transferred to the neutrals during the few hours that the neutrals spend within the polar cap. As the ions come over to the nightside and down to the geomagnetic latitude of Kiruna they are diverted east and west into the dawn and dusk convection cells. This is the region of the Harang discontinuity. However, the neutrals are carried equatorwards by their inertia and the FPI observes a large southward meridional wind in the midnight sector of the auroral oval. The 12 hour phase difference in the diurnal variation of $\langle B'_{mz} \rangle$ in spring and autumn means that the time when Kiruna passes through the Harang dis-
continuity coincides with the time of the most negative value of $< B'_mz >$ in spring but the least negative value in autumn. The value of $< B'_mz >$ reflects the strength of the solar wind-magnetospheric coupling according to the solar rectifier model, and consequently the CPCP. Therefore the difference in $< B'_mz >$ at this time means that the momentum transferred to the neutrals passing over the pole and arriving in the midnight sector, where they are observed by the FPI at Kiruna, is greater in spring than in autumn.

The response of the zonal winds to the diurnal variation of $< B'_mz >$ is more ambiguous because it depends on the position of the site with respect to the ionospheric convection pattern. The auroral oval grows broader with increasing geomagnetic activity, as shown in the plots of the geomagnetic dependence of the latitudes of the northern and southern edges of the auroral belt for the period 22GLT-02GLT in Figure 3.27 from Feldstein and Starkov [1967]. Thus at low levels of activity Kiruna is outside the auroral oval, while at very high levels thermospheric winds typical of the polar cap are observed, as shown in Chapter 4 and also described by Lloyd [1985]. In addition, the centre of the auroral oval is offset antisunwards and duskwards of the geomagnetic pole [Feldstein and Starkov, 1967, Sojka et al., 1979] which means that a site that is just within the auroral oval in the evening hours will be outside during the dawn hours. Therefore the plasma velocities from one site alone cannot be used to prove or disprove this mechanism. This might explain why the EISCAT ion velocities do not show a seasonal and diurnal variation in plasma convection in exact accord with the Russell and McPherron [1973] model.

Similarly the combination of a UT and MLT dependence complicates the investigation into the possible influence of the seasonal and diurnal variation of $< B'_mz >$ on the electron densities at the height of the 6300Å measurements. For example, the monthly variation in electron density for a height region of 200—250 km measured by EISCAT radar [Farmer, private communication, 1990] is shown in Figure 3.28 for electron densities measured during the period of solar minimum (1984-1987). A semi-annual variation in the daytime electron densities appears as expected, but the noon maximum for spring is greater than that for autumn although autumn has a more southward IMF value.

Further evidence for a seasonal and diurnal variation of the high-latitude electric field due to the variation in the orientation of the Earth's magnetic dipole may come from a recent paper by de la Beaujardière et al. [1991]. This paper has determined the high-latitude electric field for the four seasons as observed by
the Sondrestrom incoherent scatter radar during the period of solar minimum 1983-1987. The CPCP was calculated from a least squares fit to the plasma convection on the assumption that the radar is sampling an average convection pattern that is fixed with respect to the geomagnetic pole over the whole 24 hour period. The CPCPs for autumn, winter, spring and summer were calculated to be, on average, 49.93 kV, 44.79 kV, 42.02 kV and 39.75 kV. Thus there is a large equinoctial asymmetry in the average cross polar cap potential, where the autumn value is 18% larger than spring. This result is possibly consistent with the results and theory presented here for the following reason: Although the Sondrestrom database has between 27,000 and 45,000 data points per season there is no indication in the paper as to how they are distributed with time. It is probable that the results show a sampling bias with respect to time since the best radar electron scatter observations occur during the day when the electron density is high due to photoionization, and around midnight when it is high through electron precipitation. Sondrestrom is at longitude 51°W, which means that 10UT and 22UT correspond to the local times 6.6LT and 18.6LT. At the equinoxes sunrise and sunset occur at 06LT and 18LT, therefore the observations of the asymmetry at 10UT, when the solar wind-magnetospheric coupling is predicted to be greater in autumn than in spring, will be well observed because the observations are made after dawn. However, the asymmetry at 22UT, when the coupling is greater in spring than in autumn, will suffer from poor signal to noise ratios because the observations are made just after dusk. Therefore, it is likely that on average the Sondrestrom radar will see more instances of the asymmetry at 10UT than the asymmetry at 22UT and consequently on calculating an overall average of the CPCP for each season it will appear that the average CPCP in autumn is greater than in spring.

*de la Beaujardière et al.* [1991] point out that there are several reasons why there should not be a seasonal variation in the high-latitude electric field. The main reason is that the CPCP must be almost the same value in both the northern and southern hemispheres because the magnetic field lines above the ionosphere are highly conducting [Farley, 1960] and consequently the magnetic conjugate points in the ionosphere are virtually equipotentials [Hines, 1963]. In accord with this they conclude that there is no evidence that the area of the polar cap changes with season and that the 12% difference in the summer and winter CPCP is only small and may be due to larger Birkeland currents as a result of increased photoionization in the summer. The 18% difference between
the two equinoxes is overlooked since it does not fit in with this explanation. However, the theory presented here accommodates these restrictions and yet allows for a seasonal and diurnal variation. For example, although the CPCP at the spring and autumn equinoxes are equal over a 24 hour period, there is a 12 hour phase difference in the diurnal variation between the two equinox seasons. Yet the CPCP in both hemispheres are equal at all times because on the 21st March the North Pole is at the spring equinox while the South Pole is at the autumn equinox. Then when the north geomagnetic pole is at the longitude corresponding to 22UT the south geomagnetic pole is at the longitude corresponding to 10UT, and so both are simultaneously in the orientation of maximum solar wind-magnetospheric coupling.

3.8.4 Further Asymmetry Introduced by $B_{my}$


Figure 3.22a and b show the average distribution of the IMF in the $(j_m, k_m)$ plane for spring and autumn respectively. As can be seen from these two figures and also from the equations 3.7 and 3.8, the $y$- and $z$-components of the IMF at equinox are interdependent in such a way that for spring $B_{my}B_{mz} > 0$ and for autumn $B_{my}B_{mz} < 0$. Figure 3.29 shows the electric equipotential field lines from the Heppner and Maynard model for the four possible permutations and combinations of positive and negative values of $B_{my}$ and $B_{mz}$. The four models are very different from one another. The two allowed average orientations of the IMF in the $(j_m, k_m)$ plane for spring are $(B_{my} < 0, B_{mz} < 0)$ and $(B_{my} > 0, B_{mz} > 0)$, while for autumn they are $(B_{my} < 0, B_{mz} > 0)$ and $(B_{my} > 0, B_{mz} < 0)$. Thus the combination of the $B_{my}$ and $B_{mz}$ components in their influence on the ion convection patterns introduces the possibility of a further asymmetry in the autumn and spring thermospheric winds.
3.8.5 Predictions from this Model

The effect of the seasonal variation in the inclination of the Earth’s magnetic dipole to the IMF on the high-latitude ionospheric convection has been presented here as an explanation of the equinoctial asymmetry in thermospheric winds observed by the FPI at Kiruna. This mechanism allows for a seasonal and diurnal variation in the CPCP without violating the condition that the CPCP in both hemispheres must be the same because the magnetic field lines are equipotentials. Several predictions may be made from this model as follows:

1) Theoretical models do not produce an asymmetry in the thermosphere and ionosphere at the spring and autumn equinoxes, but a very large equinoctial asymmetry has been observed in the neutral winds and ion velocities at high-latitudes. The results of this analysis would suggest that a seasonally and diurnally varying high-latitude electric field should be incorporated into these models in order to simulate the experimental observations. This is particularly important at the equinoxes where, theoretically, there can be a significant variation in the solar wind-magnetospheric coupling according to the solar rectifier model. The diurnal variation in the half-wave rectified IMF, \( <B'_{mz}> \), near the equinoxes ranges from -0.4 nT to -1.0 nT, on assuming an average IMF with magnitude 5 nT [Russell and McPherron, 1973]. Experimentally there was found to be a maximum diurnal variation of around -1.2 nT to -1.5 nT for February [Wild, private communication, 1991]. The diurnal variation of \( <B'_{mz}> \) at the two solstices is not so drastic, theoretically ranging between -0.3 nT to -0.6 nT, and may not be significant.

2) It is expected that only sites close to or within the auroral oval should see an equinoctial asymmetry since the asymmetry is dependent on the response of the high-latitude electric field to solar wind-magnetospheric coupling.

3) The large equinoctial asymmetry observed at Kiruna is dependent on the fortuitous combination of the observing site at Kiruna passing through the midnight sector of the auroral oval at around 21UT. At this time it was possible to observe the residual neutral flow resulting from convection over the pole during the period when the solar wind-magnetospheric coupling was at a maximum level for spring and at a minimum level for autumn. Therefore it would be expected that a high-latitude site which is 180° longitude around from Kiruna would also see an equinoctial asymmetry, but with the autumn neutral winds...
larger than the spring, since such a site will pass through the Harang discontinuity at a time close to 10UT when the solar wind-magnetospheric coupling is at a maximum for autumn and a minimum for spring. However, the observations of the IMF have shown that the optimum coupling at 10UT in autumn is not as effective as that at 22UT in spring, and so the observed asymmetry would not be as pronounced as that observed at Kiruna.

Further, observations of the nighttime neutral winds by FPIs at high-latitude sites that are removed by 90° longitude are unlikely to see any significant equinoctial asymmetry in the meridional winds because there will be little difference in the coupling efficiency at spring and autumn at the time when they pass through the Harang discontinuity.

4) An equinoctial asymmetry in the neutral winds might not be seen at the latitude of Kiruna in the cusp region because there is insufficient time for the neutral winds to build up a large poleward velocity through momentum transfer from ions convecting over the pole. Instead the difference between the autumn and spring meridional winds around 10UT will be as ambiguous as the difference between the zonal winds at around 22UT since any difference will depend on the position of the site with respect to the ionospheric convection pattern. In which case the results from a single site will not be adequate to prove or disprove this model.

5) Finally, the results of this analysis emphasize that it will not be possible to create an fair and adequate empirical model of the predicted seasonal and diurnal variation in the high-latitude ionospheric convection pattern from one site alone due to the UT and MLT dependence of the variation.
3.9 Summary

This chapter contains results and a plausible explanation of the seasonal and solar cycle variation of thermospheric winds observed at Kiruna, which is a high-latitude site, just within the bounds of the auroral oval. It is a lengthy chapter and so a summary of the analysis follows:

1) The underlying assumption of the seasonal variation of the upper atmosphere is that the summer and winter solstices define the two extremes of behaviour, while spring and autumn lie equivalently between these two extremes. However, a strong seasonal and solar cycle variation of thermospheric winds has been observed which does not comply with the predictions of theoretical models. Thermospheric winds in autumn appear to be more ‘winter-like’ than the winds in spring. Yet solar flux densities in spring and autumn are the same. The difference in the behaviour of the upper atmosphere at spring and autumn must therefore be due to changes in the momentum transferred to the neutrals from the ions. This is confirmed by solar minimum data from the EISCAT radar which show that electron densities in the height region corresponding to the 6300Å emission peak are larger in spring and autumn than in winter, and, in addition, spring electron densities are larger than autumn. Further, the average plasma velocities at 280 km measured by the EISCAT radar at solar maximum show spring vectors that are systematically larger than the autumn vectors, and autumn vectors that are similar to the winter vectors.

2) The initial investigation considered whether the asymmetry between the high-latitude spring and autumn thermosphere and ionosphere might be due to a phase difference between the seasonal variation of solar flux and the change in the chemical composition of the thermosphere due to the global circulation of the atmosphere.

The global circulation of the thermosphere drives particles from the hot summer pole towards the cold winter pole [Roble et al., 1977, 1983, 1987c, Hedin et al., 1977a, b, Hedin, 1987]. The particles move at different rates depending on the particle mass, which results in a seasonal variation in the mean molecular mass of the thermosphere. O\(^+\) is the dominant ion constituent of the ionosphere. Its concentration is dependent on the ratio of the chemicals which determine the production and loss of the O\(^+\) ion, i.e. [O]/[N\(_2\)] Rishbeth and Setty [1961]. In the winter pole light particles dominate and in the summer pole heavy particles

The ratio [O]/[N_2] was taken from the MSIS-86 model [Hedin, 1987]. The ratio was found to have a seasonal variation which peaks before winter solstice. The curve is an asymmetric function that builds up rapidly from a minimum value in early July to a maximum in early December, then descends again, slowly, with a slight increase around February, to its minimum value in summer. The seasonal variation in the total solar flux density for the day was calculated for both an optically thick and optically thin atmosphere since the optical depth at the height of the 6300Å peak lies somewhere between these values. The combined influence of the pressure gradients and ion drag on the thermosphere was then assumed to be proportional to the product of [O]/[N_2] and the total daily flux density. However, the seasonal variation of the resulting function could not account for the observed seasonal variation in thermospheric winds nor plasma drifts. Although a small asymmetry was apparent between the spring and autumn values of the function, the difference was found to be negligible when scaled by the difference between the equinox values and value at the winter solstice.

3) The observation that ion drag appeared to be larger at the equinoxes than at winter solstice then prompted an investigation into a possible seasonal variation of the high-latitude ionospheric convection pattern. There is a semi-annual variation in geomagnetic activity, which is well-documented [e.g. Cortie, 1912, Bartels, 1932]. According to the solar rectifier model by Arnoldy [1971], the efficiency of solar wind-magnetospheric coupling is proportional to the size of the component of the IMF along the z-axis in the GSM coordinate system for $B_{mz} < 0$. The seasonal variation in the orientation of the Earth's geomagnetic pole to the average IMF vector consequently results in a semi-annual variation in the solar wind-magnetospheric coupling efficiency. The most efficient coupling occurs on April 5th and October 5th, which is compatible with the phase of the semiannual variation in geomagnetic activity [Russell and McPherron, 1973].

There is also a diurnal variation in the coupling efficiency due to the rotation of the geomagnetic pole about the geographic pole. This diurnal variation is phase shifted by 12 hours between the two equinoxes. On April 5th the most efficient coupling is predicted to occur at 22UT while on October 5th it should occur at 10UT.
4) The seasonal and diurnal variation of the solar wind-magnetospheric coupling will result in a seasonal and diurnal variation in the high-latitude electric field. However, the CPs in both hemispheres must be the same at all times in accord with the principle that the magnetic field lines joining the North and South geomagnetic poles are near equipotentials [Hines, 1963]. This is accommodated by the seasonal dependence of the phase of the diurnal variation. For example, on the 21st March the Northern hemisphere is at the spring equinox while the Southern hemisphere is at the autumn equinox. Then when the North geomagnetic pole is at the longitude corresponding to 22UT the South geomagnetic pole is at the longitude corresponding to 10UT, and both poles are in the best orientation for maximum solar wind-magnetospheric coupling.

5) FPI observations can only be made during the night, and the largest equinoctial asymmetry observed at Kiruna occurs in the meridional winds in the period around midnight. It was suggested that this asymmetry is caused by an asymmetry in the momentum transferred to the neutrals travelling over the polar cap by ions which have been largely generated on the dayside of the Earth. At 22UT the cross polar cap plasma velocity is boosted in spring by the solar wind-magnetosphere system being in a state of optimum coupling, while coupling is at its minimum value in autumn. Therefore, during the time period around 22UT, the momentum transferred from the plasma to the neutral gas as it passes over the poles is greater in spring than in the autumn. At this time Kiruna is passing under the Harang discontinuity and so can observe the outflow of neutrals which have passed over the polar cap when the equinoctial asymmetry is at its greatest difference.

6) Although the meridional neutral winds are larger in spring than in autumn for both solar maximum and solar minimum conditions, only the ion velocities at solar maximum show the same asymmetry. At solar minimum the ion velocities are generally larger in autumn than in spring. This is presumed to be because the auroral oval is contracted north of Kiruna and Tromsø at solar minimum. Thus the ion velocities observed above Tromsø are those of a mid-latitude site which is not within the influence of the ionospheric convection field. Therefore Tromsø does not see the same equinoctial asymmetry in the ion velocities. However, the momentum given to the meridional winds as they pass over the pole is sufficiently large that the neutral gas overshoots the confines of the auroral oval and consequently the asymmetry can still be observed at
Kiruna, despite it being outside the auroral oval.

7) The Earth’s rotational axis is tilted in symmetrically opposite directions with respect to the \( j_\alpha, k_\alpha \) plane of the GSE coordinate system at spring and autumn (Figure 3.22a,b). As a result the average orientation of the \( y- \) and \( z- \)components (in GSM coordinates) of the IMF is determined by \( B_{my}B_{mz} > 0 \) for spring and \( B_{my}B_{mz} < 0 \) for autumn. Therefore, although there is an immense variability in the IMF vector, the geometry between the Earth’s axis and the mean IMF vector is such that on average the two permutations and combinations of \( B_{my}B_{mz} \) evident in spring are \((B_{my} < 0, B_{mz} < 0)\) and \((B_{my} > 0, B_{mz} > 0)\), while in autumn they are \((B_{my} < 0, B_{mz} > 0)\) and \((B_{my} > 0, B_{mz} < 0)\). This is a further source of asymmetry between the spring and autumn ionosphere, and consequently the thermospheric winds.

The UCL-Sheffield three-dimensional time-dependent coupled thermosphere-ionosphere model [Fuller-Rowell et al., 1987, Rees et al., 1988, Rees and Fuller-Rowell, 1989] will be used in a future project to provide a qualitative test of the combined influence of the \( B_{my} \) and \( B_{mz} \) components on the thermosphere and ionosphere, using the Heppner and Maynard [1987] empirical high-latitude electric field models for the four permutations and combinations of \( B_{my}B_{mz} \).
Figure 3.1
Figure 3.3

DATA SAMPLE TAKEN FROM:
FEBRUARY TO APRIL
AVERAGE NEUTRAL WINDS
KIRUNA 6300A

UT 12 18 00 06 12

17/11/81 to 10/4/90
Figure 3.4
Figure 3.5
Figure 3.6

Representative electron density profiles at high-, middle- and low-latitudes for three levels of solar activity ($R = 0, 100, 200$) based partly on bottomside ionograms and partly on theory [Wright, 1962].
Seasonal variation in plasma drifts at Tromsø at solar maximum

[Farmer and Jarvis, 1991]

AUTUMN: 1988 - 1990  \( \cdot \cdot \cdot V = 100 m/s \) Northwards CP-1 + CP-2 NUN-1

WINTER: 1988 - 1990  \( \cdot \cdot \cdot V = 100 m/s \) Northwards CP-1 + CP-2 NUN-1

SPRING: 1988 - 1990  \( \cdot \cdot \cdot V = 100 m/s \) Northwards CP-1 + CP-2 NUN-1

SUMMER: 1988 - 1990  \( \cdot \cdot \cdot V = 100 m/s \) Northwards CP-1 + CP-2 NUN-1
Seasonal variation in plasma drifts at Tromsø at solar minimum
[Farmer and Jarvis, 1991]

AUTUMN: 1984 - 1987
- O- $v = 100$ m/s Northwards CP-1 + CP-2 RNN-1

WINTER: 1984 - 1987
- O- $v = 100$ m/s Northwards CP-1 + CP-2 RNN-1

SPRING: 1984 - 1987
- O- $v = 100$ m/s Northwards CP-1 + CP-2 RNN-1

SUMMER: 1984 - 1987
- O- $v = 100$ m/s Northwards CP-1 + CP-2 RNN-1
MONTHLY VARIATION OF SOLAR FLUX AND POSITION OF SOLAR TERMINATOR AT LATITUDE 68N

Figure 3.9

150
Figure 3.10
MONTHLY VARIATION OF [O]/[N2] AND DAYLIGHT HOURS
AT PRESSURE LEVEL 11, f10.7 = 179, latitude: 68 N

Figure 3.11
MONTHLY VARIATION OF [O] AND [N2]
AT PRESSURE LEVEL 11, $f_{10.7} = 179$, latitude: 68 N

Figure 3.12

153
Figure 3.13
Typical 'butterfly' diagram showing the yearly variation in distribution of sunspots during a sunspot cycle. The distribution migrates toward the solar equator. Handbooks of Geophysics, 1955.

Figure 3.14

a) weekly mean maximum, b) weekly mean, c) weekly mean minimum

[McIntosh, 1959]

Annual variation of the daily U index for active (top curve), intermediate (middle curve) and quiet (bottom curve) years averaged over the years 1872-1930 [Chapman and Bartels, 1940]
Figure 3.17

a) Northward IMF resulting in a closed magnetosphere
b) Southward IMF resulting in an open magnetosphere

[Russell, 1972]
Figure 3.18

Satellite observations of the Interplanetary Magnetic Field. The orientation is given in terms of the Geocentric Solar Ecliptic coordinate system.
Figure 3.19

View of the ecliptic plane illustrating the average orientation of the Interplanetary Magnetic Field vector and the position of the Earth's North Geographic Pole at the solstices and the equinoxes.

$X =$ position of the North Geographic Pole

Figure 3.20

Schematic illustration of the warped solar current sheet which divides interplanetary space into regions of magnetic field pointing towards or away from the sun. The region illustrated by the diagram covers approximately 20 astronomical units.
Figure 3.21

A contour plot of the diurnal and annual variation of the effective average southward component, $B'_{mz}$, of the IMF in the GSM coordinate system due to the towards and away sectors along an ideal spiral field. The more negative the contour the more geomagnetically active the interval should be on average [Russell and McPherron, 1973]. $B'_{mz}$ has been calculated in the following manner:

$$B'_{mz} = \begin{cases} 
B_{mz}, & \text{for } B_{mz} < 0 \\
0, & \text{for } B_{mz} \geq 0 
\end{cases}$$
The average orientation of the IMF with respect to the GSM coordinate system is such that in spring $B^z > 0$ while in autumn $B^z < 0$. However, the orientation of the GSM axes (i.e., $J^m$, $J^n$, $J^m$) with respect to the GSE axes is different at the spring and autumn equinoxes owing to the tilt of the Earth's rotational axis with respect to the ecliptic plane. Thus, the average orientation of the IMF in the $J^m$, $J^n$, $J^m$ plane is inclined at an angle $B = (-10 \pm 5)^\circ$. The mean distribution of the IMF in the $J^m$, $K^m$ plane is given by $B = (10 \pm 5)^\circ$. The average vector total plane. Thus, in the GSE coordinate system, the average vector lies in the solar equator.

**Figure 3.22**

[Diagram showing IMF orientation in spring and autumn equinoxes]
Figure 3.23

Seasonal variation of the effective average southward component, $B_{mz}$, of the IMF in the GSM coordinate system calculated from the one-hour averages of satellite data collected between 1963 and 1990 [Wild, 1991]. The solar rectifier model was used to calculate $B'_{mz}$, as for Figure 3.21:

$$B'_{mz} = \begin{cases} B_{mz}, & \text{for } B_{mz} < 0 \\ 0, & \text{for } B_{mz} \geq 0 \end{cases}$$
Seasonal variation of the effective average southward component, $B_{ms}$, of the IMF for six 4-hour intervals, averaged over the years 1963-1990. (Wild, 1991)
Normalised reconnection voltage between closed geomagnetic field and open IMF vs. IMF clock angle, $\theta$. Also shown for reference are the curves $\sin(\theta/2)$ and $\sin^2(\theta/2)$. [Fedder et al., 1991]
Figure 3.26

Equipotential convection contours derived from Millstone Hill radar observations for four levels of high-latitude particle precipitation. A 2kV contour spacing is shown in polar apex magnetic latitude/local time coordinates. [Foster et al., 1986]
Figure 3.27

The dependence of the location of the northern and southern edges of the auroral belt on geomagnetic activity as represented by the $Kp$ index, in the period around midnight (22-02 hours) in Local Geomagnetic Time. $\psi'$ is the corrected geomagnetic latitude. The dashed line represents the centre line of the auroral belt. [Feldstein and Starkov, 1967]
Figure 3.28

Monthly and diurnal variation in electron densities at solar minimum measured by the EISCAT radar at Tromsø [Farmer, 1990].
Figure 3.29

With reference to Figure 3.22, this figure shows models of the four high-latitude electric field patterns for a) $B_y < 0, B_z > 0$, b) $B_y > 0, B_z > 0$, c) $B_y < 0, B_z < 0$ and d) $B_y > 0, B_z < 0$ [Heppner and Maynard, 1987]. In terms of the GSM coordinate system, the $y$- and $z$- components of the IMF are ordered so that in spring $B_{my}B_{mz} > 0$ while in autumn $B_{my}B_{mz} < 0$. This should therefore introduce a further source of asymmetry to the behaviour of the high-latitude ionosphere and consequently the thermosphere at spring and autumn.
Geomagnetic and Solar Cycle Variations in High-Latitude Thermospheric Winds

4.1 Introduction

The $K_p$ index is a measure of the disturbance of the geomagnetic field due to the input of energy from the solar wind to the magnetosphere. The index is based on readings from standard magnetometers [Bartels, 1957]. The variation of high-latitude thermospheric winds with geomagnetic activity has been demonstrated previously by this and other groups, though with much smaller data bases [Lloyd, 1985, Sica et al., 1986a, Rees et al., 1987b,c]. These past papers, and others such as Hays et al. [1984] and McCormac et al. [1987], have already established a link between geomagnetic activity and the size of the auroral oval and the consequent effect on thermospheric winds of the convective electric fields and Joule heating at auroral latitudes.

In the study presented here the gross effect of solar flux, in addition to that of geomagnetic activity, is investigated. Observations of thermospheric winds at low- and mid-latitude sites, using FPI and incoherent scatter radar, have not shown solar cycle variations [Burnside and Tepley, 1989, Sipler et al., 1982, Babcock and Evans, 1979]. This can be accounted for in terms of the competition between solar heating and ion drag. At solar minimum the EUV flux is less than at solar maximum, therefore the level of solar heating is lower and the pressure gradients are smaller. Due to the smaller amount of photoionization from the lower flux levels the plasma density, and consequently the ion drag, is also lower than at solar maximum. At solar maximum the EUV flux is 2-3 times larger than at solar minimum [Torr et al., 1980c, Lean, 1988], resulting in larger pressure gradients. However, the plasma density is also larger due to the increased levels of photoionization. As a result the balance between solar heating and ion drag is maintained approximately, as evidenced by the lack of a solar cycle effect on the thermospheric winds at low- and mid-latitudes.

The high-latitude thermosphere presents a different situation. At the poles a horizontal electric field is induced by the flow of the solar wind plasma across the geomagnetic field lines. A two cell convection flow of ions is set up such that in the dawn and dusk sectors of the auroral oval the ions move in a sunward
direction, contrary to the direction of the global pressure gradients, while within the polar cap the ions travel almost parallel to the pressure gradients in an anti-sunward direction. Thus at high latitudes ion drag acts as a driving force that can act both with and against the pressure gradients rather than consistently acting against them, as is the case at low- and mid-latitudes. Therefore the additional acceleration on the neutral winds due to an increase in the pressure gradients through greater solar flux levels will not be negated necessarily by the concurrent rise in ion drag. As such there is reason to expect to see a solar cycle variation in the thermospheric winds at high-latitudes.

4.2 Description of the Database

The data were first separated according to solar activity: into solar maximum and solar minimum periods, just as for the seasonal and solar activity analysis of Chapter 3. Data were then further sorted according to $K_p$ value. $K_p$ indices range from 0 to 9 and are given for each three hour period of a Greenwich day, starting with 00UT-03UT and ending with 21UT-24UT, thereby giving eight values per day. The Kiruna data were binned by the corresponding three hour period into which each data point fell. No attempt was made to interpolate the $K_p$ values when binning the data in order to give a coarse measure of the level of activity; this may then be used to represent the time lag of 1-2 hours in the response of the thermosphere to any change in the ionosphere [Killeen et al., 1984a, Thayer et al., 1987, Sica et al., 1986b]. For each level of solar activity the data were ordered according to the following three bands of geomagnetic activity:

\[
\begin{align*}
0^\circ \leq K_p < 2^\circ & : \text{geomagnetically quiet data} \\
2^\circ \leq K_p < 5^\circ & : \text{geomagnetically moderate data} \\
5^\circ \leq K_p < 8^\circ & : \text{geomagnetically active data}
\end{align*}
\]

Thus there are six categories altogether of solar and geomagnetic activity. The values of the $K_p$ indices used to bracket the geomagnetically ‘quiet’, ‘moderate’ and ‘active’ levels are approximately those in general use [e.g. Sica et al., 1986a]. One of the consequences of using these ranges in $K_p$ is that the data are not evenly distributed amongst the six categories so that the geomagnetic activity range of $2^\circ \leq K_p < 5^\circ$ is the best represented and the range of $5^\circ \leq K_p < 8^\circ$ is the poorest represented. The average number of nights contributing to each 15 minute interval in each category, $N$, is given in Table 4.1 which in addition contains the average values of various parameters defining the observing
conditions of this database. The data are also unevenly distributed amongst the hours through the night. The least well represented times are at the beginning and end of each night due to the seasonal variation in the length of a night. Thus, the first and last couple of hours are biased to represent the longest nights in the sample category in the same way as occurs with the thermospheric winds analysed with respect to season in Chapter 3.

As with the seasonally averaged data, the descriptions of the thermospheric wind behaviour will concentrate on the observations made to the North and West of Kiruna because these data are averaged from a larger number of samples than the observations to the South and East, as explained in Chapter 3.

4.3 Geomagnetic and Solar Cycle Variations

Figures 4.1 and 4.2 show averaged horizontal wind components from the Kiruna FPI database for solar maximum and solar minimum conditions, respectively, at three levels of geomagnetic activity. Figures 4.1a and 4.2a show the average horizontal and vertical wind components under geomagnetically quiet conditions where $0^\circ < Kp < 2^\circ$. Figures 4.1b and 4.2b show average winds under moderate geomagnetic activity where $2^\circ < Kp < 5^\circ$. Figures 4.1c and 4.2c show the average winds under geomagnetically active conditions where $5^\circ < Kp < 8^\circ$. The coordinates are geographic and the time is Universal Time (UT). An arrow on each plot indicates the direction of positive wind magnitude. As with the seasonally averaged data in Chapter 3, the vertical bars are the standard deviation of the average wind speeds for each data point and are not error bars.

Tables 4.2 and 4.3 give smoothed average meridional and zonal winds at four sample times for solar maximum and solar minimum respectively. The winds are an average of the mean speeds from five consecutive 15 minute intervals centred on 18UT, 21UT, 24UT, and 03UT.

$0^\circ < Kp < 2^\circ$: geomagnetically quiet conditions

The thermospheric winds observed during geomagnetically quiet conditions at solar maximum and solar minimum are qualitatively very similar.

The meridional wind at solar minimum turns southward at around 18UT and remains so for the next 11 hours. The meridional wind at solar maximum also turns southward at 18UT, but remains southward for only just over 10 hours. The peak nighttime meridional wind is, on average, 91 m/s southward at solar minimum and 129 m/s at solar maximum, at around midnight. This is a 42% difference in magnitude.
The average zonal winds at both solar maximum and solar minimum are eastward from the beginning of the observations at around 15UT until about 19UT, which indicates that the neutral winds above Kiruna show little or no response to ion drag from within the dusk auroral oval. Yet, although qualitatively similar in direction there is a solar cycle dependence of the magnitude of the winds: the zonal winds are smaller at solar maximum. For example, at 18UT to the west of Kiruna, the average zonal wind at solar minimum is 103 m/s eastward, while at solar maximum it is only 27 m/s.

The winds are westward at both levels of solar activity just before entering the midnight sector, which shows that Kiruna is now under the influence of the westward flow of the dusk cell of the auroral oval. Again this is to be expected owing to the distorted shape of the oval which is broadest in the midnight sector and narrowest in the midday sector, as illustrated in Figure 1.13. However, the zonal winds turn westward earlier at solar maximum and remain westward for a longer period of time. Thus the period of westward flow at solar maximum occurs between 18UT and 22UT, while for solar minimum it occurs between 19:45UT and 22:45UT. The maximum westward wind speed attained is 56 m/s at solar maximum and 35 m/s at solar minimum.

The zonal winds subsequently turn eastward and remain flowing eastward until 03:15UT at solar maximum and 02:30UT at solar minimum, reaching a peak speed of 51 m/s at solar maximum and 57 m/s at solar minimum. In the morning hours, when the zonal winds turn strongly westward again, it is probable that Kiruna has moved outside the vicinity of the eastward flow of the dawn cell of the auroral oval, and is now once again observing the response of the neutral gas to the antisunward forcing of the pressure gradients.

\[2^\circ \leq Kp < 5^\circ: \text{moderate geomagnetic activity}\]

The magnitude of thermospheric winds under moderately active geomagnetic conditions are generally larger than under quiet conditions. The meridional wind components are southward between 16:15UT and 05:15UT at solar minimum, while at solar maximum they are southward between 17UT and 05:30UT.

At this level of geomagnetic activity the auroral oval is extended overhead of Kiruna during the whole nighttime period. As a result the zonal wind components are westward for both levels of solar activity until about 22UT. However, the zonal winds at solar maximum are more strongly westward than at solar minimum, reaching a peak value of 158 m/s, while at solar minimum the peak
value is merely 86 m/s. Between 22UT and 03:30UT the solar minimum zonal winds are eastward, reaching a peak value of 37 m/s. In comparison the zonal winds at solar maximum are eastward between 21:30UT and 03:45UT, which is 45 minutes longer than at solar minimum, and attain a larger maximum speed of 72 m/s. By dawn the zonal winds have turned westward once more.

$5^\circ \leq K_p < 8^\circ$: geomagnetically active conditions

During geomagnetically active conditions the magnitudes of the average thermospheric winds are extremely large. The magnitude of the winds at solar maximum can be up to 2-3 times larger than at the same time at solar minimum. At solar maximum the peak meridional wind speed is 236 m/s, which is 53% larger than the peak meridional wind speed at solar minimum. There is a southward component for $13\frac{1}{2}$ hours at solar maximum. The average meridional wind at solar minimum changes direction at 15:45UT to become southward but does not turn northward again during the period of observation. Thus it is not possible to know the average length of time for which there is a southward component of the thermospheric wind during geomagnetically active conditions at solar minimum.

The database for geomagnetically active conditions at solar minimum is the smallest of all the categories, and so the average winds do not vary as smoothly with time as do the average winds in the other categories. The times when the wind velocity components turn from east to west, or from north to south, are consequently not well defined. The zonal wind observed to the east of Kiruna shows a small eastward excursion between 17UT and 18UT, then a small westward excursion until 19:15UT, before turning eastward for the next $9\frac{1}{4}$ hours. The general shape of the variation between the beginning of observation until about 20UT is similar for the winds observed to the west of Kiruna, but the zonal winds remain westward throughout.

The zonal wind components in the dusk sector during geomagnetically active periods at solar maximum are very large and westwards, reaching speeds of over 300 m/s. Then between 19:45UT and 05UT the zonal winds are eastward, with a peak value of 363 m/s.

There is an interesting systematic change in the length of time for which the nighttime meridional winds remain southward. This time increases with geomagnetic activity from a minimum of 11 hours for geomagnetically quiet periods during solar minimum, to $13\frac{1}{2}$ hours for geomagnetically active periods at solar
maximum. The length of time during which there is a southward component appears also to depend on the level of solar activity. For a given level of geomagnetic activity the average length of time that a southward component exists is consistently a half hour less for solar maximum data than for solar minimum.

Table 4.1

AVERAGE PARAMETERS ASSOCIATED WITH $Kp$
AVERAGED THERMOSPHERIC WINDS
FOR SOLAR MAXIMUM $F_{10.7} > 110$

<table>
<thead>
<tr>
<th>$Kp$ range</th>
<th>$&lt; Kp &gt;$</th>
<th>average date</th>
<th>$N$ per 15 min.interval</th>
<th>$F_{10.7}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$0^\circ \leq Kp &lt; 2^\circ$</td>
<td>$1^\circ$</td>
<td>26th Dec</td>
<td>20±10</td>
<td>178±23</td>
</tr>
<tr>
<td>$2^\circ \leq Kp &lt; 5^\circ$</td>
<td>$3^+$</td>
<td>8th Jan</td>
<td>82±38</td>
<td>180±23</td>
</tr>
<tr>
<td>$5^\circ \leq Kp &lt; 8^\circ$</td>
<td>$6^\circ$</td>
<td>9th Feb</td>
<td>13±6</td>
<td>179±26</td>
</tr>
</tbody>
</table>

AVERAGE PARAMETERS ASSOCIATED WITH $Kp$
AVERAGED THERMOSPHERIC WINDS
FOR SOLAR MINIMUM $F_{10.7} < 110$

<table>
<thead>
<tr>
<th>$Kp$ range</th>
<th>$&lt; Kp &gt;$</th>
<th>average date</th>
<th>$N$ per 15 min.interval</th>
<th>$F_{10.7}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$0^\circ \leq Kp &lt; 2^\circ$</td>
<td>$1^\circ$</td>
<td>31st Dec</td>
<td>26±13</td>
<td>89±17</td>
</tr>
<tr>
<td>$2^\circ \leq Kp &lt; 5^\circ$</td>
<td>$3^+$</td>
<td>27th Dec</td>
<td>69±41</td>
<td>93±17</td>
</tr>
<tr>
<td>$5^\circ \leq Kp &lt; 8^\circ$</td>
<td>$6^-$</td>
<td>9th Dec</td>
<td>8±4</td>
<td>97±20</td>
</tr>
</tbody>
</table>
Table 4.2

*Kp AVERAGED THERMOSPHERIC WINDS FOR SOLAR MAXIMUM* $F_{10.7} > 110$

averaged meridional winds observed to the north of Kiruna
( m/s, positive South )

<table>
<thead>
<tr>
<th>$Kp$</th>
<th>18:00</th>
<th>21:00</th>
<th>24:00</th>
<th>03:00 UT</th>
</tr>
</thead>
<tbody>
<tr>
<td>$0^\circ \leq Kp &lt; 2^\circ$</td>
<td>6.4</td>
<td>79.2</td>
<td>122.3</td>
<td>51.0</td>
</tr>
<tr>
<td>$2^\circ \leq Kp &lt; 5^\circ$</td>
<td>38.9</td>
<td>125.3</td>
<td>146.5</td>
<td>74.7</td>
</tr>
<tr>
<td>$5^\circ \leq Kp &lt; 8^\circ$</td>
<td>67.3</td>
<td>207.5</td>
<td>213.9</td>
<td>87.2</td>
</tr>
</tbody>
</table>

averaged zonal winds observed to the west of Kiruna
( m/s, positive East )

<table>
<thead>
<tr>
<th>$Kp$</th>
<th>18:00</th>
<th>21:00</th>
<th>24:00</th>
<th>03:00 UT</th>
</tr>
</thead>
<tbody>
<tr>
<td>$0^\circ \leq Kp &lt; 2^\circ$</td>
<td>-10.4</td>
<td>-36.7</td>
<td>38.0</td>
<td>-1.3</td>
</tr>
<tr>
<td>$2^\circ \leq Kp &lt; 5^\circ$</td>
<td>-131.8</td>
<td>-20.4</td>
<td>65.1</td>
<td>36.5</td>
</tr>
<tr>
<td>$5^\circ \leq Kp &lt; 8^\circ$</td>
<td>-186.7</td>
<td>28.3</td>
<td>51.6</td>
<td>68.4</td>
</tr>
</tbody>
</table>

Table 4.3

*Kp AVERAGED THERMOSPHERIC WINDS FOR SOLAR MINIMUM* $F_{10.7} < 110$

averaged meridional winds observed to the north of Kiruna
( m/s, positive South )

<table>
<thead>
<tr>
<th>$Kp$</th>
<th>18:00</th>
<th>21:00</th>
<th>24:00</th>
<th>03:00 UT</th>
</tr>
</thead>
<tbody>
<tr>
<td>$0^\circ \leq Kp &lt; 2^\circ$</td>
<td>3.7</td>
<td>49.0</td>
<td>79.9</td>
<td>66.4</td>
</tr>
<tr>
<td>$2^\circ \leq Kp &lt; 5^\circ$</td>
<td>27.7</td>
<td>90.8</td>
<td>87.7</td>
<td>58.1</td>
</tr>
<tr>
<td>$5^\circ \leq Kp &lt; 8^\circ$</td>
<td>45.6</td>
<td>116.6</td>
<td>114.1</td>
<td>53.1</td>
</tr>
</tbody>
</table>

averaged zonal winds observed to the west of Kiruna
( m/s, positive East )

<table>
<thead>
<tr>
<th>$Kp$</th>
<th>18:00</th>
<th>21:00</th>
<th>24:00</th>
<th>03:00 UT</th>
</tr>
</thead>
<tbody>
<tr>
<td>$0^\circ \leq Kp &lt; 2^\circ$</td>
<td>76.8</td>
<td>-24.9</td>
<td>41.8</td>
<td>-24.9</td>
</tr>
<tr>
<td>$2^\circ \leq Kp &lt; 5^\circ$</td>
<td>-73.7</td>
<td>-27.7</td>
<td>34.0</td>
<td>8.4</td>
</tr>
<tr>
<td>$5^\circ \leq Kp &lt; 8^\circ$</td>
<td>-76.3</td>
<td>1.7</td>
<td>33.8</td>
<td>59.1</td>
</tr>
</tbody>
</table>

175
4.4 Discussion

The results of this chapter are a continuation and extension of the work of Lloyd [1985] on the database of FPI measurements of thermospheric winds above Kiruna. There are two main differences between Lloyd's original work and this. Firstly, the Kiruna FPI data presented by Lloyd [1985], and also interpreted in terms of the UCL 3D TD thermospheric model by Rees et al. [1987a,b], were obtained between November 1981 and December 1984, in the decaying phase of solar cycle 21, but encompassing the peak levels of solar activity in 1981. Therefore the thermospheric wind samples were biased towards high solar activity levels. In contrast the database presented here covers most of an eleven year solar cycle and the data are consequently evenly disposed between solar maximum and solar minimum levels.

Secondly, the Lloyd database was analysed using the nightly sum of $K$ values ($K_{sum}$) while the present database has been analysed in terms of the 3-hour $Kp$ values. The $Kp$ index refers to the planetary geomagnetic activity. It is derived from weighted, averaged measurements of local activity at several magnetic observatories scattered over the world, which are given in terms of the $K$ index. The Lloyd database used the $K$ index compiled at the Swedish Institute of Space Physics at Kiruna, where the FPI is based. Data sorted according to the $Kp$ index or $K$ index are very similar because the differences between the $Kp$ and $K$ index values are generally relatively small, and are not systematic. The differences reflect the localised structures of auroral activity and electrojets, which are fairly arbitrary occurrences and are consequently smoothed out of a large compilation of data. There is, however, a difference between $Kp$ and $K_{sum}$ averaged data because the $K$ and $Kp$ indices vary in a quasi-logarithmic manner with respect to the level of geomagnetic disturbance, thus use of the $K_{sum}$ can give a distorted view of the activity of that day. As an example Lincoln [1967] hypothesised two days with $K_{sum} = 8$, where one day has $K$ indices of 1111 1111 and the other has $K$ indices of 0000 0008. By using $K_{sum}$ both the days would be assumed to be quiet days when in fact only the first is a quiet day, with a total range of 10 nT, whereas the second is highly disturbed with a range of over 300 nT.

Lloyd [1985] analysed the Kiruna data using $K_{sum}$ in an attempt to take into account the time lag between a change in the geomagnetic conditions and the response of the thermosphere i.e. $K_{sum}$ would incorporate the history of that day's activity. Unfortunately, as shown above, $K_{sum}$ can also give a misleading
picture of the history, which may contribute to the discrepancies between his observational results and the predictions of the thermospheric model simulations used in the comparison by Rees et al. [1987b]. For instance, it was noticed that the averaged neutral winds during periods of low activity ($K_{sum} < 15$) were stronger than expected. This can easily be attributed to the variety of permutations and combinations of $K$ values that are possible to achieve $K_{sum} < 15$, which, as with the example from Lincoln [1967], can encompass all levels of activity. Even at the highest values of $K_{sum}$, where the number of possible permutations and combinations of $K$ values are smaller, a pure response to geomagnetic activity was obscured in the Lloyd [1985] database by the need to use a wide range of $K$ values. Such a wide range was chosen in order to include a quantity of data comparable to that in the other ranges. In addition, the rapid response of thermospheric winds to large fluctuations in the auroral and geomagnetic activity was lost, smoothed out by the averaging process. The latter point is a problem with all averaging procedures. Therefore, as a measure of the geomagnetic activity associated with an observation, the three hour $Kp$ value is probably better than $K_{sum}$ and is the parameter of geomagnetic activity used in this analysis.

The general variation of the neutral thermospheric winds above Kiruna with $Kp$ has been explained and interpreted already. Lloyd [1985] has shown that at low levels of geomagnetic activity the neutral wind flow above Kiruna is dominated by the global effect of solar heating. This corresponds to the auroral oval being contracted polewards of Kiruna. With increasing $Kp$ the auroral oval expands equatorwards and the thermosphere above Kiruna comes under the direct influence of magnetospheric convection and particle precipitation associated with the auroral oval. Within the auroral oval ions are driven by magnetospheric convection and through momentum transfer from collisions between ion and neutrals the neutral gas is drawn into the convection pattern. Generally the higher the value of $Kp$ the stronger the ion drag and Joule heating in the vicinity of the auroral oval, and consequently the neutral flow becomes increasingly diverted from the simple flow patterns developed in response to pure solar heating and ionization.

The same trends in behaviour were shown by Sica et al. [1986a] for the variation of thermospheric winds with respect to the $Kp$ index over College, Alaska (65°N, 148°W). Figure 4.3, taken from this paper, shows the average neutral wind vectors obtained for low, medium and high levels of geomagnetic
activity. The vectors are plotted in geomagnetic coordinates, on a dial with geomagnetic north at the centre. Both of the sites: College, Alaska and Kiruna, Sweden, are at 65° magnetic latitude, but the orientation of the geomagnetic pole with respect to geographic north is different for each site. At Kiruna the geomagnetic pole is oriented at 12° west of geographic north, while at College the geomagnetic pole is about 20° east of geographic north. Yet changes in geomagnetic activity affect the College and Kiruna data in a similar manner despite this difference.

The magnitude of the influence of geomagnetic activity on the thermosphere has been shown to be very similar to that of solar flux by Sica et al. [1986a]. They found that the average increase in thermospheric temperature due to an increase in geomagnetic activity from low geomagnetic activity ($0 \leq Kp < 3^-$) to high ($Kp > 5^-$) is around 250 K, which is similar to the temperature increase due to the change in solar flux between solar minimum and solar maximum. It might be expected that the response of the composition and dynamics of the thermosphere to these (almost) equivalent temperature changes is also to the same degree. The problem now is to identify how the combination of geomagnetic activity, as represented by the $Kp$ index, and the variation of the solar flux throughout the solar cycle influences the thermosphere.

Solar EUV radiation increases by approximately threefold between solar minimum and solar maximum [Torr et al., 1980, Lean, 1988]. This increase has a direct effect on the neutral gas in two ways: increased dayside temperatures generate larger pressure gradients which accelerate the neutral particles to greater speeds; but this is counteracted by the greater ion drag due to more photoionization and a consequently higher ion density. Burnside and Tepley [1989] used this argument to explain why they did not see any significant solar cycle influence on thermospheric winds observed using a FPI between 1980 and 1987 at Arecibo, which is a low-latitude site, dominated by solar heating. This argument does not apply to the thermospheric winds observed over Kiruna, which is a high-latitude site under the influence of the auroral oval.

The auroral oval is an additional source of heating and ionization localised at high-latitudes. The solar cycle variation shown in the thermospheric winds at Kiruna, but not in the winds at Arecibo, indicate that the auroral oval must, therefore, introduce new conditions that throw out of kilter the careful balance between solar heating and ion drag which at low-latitudes appears to defy the cyclical variation in solar flux.

178
Why the average thermospheric winds observed at Kiruna for geomagnetic activity levels in the region $0^\circ \leq Kp < 2^\circ$ are qualitatively very similar at solar minimum and solar maximum, yet very different at higher levels of geomagnetic activity might be explained in terms of the latitudinally-selective influence of the solar cycle, due to the presence of the auroral oval. Figure 4.4 shows the calculated mass flow for December solstice conditions at solar maximum and solar minimum from the numerical model by Roble et al. [1977]. The predicted general flow is from the summer hemisphere to the winter hemisphere, but at solar maximum a reverse cell exists at high-latitudes in the winter hemisphere due to high-latitude heating. At solar minimum, according to this model, the auroral activity is too low to drive the reverse cell. A comparison of the solar maximum and minimum mass flows reveals that below about latitude $40^\circ$ the general flow is not much altered by an increase in solar flux. In contrast, the high-latitude thermosphere is radically different.

The latitudinal extent of the reverse cell is also dependent on geomagnetic activity which affects the radius of the auroral oval. For $Kp \leq 2$ the auroral oval is contracted polewards of Kiruna [Feldstein and Starkov, 1967, Evans et al., 1987], so the conditions in the thermosphere above Kiruna will be more typical of a low- or mid-latitude site, and as such might not be expected to show a solar cycle variation according to the Roble et al. [1977] model. With increasing $Kp$ the auroral oval expands, until at medium to high values of $Kp$ the auroral oval begins to dominate the behaviour of the thermosphere above Kiruna and as a consequence a solar cycle dependence appears.

However, the most interesting aspect of the combined influence of $Kp$ and the solar cycle is not that the neutral winds are on average larger at solar maximum but the observation that the solar cycle appears to produce a separation in the response of the neutral winds to $Kp$. Figures 4.5a and b show the variation of neutral wind vectors with $Kp$ for Kiruna in the same format as Figure 4.3, but here the data is separated into solar minimum and solar maximum averages. There is still the same general increase in the response of the neutral wind to each increment in $Kp$ for both solar maximum and minimum, but the higher levels of EUV flux at solar maximum appear to precipitate this increase in the response.

The different response of the neutral winds at solar maximum and minimum is better illustrated when the $Kp$ dependence is broken down into smaller categories, so that each level of geomagnetic activity is separated from the next
by one unit of $Kp$ value. It was found that at solar minimum ion drag begins to dominate over the driving force of the pressure gradients at about $Kp = 4$, while at solar maximum this occurs at $2 < Kp < 3$ [Aruliah et al., 1991]. The manifestation of this is that for quieter geomagnetic activity levels the neutral winds in the dusk sector, between 18 UT and 24 UT, are generally quite strongly eastward (antisunward), which shows the domination of the pressure gradients over ion drag, while for more active levels the neutral winds turn increasingly westward (sunwards), which shows that the neutrals are being drawn into the ion convection pattern through ion drag.

The progress of the response of the thermosphere to geomagnetic activity at solar maximum and solar minimum can be seen by finding the best matches of the meridional and zonal winds and comparing the associated range in $Kp$. For example, comparison of the winds from solar minimum and solar maximum reveals that the winds are qualitatively the most similar when $2 \leq Kp < 3$ for solar minimum and $1 \leq Kp < 2$ for solar maximum. Figure 4.6 focuses on the zonal winds for this comparison. In this case the activity level of the thermosphere at solar minimum is about one unit of $Kp$ behind solar maximum. Moving on to a comparison of winds at a more geomagnetically active level reveals that the best match between solar minimum and solar maximum winds has a lag of almost two units in $Kp$ value, as illustrated in Figure 4.7 which compares the zonal winds for solar minimum where $4 \leq Kp < 5$ and solar maximum where $2 \leq Kp < 3$.

The most obvious and simplest conclusion of this investigation is that the solar cycle dependence of the response of the high-latitude thermosphere to geomagnetic activity as measured by the $Kp$ index would indicate that the $Kp$ index is not a unique and sufficient measure of the transfer of energy and momentum from the magnetosphere to the high-latitude ionosphere and thermosphere. This conclusion has been reached by several others already, resulting in new indices of geomagnetic activity being developed, such as the Auroral Electrojet ($AE$) index which is specific to high-latitudes [Davies and Sugiura, 1966]. However, these results should also provoke investigation into why the $Kp$ index is an insufficient measure when the solar cycle is taken into account.

The main criticism of the $Kp$ index is that the network of monitoring stations is biased to the Northern Hemisphere and has a limited coverage of Local Times, as is shown by the distribution of observatories measuring $Kp$ in Fig-
Apart from this criticism of the global sampling statistics the results of the analysis presented in this chapter may indicate a bias in the sampling with respect to height. The measurements of fluctuations in the geomagnetic field are made by ground-based magnetometers and therefore may be dominated by the behaviour of induced currents in the lower part of the ionosphere rather than being typical of the whole ionosphere and magnetosphere. Due to the presence of the Earth's magnetic field electric conductivity within the upper atmosphere is anisotropic. The direction of currents depends on the balance between electron and ion gyrofrequencies and also on the collision frequencies between ions and neutrals [Rishbeth and Garriott, 1969]. Although the gyrofrequencies are constant throughout the height region from about 80 km to 300 km, the collision frequencies vary rapidly with height, as is shown in Figure 4.9, which leads to a strong height dependence of the magnitude and direction of the current flow.

However, there is a solar cycle variation in solar EUV and UV flux, as well as particle precipitation, which affects the plasma concentrations in the ionosphere, and consequently the collision frequencies between ions and neutrals. The altitude at which absorption in the atmosphere is at a maximum is different for UV and EUV radiation, as is illustrated in Figure 1.4 [Chamberlain, 1978]. This is due to the height distribution of the chemical species that have large absorption cross sections at these wavelengths. EUV absorption reaches a peak at between 130 – 140 km altitude due to absorption by O\textsubscript{2}, N\textsubscript{2} and O, while UV absorption peaks at between 110 – 120 km due to absorption by O\textsubscript{2}. In addition, the penetration into the atmosphere of energetic solar particles depends on the characteristic energy of the velocity distribution which further affects the height dependence of the ionospheric conductivity [Evans et al., 1987, Fuller-Rowell and Evans, 1987]. Figure 4.10 shows the height dependence of the ionization rate for particle precipitation in five sample energy bands where band 2 has a characteristic energy of 0.379 keV and band 22 has a characteristic energy of 17.479 keV.

Although both categories of solar fluxes are solar cycle dependent, EUV flux varies by threefold between solar minimum and solar maximum while UV flux increases by only 30% at solar maximum [Lean, 1987]. Meanwhile the characteristic energies of precipitating electrons and ions are also likely to vary with the solar cycle. In this way there may be a different solar cycle variation of the ionosphere at different heights, so that the induced currents of the lower
part of the ionosphere do not adequately reflect the behaviour of the upper part of the ionosphere which affects the thermospheric winds measured by the FPI. Through this argument it may then be possible to account for the failure of the thermospheric winds above Kiruna to observe a consistent response to geomagnetic activity as measured by ground-based magnetometers in terms of the $K_p$ index.
Figure 4.1a
Figure 4.1b

Data sample taken from:
29 < KP < 50
Average Neutral Winds
KIRUNA 6300A

UT 12 18 00 06 12

M/S

M/S

M/S

M/S

M/S

M/S

Figure 4.1b
Figure 4.1c
Figure 4.2a
Figure 4.2b
Figure 4.2c

188
Figure 4.3

Thermospheric wind vectors observed at College, Alaska for (a) low, (b) moderate and (c) high geomagnetic activity. The location of the magnetic north pole is identified as MNP. [Sica et al., 1986]
Calculated contours of mass flow stream function (in grams per second) for December solstice during (a) solar minimum and (b) solar maximum. [Roble et al., 1977]
Figure 4.5a
Figure 4.5b
Figure 4.6
AVERAGE NEUTRAL WINDS AT KIRUNA 6300A

2 < Kp < 3
data sample taken from: 17/11/81 to 15/4/83
and 24/9/88 to 10/4/90

Figure 4.7
Corrected Observatory Code Geomagnetic Latitude

<table>
<thead>
<tr>
<th>Observatory</th>
<th>Code</th>
<th>Geomagnetic Latitude</th>
</tr>
</thead>
<tbody>
<tr>
<td>Meanook</td>
<td>MEA</td>
<td>62.5°</td>
</tr>
<tr>
<td>Sitska</td>
<td>SIT</td>
<td>60.0°</td>
</tr>
<tr>
<td>Lerwick</td>
<td>LER</td>
<td>58.9°</td>
</tr>
<tr>
<td>Ottawa</td>
<td>ott</td>
<td>58.9°</td>
</tr>
<tr>
<td>Lovel*</td>
<td>Lov</td>
<td>56.5°</td>
</tr>
<tr>
<td>Eskdalemuir</td>
<td>ESK</td>
<td>54.3°</td>
</tr>
<tr>
<td>Belfield</td>
<td>BJF</td>
<td>52.7°</td>
</tr>
<tr>
<td>Fredericburg</td>
<td>FRD</td>
<td>51.8°</td>
</tr>
<tr>
<td>Wingst</td>
<td>WNG</td>
<td>50.9°</td>
</tr>
<tr>
<td>Witteveen</td>
<td>WIT</td>
<td>50.2°</td>
</tr>
<tr>
<td>Hartland</td>
<td>HAD</td>
<td>50.0°</td>
</tr>
<tr>
<td>Eyrewell</td>
<td>EYR</td>
<td>50.2°</td>
</tr>
<tr>
<td>Canberra†</td>
<td>CAN</td>
<td>45.2°</td>
</tr>
</tbody>
</table>

*Observatory added to the network in 1954.
†Observatory added to the network in 1970.

Figure 4.8

The global network of observatories involved in measuring the $K_p$ index in 1988. [Menvielle and Berthelion, 1991]
Electron collision frequency, $\nu_e$, and gyrofrequency, $\omega_e$, for the height region 50—150 km. The ion collision frequency is 10—100 times smaller than $\nu_e$, and the ions gyrofrequency is $30 - 50 \times 10^3$ times smaller than $\omega_e$, depending on the nature of the ion. [Thrane and Piggott, 1966]

Height dependence of the mean ionization rate per unit volume (m$^{-3}$) per unit incident energy flux (1 mW/m$^2$) for precipitating electrons with spectra with peaks in five different energy bands, where band 2 has a characteristic energy of 0.379 keV and band 22 has a characteristic energy of 17.479 keV. [Fuller-Rowell and Evans, 1987]
Simultaneous Observations of Thermospheric Winds from Three Stations in Northern Scandinavia

5.1 Introduction

During the winter of 1987 to 1988 several nights of good quality FPI observations of the 6300Å emission in the aurora and airglow were collected simultaneously from all three sites in Northern Scandinavia. Lack of funds and the difficulty of operating instruments automatically at remote sites have plagued this work so that only the instrument at Kiruna has consistently provided good results since November 1981 because it is routinely monitored by our colleagues at the Swedish Institute of Space Science. Thus although there were a few winters when two of the three FPIs were in operation, there was only one winter when all three were running simultaneously. Even then the necessity of having clear skies at all three sites has cut down the size of the database of simultaneous observations quite considerably. As a result there were only twenty or so nights with significant periods of joint observations from all three sites. Of these, four nights have been picked to serve as interesting case studies of the behaviour of thermospheric winds in the auroral region. The nights chosen illustrate mid-winter thermospheric winds near solar minimum for three levels of geomagnetic activity. Comparison is made with standard model simulations from the UCL-Sheffield three-dimensional, time-dependent global model of the thermosphere/ionosphere (UCL-Sheffield 3DTD) [Fuller-Rowell et al., 1987, Rees et al., 1988, Rees and Fuller-Rowell, 1989] which serves to test both the analysis of the empirical data and the level of sophistication of the model.
5.2 Thermospheric Winds at Auroral Latitudes

Thermospheric winds observed on individual nights are very different from the average winds shown in the previous two chapters. The long-term averages show smooth changes in the wind patterns since the data are pre-sorted so that the conditions controlling the wind flow are fairly consistent throughout the night. The averaged winds display the most basic behaviour of high-latitude thermospheric winds, which is determined by the pressure gradients set up by solar heating and momentum transfer from convecting ions in the auroral oval. Localised, sporadic conditions which cause sharp changes in the winds are smoothed out. In contrast, thermospheric winds observed on any one particular night can be highly variable because geomagnetic activity can change dramatically over a very short period of time. The plasma drift responds immediately to changes in the ionospheric electric field. Collisions between ions and neutrals result in an exchange of energy and momentum so that the neutrals display a delayed and averaged response. The neutral response is that of a low-pass filter to the dynamic response of the ionosphere. The response time of the neutrals depends on the size of the electric field, the plasma density and the ion-neutral collision frequency. Fedder and Banks [1972] calculated that the average response time for the zonal neutral wind in the $F$-region to an electric field of 20 mV/m would be of the order of 1 hour, though this time lag will vary with the plasma density.

The wind components observed to the north, north-east, east, south, west, north-west and vertical, together with the calibration lamp, when available, are shown in Figures 5.2, 5.4, 5.6 and 5.8 for each of the three sites: Longyearbyen, Kilpisjarvi and Kiruna, observed on the nights of the 11th January 1988, 15th December 1987, 21st December 1987 and 14th January 1988 respectively. Longyearbyen is the only one of the three sites able to collect data from around 13UT through to around 10UT on the following day, since at this latitude the sun is below the horizon for almost 24 hours in the middle of winter. The observations from Kiruna and Kilpisjarvi only span the hours between 14UT and 07UT owing to their lower latitude.

The integration time at Kiruna was 60 seconds that winter, while for Kilpisjarvi it was 180 seconds and at Longyearbyen it was 360 seconds. In order to maintain the time resolution of the Longyearbyen instrument to a realistic period the mirror sequence was reduced so that observations were made only to the four cardinal positions plus the zenith and calibration lamp. Unfortunately for the 1987-1988 Longyearbyen data the calibration lamp failed without warning
soon after the operators left the site after setting up the FPI for the winter’s observations. The calibration lamp is used to monitor the stability of the FPI and the baseline is derived from the calibration lamp peak positions, as is described in Chapter 2. In the absence of calibration lamp data the zero Doppler shift baseline for Longyearbyen observations was found by fitting a second order polynomial to the zenith peak positions under the assumption of a uniform wind field with no convergence or divergence of the winds, and consequently no vertical winds. This means that the zero Doppler shift baseline for Longyearbyen data is not as reliable for determining the Doppler shifts as the Kiruna baseline for two main reasons. Firstly, the assumption of zero vertical winds may apply well to the average vertical wind over 24 hours or even over a 12 hour period of nighttime observations due to the physical restrictions imposed by the conservation of mass on the bulk movement of air. However, even allowing for any instrumental drifts (which cannot be done at Longyearbyen without the calibration lamp) there can be considerable variability in the vertical winds over short time periods. Secondly, the intensity of the 6300Å emission observed in the zenith is less than in the other directions since the line-of-sight path through the emission profile is shorter. This results in a poorer signal to noise ratio for the zenith peaks and consequently a larger error associated with the peak position. Despite this, the data presented in this chapter shows an overall compatibility of the Longyearbyen winds with the winds obtained independently from Kiruna and Kilpisjarvi, which thereby increases the confidence in the Longyearbyen results.

The winds observed at Kiruna, Kilpisjarvi and Longyearbyen are presented together also in a two-dimensional grid of neutral wind vectors where the ordinate is latitude and the abscissa is Universal Time. Figure 5.1 shows the wind vectors for every hour from each of the three sites observed on the night of the 11th January 1988. Similarly Figures 5.3, 5.5 and 5.7 show 3-site wind vectors for the nights of the 15th December 1987, 21st December 1987 and 14th January 1988 respectively. Time runs from midday on one day through to midday on the next day. The data plotted are limited to one wind vector every hour for the sake of clarity of presentation. At a given time there are three vectors calculated for each site: to the north, overhead and to the south. When plotting these winds in a latitude versus time graph, account is taken of the curvature of the Earth. This means that the FPI has an observing circle of radius 380 km that subtends an angle of 3.4° of a great circle. The estimate of the radius of
the observing circle depends on the assumption that the peak 6300Å emission occurs at a constant height of 240 km, which may not always be true.

11th January 1988 - geomagnetically quiet night
The monthly average solar flux index for January 1988 was $F_{10.7} = 108$. The magnetometer readings for this night were almost without perturbation throughout the night, which would imply a weak high-latitude electric field resulting in an auroral oval that was contracted polewards of Kiruna. At the start of the observations at midday on the 11th January the thermospheric winds at Longyearbyen were small (approximately 50 m/s) and poleward, as illustrated in Figure 5.1. Then by between 01UT and 02UT the winds observed at all three sites were flowing equatorwards. Finally, at the end of the observing period for Longyearbyen, at 11UT of the 12th January, the winds had rotated round to a westward direction. Thus between 12UT on the 11th January until about 03UT on the 12th January the wind vectors for all three sites were small and rotated in a clockwise fashion throughout the night in a manner that is generally typical of the behaviour expected of a geomagnetically quiet thermosphere, where the prevailing driving force is due to the pressure gradients. However, from 01UT on the 12th January onwards the geomagnetic activity slowly picked up from quiet to moderately active levels, resulting in larger wind magnitudes of the order of 200 m/s for the latter half of the night. By 10UT on the 12th January the winds observed at Longyearbyen were predominantly westwards with a speed of around 400 m/s.

Of note is the difference in the winds observed at Kiruna and Kilpisjarvi. There is a considerable overlap of the fields of view since Kilpisjarvi is merely 1.3° latitude north of Kiruna, and the radius of the field of view is about 3.4° latitude. The wind components shown in Figures 5.2b and c match well in a general way, except that the magnitude of the winds observed from Kilpisjarvi are, on the whole, smaller than those observed from Kiruna. This is not due to a systematic error in the factors used to convert Doppler shifts to wind speeds for the two instruments, as can be seen from a comparison of the winds at individual times which do not show any systematic difference in magnitude. It is accepted that there can be a genuine variability in the winds with a scale much less than the 400 km diameter of the field of view, but these conditions do not warrant such an explanation. Instead the difference is attributed to cloud scatter.

The presence of cloud was monitored at Kiruna with an all-sky camera, but
this facility was not available at the other two sites. The Kiruna all-sky camera indicated cloud overhead from 18UT onwards, but since there were significant Doppler shifts observed in all directions, the cloud cover cannot have been complete and so the data for the whole night are presented here for inspection. (N.B. for the purposes of the long-term averages in Chapters 2 and 3, data occurring in periods that are labelled ‘cloudy’ by the Kiruna all-sky camera are rejected from the database used.) The effect of cloud cover on the apparent Doppler shifts is difficult to determine. A totally overcast sky causes isotropic scattering of light, thereby averaging the Doppler shifts and giving the appearance of zero winds in all directions. However, Abreu et al. [1983, 1985] warned that localised cloud cover and large gradients in the intensities over the field of view may result in anisotropic scatter of light. This can manifest itself in an apparent convergence and divergence in the wind field as explained in Chapter 2.

On examining the intensities of the 6300Å emission it was found that there was indeed a very large gradient in the emission intensity over the region covered by Kiruna and Kilpisjarvi. The emission was concentrated to the north and reached a peak value around midnight. The gradient in the emission was much stronger over the field of view for Kiruna, despite the overlap with Kilpisjarvi. The emission overhead and south of Kiruna was very low, between 10-20% of the peak emission to the north, while the emission above Kilpisjarvi was around 25% of the peak emission observed to the north. Therefore the most plausible explanation of the difference between the Kiruna and Kilpisjarvi winds is that localised and intermittent cloud over Kiruna together with the large intensity gradient have distorted the Doppler shifts observed and consequently the derived winds. Since clouds occur in a height range of 5 – 10 km the cloud cover need only extend over a radius of around 20 km to affect the FPI observations which are made at a zenith angle of 60° collecting photons from within a 1° field of view. Since Kilpisjarvi is nearly 150 km away from Kiruna it is possible that one site can be under cloud while the other is cloud-free.

Theoretically there should be very little vertical movement of air during quiet conditions like this when there are no localised sources of heating, and the wind flow should be fairly uniform. However, although the vertical winds over Kilpisjarvi and Longyearbyen were generally very small throughout the night, at around ±30 m/s, the vertical wind component observed over Kiruna appeared to be very variable, with winds reaching around ±80 m/s. In addition there were large gradients observed in the meridional and zonal winds at Kiruna,
with differences in the wind components observed to the north and south, or to the east and west, of up to 200 m/s. In contrast any gradients observed in the wind field over Kilpisjarvi and Longyearbyen were no more than 50 m/s in size. This would appear to be further evidence of the presence of cloud at Kiruna on this night, thus the wind components and vectors observed at Kilpisjarvi are probably more reliable than those observed at Kiruna.

15th and 21st December 1987 - geomagnetically moderate nights
The prevailing conditions were similar on both these nights. Both were mid-winter nights, near solar minimum (monthly average solar flux index: $F_{10.7} = 102$) with moderate to high geomagnetic activity averaging at $K = 4 - 5$. On both nights the geomagnetic activity built up to a peak in the time period 00 – 03 UT, and then died away. Yet the winds were appreciably different.

Figure 5.3 and Figure 5.5 show the wind vectors for the 15th and 21st December respectively. The winds in the evening sector, between 14UT and 18UT, were more westward on the 21st December, thereby implying stronger ion drag in the dusk cell of the auroral oval, despite $K$ indices that were one unit lower. Yet the magnitudes of the winds in this period were similar, being about 300 m/s. On the 21st December the winds in the midnight sector were about 10% larger than on the 15th December, but the vector bearings were generally the same between 19UT and 02UT.

Then from 03UT onwards the winds observed at Longyearbyen on these two nights became radically different. On the 15th December the wind vectors at Longyearbyen were up to 10 times larger at 400 – 500 m/s, although in the same westwards direction, as the equivalent winds in the morning sector on the 21st December. This discrepancy may be resolved by a closer look at the level of geomagnetic activity during the latter halves of these two nights. Although the 24 hour average $K$ index was almost the same for both nights, as was the sum of the $K$ indices for the night i.e. $K_{\text{sum}} = 38$ on the 15th December and $K_{\text{sum}} = 34$ on the 21st December, the $K$ and $Kp$ indices are semi-logarithmic scales and therefore a difference of 1 unit can represent a significant difference in the geomagnetic activity. Thus between midnight and 12UT on the following day the average value of $K$ for the night of the 15th December was $< K > = 5.25$, while on the 21st December $< K > = 4.0$. Converting the $K$ values to the $Ap$ index, which is a linear scale (reference may be made to Table 3.3) shows that the average geomagnetic activity on the 15th December ($< Ap > = 56$) was
double that on the 21st December ($< Ap > = 27$).

Figures 5.4a, b and c show the line-of-sight winds for the 15th December observed at Longyearbyen, Kilpisjarvi and Kiruna, respectively. Similarly, Figures 5.6a, b and c show the line-of-sight winds for the 21st December. There was a great deal of structure in the variation of the winds through the night on the 15th December, while the variation was quite smooth on the 21st December. This tallies well with the variation observed in the $K$ index. On the 21st December the $K$ index moved smoothly up and down from a minimum value of $K = 3$ to a maximum of $K = 6$ and then back down to $K = 3$. In contrast, on the 15th December the $K$ index varied a great deal in its progress from a minimum of $K = 2$ to a maximum of $K = 7$ and then a decline to $K = 4$. As a consequence, on the 15th December there were several pulses of high speed winds, which appeared consistently at all 3 sites in the observations to the north of the sites. The general wind speeds and times of occurrence of the pulses were the same for Kiruna and Kilpisjarvi, thereby confirming the almost complete overlap of the two fields of view.

The observations to the south of each of the sites showed a smoother variation than to the north. The smoothness of the wind variation to the south of both Kiruna and Kilpisjarvi is likely to indicate that the neutral winds at that latitude were equatorward of the auroral oval and so less affected by the dynamic variability within the auroral oval, which is the dumping site for magnetospheric energy. The smooth wind variation observed to the south of Longyearbyen may imply that that portion of the thermosphere is in the ‘eye’ of the dusk convection cell, and just happened to continue the night in a region that was little disturbed.

There is no reason to suspect that the data analysis procedure could introduce a bias into the line-of-sight wind speeds, causing the winds to the south to appear to be always smoother than winds to the north of the site, because there is no bias in the handling of the data from the azimuthal directions. As confirmation of this impartiality in the data processing Figures 5.8a, b and c show that on the 14th January 1988 there was as much structure in the data observed to the south of each of the sites as to the north.

The general change in the wind speeds throughout the night of the 21st December was much smoother than for the 15th December. However, there was an interesting pulse in the wind field that was highly localised in space as can be seen from a comparison of the line-of-sight winds observed to the east and
west of Kiruna and Kilpisjarvi. The general trend of the zonal winds in this part of the night showed an increase in magnitude to a maximum westwards zonal wind between 16 – 17 UT, followed by a gradual drop in magnitude. The pulse caused a rapid drop in wind speed which occurred between 17 – 19 UT and was very clearly apparent in the Kilpisjarvi data. A hint of the pulse may be seen to the east of Kiruna between 17 – 18 UT, but not to any certainty to the west. The zonal wind was observed to reach a minimum westward value at around 17:45UT to the east of both Kiruna and Kilpisjarvi. The pulse was observed to the east of Kilpisjarvi one hour before it was observed to the west. Spatially this might represent a highly localised perturbation of the wind field lying along a line of longitude, which only came into view of the FPI, first in the observation to the east and then in the observation to the west, as the site passed under it. The perturbation petered out at the latitude of Kiruna, which is 150 km south of Kilpisjarvi.

On the 21st December, there was a general increase in the equatorward meridional wind to the north of Kiruna and Kilpisjarvi. However, to the south there was a drop of up to 100 m/s in the magnitude of the southward meridional wind between 18 – 20 UT, which appears to match the drop in the westward zonal winds observed at Kilpisjarvi. Such a large change in the magnitude of the zonal winds was not observed at Kiruna, which is about 130 km south of Kilpisjarvi. However, the thermosphere observed to the south of Kiruna is about 400 km away. Therefore this might indicate that the pulse observed to the south could have been a continuation of the perturbation observed at the latitude of Kilpisjarvi in the zonal winds, with a break in its influence at the latitude of Kiruna.

Associated with the pulse observed in the horizontal winds was an oscillation in the vertical winds at Kilpisjarvi between 14 – 20 UT. This oscillation appeared to a smaller extent in the vertical winds observed over Kiruna, but the amplitude was of a similar magnitude to the wind errors and so cannot be resolved definitely.

A short lull in the meridional winds between 00 – 01 UT was observed to the north of both Kiruna and Kilpisjarvi. Then from 03 – 06 UT the meridional wind in the region between latitudes 65° – 75° turned poleward, as is shown in Figure 5.6 in the meridional winds observed to the north of Kiruna and Kilpisjarvi, and to the south of Longyearbyen. During this same period there was a convergence in the wind field over Longyearbyen, accompanied by downward
vertical wind component observed in the zenith.

14th January 1988 - geomagnetically active night

Figure 5.7 shows the high-latitude thermosphere under major storm conditions as observed by the three sites. $K_{sum} = 51$ and the average value of the $K$ index was 6. The geomagnetic activity increased dramatically from a minimum value of $K = 2$ after midday, through to a maximum of $K = 9$ in the 3-hour period before midnight, then dropped to $K = 5$ in the 3-hour period before midday of the next day. Such highly disturbed conditions are rarely observed.

Under moderately active geomagnetic conditions Kiruna lies on the outer edge of the auroral oval where the ionospheric convection in the dusk cell is sunward. In which case the neutral wind flow is south-westward to westward due to ion drag, as was demonstrated by the winds observed on the 15th and 21st December 1987 (Figures 5.3 and 5.5). However, between 17UT and 21UT on the 14th January 1988 the wind vectors from all three sites were large, with an average magnitude of 300 m/s, and pointed in a south-east direction. The observation of south-eastward winds at Kiruna and Kilpisjarvi would indicate that the auroral oval had expanded far equatorward of Kiruna, so that Kiruna lay within the polar cap where the ionospheric convection is anti-sunward. EISCAT radar observations of plasma flow were then consulted in order to confirm the FPI observations. Figure 5.9 shows the plasma flows observed between latitudes 71°N to 75°N during the period 15UT to 21UT. Strong south-eastward plasma flows were measured until nearly 17UT which agrees with the observations at Longyearbyen and also with the observations to the north of Kiruna and Kilpisjarvi, since the range of the radar observations extends far enough south to overlap with the observations to the north of Kiruna and Kilpisjarvi.

The vertical winds observed over Kiruna and Kilpisjarvi were always below 50 m/s (Figure 5.8). There was a change in the direction of the vertical flow from a downward wind before 16UT to a small upward wind which lasted until 19UT. This feature appeared in both the Kiruna and Kilpisjarvi data, and its observation at both sites independently confirms that it was real and not an artifact of the individual data sets.

In contrast, the vertical winds observed over Longyearbyen appear to have been highly variable, having reached unusually high values of over 150 m/s. However, in this case the determination of the zero Doppler shift baseline may have been less reliable. The assumption of a uniform wind field, which is used
to determine the baseline, loses validity during highly unstable conditions such as those prevalent on the night of the 14th January 1988. Therefore the large vertical winds calculated may be either due to a poor choice of baseline, created by forcing the average vertical wind over the whole night to be zero, or to a genuine upheaval in the thermosphere. Such a problem can only be resolved by an independent measurement of the thermosphere or ionosphere. True vertical winds of around 100 m/s would be accompanied by a large increase in temperature of 200 – 300 K [Hernandez et al., 1982, Abreu et al., 1983]. Temperature measurements will not be possible with the UCL FPIs until a stabilised single-mode, mode-locked laser can be bought and installed to provide a reference δ function line peak from which the FPI instrument profile can be derived.

After 21UT the winds suddenly decreased in magnitude at all three sites despite the continuing increase in geomagnetic activity to a maximum possible value of $Kp = 9$. It is possible that the apparent drop in wind speeds may have been due to cloud scatter of the airglow and auroral emissions since the all-sky camera at Kiruna indicated cloud overhead after 22UT. However, cloud cover need not have extended as far as Kilpisjarvi, let alone Longyearbyen. Further, after about 22UT the winds observed at Kilpisjarvi and Longyearbyen reached 200 – 300 m/s, which would imply that if any cloud had been present the consequent light scatter had merely subdued the apparent Doppler shifts rather than wiped them out. The sudden drop in wind speed might be accounted for by the behaviour of the ionosphere. EISCAT radar data show a rapid double reversal of the plasma flows that occurred just after 21UT. The flow direction changed first to the west and then to the east between 21:30UT and 22UT, and then eventually settled down to a south-eastward flow, as is corroborated by the Longyearbyen and Kilpisjarvi neutral wind data.

In the morning sector, between 06–09 UT, thermospheric winds observed at Longyearbyen were westwards turning to north-westwards, while the winds observed at Kilpisjarvi were eastwards turning to northwards. The EISCAT radar observations between 08-09UT (Figure 5.10) confirm the northward plasma flows in the latitude region 71°N to 75°N. This might imply that the eye of the dawn cell was in the region between Longyearbyen and Kilpisjarvi, at a latitude of about 75°N.
5.3 Comparing Model Data With FPI Data

The input parameters used in a model to simulate specific solar, geomagnetic and seasonal conditions appropriate for comparison with empirical data must be chosen very carefully. The parameters will be interrelated in various complex ways, and similar wind fields may be produced by different combinations of these parameters. An impartial choice of model is required, which may highlight better the similarities and discrepancies between model and observation. However, with experimental results that are limited in quantity, location and time, as is the usual case, there is an element of personal judgement needed to set up a matching model simulation. Thus a simulation can be restricted by both its level of sophistication and the experience of the modeller, both of which, of course, are determined by the existing theoretical and empirical work.

If data could be classified into unique groups of geophysical conditions it may then be easier to attribute behaviour to a particular cause. The major drawback to this being that not all empirical parameters are known, and, in particular, that within the limits of error a data set may match with more than one simulation. The latter problem may be reduced when there are several independent observations which can be combined to create the boundary conditions within which the model must fit. As such, the 3-site wind data plots of Figures 5.1, 5.3, 5.5 and 5.7 allow a more rigid comparison of results with 3-D models since there are independently observed wind vectors over a wide range of latitude and local time to be matched.

The data presented here are compared with the UCL three-dimensional, time-dependent thermospheric model which simulates the wind, temperature, density and composition of the neutral atmosphere by solving the non-linear equations of momentum, energy and continuity numerically [Fuller-Rowell and Rees, 1980, 1983, Rees and Fuller-Rowell, 1988]. The model works in a spherical coordinate system, with the thermosphere divided up by geographic latitude, longitude and pressure. The resolution is 2° in latitude and 18° in longitude. The atmosphere is divided into 15 levels in the vertical direction according to a logarithmic scale of pressure. Each level is equivalent to one scale height in depth. The lower boundary is fixed at an altitude of 80 km, but the upper boundary is temperature dependent, and can vary from 300 km during geomagnetically quiet periods at solar minimum, to over 700 km during disturbed periods at solar maximum. Each grid point rotates with the Earth, making this a non-inertial frame of reference.
The strong coupling between the neutral thermospheric particles and the ionospheric plasma was not sufficiently modelled by using the ionospheric model by Chiu [1975] at high-latitudes. Enhancement of [N₂]/[O] ratio due to polar upwellings in the summer polar cap, the large anti-sunward winds over the polar cap and the large sunward winds in the dusk cell of the auroral oval all required an increase in the polar plasma density [Quegan et al., 1982, Rees et al., 1983, 1985]. In order to solve this problem a collaboration between UCL and the University of Sheffield has led to a self-consistent fully-coupled numerical high-latitude ionospheric and global thermospheric model [Fuller-Rowell et al., 1987].

The neutral composition is calculated within the thermospheric model, except for the density of hydrogen, which is taken from the MSIS model [Hedin, 1987]. While the effects of geomagnetic activity, particle precipitation and dependence on the orientation of the IMF can be accounted for by using one of a selection of electric fields from Heppner and Maynard [1987], Foster [1984] and Heelis [1986].

Figures 5.11 and 5.12 show the neutral wind vectors derived from the UCL-Sheffield 3D model between latitudes 60°N and 82°N at pressure level 12, which is in the region of the maximum 6300Å emission. These simulations are chosen from a database of simulations of steady state conditions. Both simulations are for solar minimum conditions, with \( F_{10.7} = 95 \). Figure 5.11 represents geomagnetically quiet conditions and Figure 5.12 represents moderately active conditions. Extension of this analysis would be to customise time-dependent simulations to follow closely the prevailing conditions for each of these nights. However, such an analysis is beyond the scope of this thesis.

11th January 1988 - geomagnetically quiet night
As explained in section 5.2, the winds derived from the observations at Kilpisjarvi are probably more reliable than those derived from the observations at Kiruna on this night owing to possible cloud scatter. Consequently the Longyearbyen and Kilpisjarvi winds are more likely to be representative of the thermosphere in the latitude range between 64°N – 82°N.

Before midnight the vector winds portrayed in Figure 5.1 are very similar to those of the model simulation called \( LJ3 \) (Figure 5.11), which simulates very quiet geomagnetic conditions. The \( LJ3 \) simulation uses the electric field derived from Millstone Hill radar observations [Foster et al., 1986a]. The activity level chosen represents an average geomagnetic activity of \( Kp = 1^+ \), which is very
similar to that of the first half of the night of the 11th January. The average solar flux index used in the simulation is 13 units less than the monthly average for January 1988 ($F_{10.7} = 108$), but this is unlikely to have a significant effect.

The first half of the night was geomagnetically quiet and then the geomagnetic activity started to increase after midnight. The $Kp$ index rose steadily to a value of $Kp = 5^-$ by midday of the 12th January, so that the average value of $Kp$ for the second half of the night was around $2^+$, which is a moderately active level. Consequently the magnitude of the winds observed at Longyearbyen in the second half of the night were between $150 - 200$ m/s, which is at least 50% larger than the model prediction.

None of the standard coupled model simulations for geomagnetic activity levels of $Kp = 2^+$ and low solar activity match the FPI winds observed after midnight. For these input parameters the simulations all predict westward winds from 05UT until around 11UT at the latitude of Longyearbyen while the FPI observations showed strong south-west flow during this period of time.

15th and 21st December 1987 - geomagnetically moderate nights

Both of these nights have similar prevailing conditions. However, a comparison of the winds illustrates how though there can be an overall similarity, there may still be significant differences that need to be accounted for. The model simulation $KA3$, shown in Figure 5.12, is the closest match to both these nights. The model simulation conditions are $F_{10.7} = 95$, $Kp = 3$ and $B_y$-ve. The change in the Longyearbyen winds from westward to eastward at around 15UT is predicted by the simulation. The magnitude of the simulated winds at this time and latitude is closer to the Longyearbyen winds observed on the 15th December; the Longyearbyen winds observed on the 21st December are less than half the size of the simulated winds. Even at the latitude of Kiruna and Kilpisjärvi the simulation is a better approximation to the winds observed on the 15th December, although on both nights the winds changed from a westward to eastward direction at around 21UT, in agreement with the simulation. Between 06-12UT the predicted westward flow at the latitude of Longyearbyen is intermediate in magnitude to the winds of the 15th and 21st December, though in agreement with the observed vector bearings.

The comparison of the vector wind plots of Figures 5.3 and 5.5 with the simulated winds shown in Figure 5.12 shows up an inadequacy in the procedure for deriving vectors to the north and south of each site. The wind fields created
by this method do not allow for a smooth transition between the vectors south of Longyearbyen and the vectors north of Kiruna and Kilpisjarvi before 21UT. For one thing it has been assumed in these plots that the height of the 6300Å emission remains constant and is unaffected by the ambient conditions, in which case the radius of the field of view remains constant. This is probably not true, as is investigated in Chapter 3, but there has been insufficient analysis of the dependence of the height of the emission layer on the prevailing geophysical conditions to justify a more sophisticated approach.

Perhaps most importantly the method of combining the line-of-sight observations from all directions to derive a wind vector should be examined. With this method there is little difference in the three vectors, which is consistent with the underlying assumption of a uniform wind field. However, with an average distance of around 800 km separating the observations to the north and south of each site it is unlikely that the wind vectors to the north and south are the same, especially for high latitude sites within the auroral oval where the gradients in the electric field can be very steep and variable. A comparison of the line-of-sight wind components does in fact show such differences.

14th January 1988 - geomagnetically active night

The behaviour of the thermosphere and ionosphere on this night is extremely interesting. Under moderate conditions of geomagnetic activity the outermost boundary of the polar cap lies at a geomagnetic latitude of about 73°N (Figure 3.27) [Feldstein and Starkov, 1967], which places Longyearbyen within the polar cap but Kiruna and Kilpisjarvi outside it. However, on the 14th January 1988, large anti-sunward thermospheric winds were observed during the evening at all three sites, and at the same time large anti-sunward plasma drifts were observed with the EISCAT radar. Such behaviour is typical of the thermosphere and ionosphere within the polar cap.

A polar cap that extends to latitudes lower than that of Kiruna is very rarely observed. As a consequence there are no empirical electric field models at present that can simulate the conditions on this night. Not even the highest level of the electric fields derived from Millstone Hill radar observations by Foster et al. [1986] has a polar cap that extends equatorward of geomagnetic latitude 70°N. However, Rich and Maynard [1989] have modelled the convection electric field of Heppner and Maynard [1987] using simple analytic functions. The original Heppner and Maynard models were developed for average geomagnetic
conditions of $K_p = 3^+$ and a southward IMF. Since then Rich and Maynard have digitized the models and fitted polynomial functions to the digitized patterns. The resulting spherical harmonic representations will now allow the patterns to be manipulated to simulate geomagnetic conditions other than average conditions. It may then be possible to simulate the conditions experienced by the thermosphere on the night of the 14th January 1988 by use of this harmonic model of the electric field.
Figure 5.2

Line-of-sight neutral winds observed on the 11th January 1988 at (a) Longyearbyen, (b) Kilpisjarvi and (c) Kiruna.
Figure 5.3
Figure 5.4

Line-of-sight neutral winds observed on the 15th December 1987 at (a) Longyearbyen, (b) Kilpisjarvi and (c) Kiruna.
Figure 5.5
Figure 5.6

Line-of-sight neutral winds observed on the 21st December 1987 at (a) Longyearbyen, (b) Kilpisjarvi and (c) Kiruna.
Figure 5.7
Figure 5.8

Line-of-sight neutral winds observed on the 14th January 1988 at (a) Longyearbyen, (b) Kilpisjarvi and (c) Kiruna.
Figure 5.9

Plasma velocities observed with the EISCAT radar for the period 15-21UT on the 14th January 1988.

Figure 5.10

Plasma velocities observed with the EISCAT radar for the period 03-09UT on the 15th January 1988.
UCL-Sheffield model simulation *LJ3* of thermospheric winds observed through 24 hours at a longitude of 18°E, over a latitude range of 60°N to 82°N for geomagnetically quiet conditions at solar minimum.

UCL-Sheffield model simulation *KA3* of thermospheric winds observed through 24 hours at a longitude of 18°E, over a latitude range of 60°N to 82°N for geomagnetically moderate conditions at solar minimum.
6.1 Introduction

The physics of the middle atmosphere is quite different from that of the upper atmosphere in that the mesosphere and lower thermosphere appear to be primarily governed by tidal motion generated in the troposphere and stratosphere. The eddy dissipation of propagating waves and tides is a source of heating that is often greater than that due to solar UV and EUV heating [Hines, 1965].

There has been much theoretical work on the behaviour of the middle atmosphere but relatively little experimental work owing to the difficulties of observing this region. The altitudes involved are too low for in situ observation by satellite and too high for balloons and rocket-sondes. As a result, most of the systematic data on tidal winds in the mesosphere have been provided by remote sensing. For example, the height region between 80 km and 110 km has been monitored for many years since the late 1950's by radar tracking of meteor trajectories, and has consequently been labelled the 'meteor region' [Greenhow and Neufeld, 1959, Tsuda et al., 1980, Bernard, 1981, Ahmed and Roper, 1983].

The Atmospheric Tides Middle Atmosphere Program (ATMAP) has organised several campaigns and workshops using data from this source and from partial reflection radar to advance the understanding of the tides of the middle atmosphere [Forbes, 1987a, 1985, 1986, 1987b]. Incoherent scatter radar is a more modern major source of middle atmosphere data [Vasseur, 1969, Johnson et al., 1987, Williams and Virdi, 1989]. Chemical releases from rocket probes [Manring et al., 1961, Groves, 1963, Pereira et al., 1980] have also provided their share of data, but they are an expensive experimental procedure, giving data that is limited in time and horizontal location although collecting important observations of the vertical structure. Now recent advances in imaging photon detectors have allowed direct measurements of the winds in the mesosphere using the Doppler shifts of the infra-red emission of the hydroxyl molecule [Rees et al., 1990]. The results of the work so far are presented in this chapter.
6.2 Tidal and Gravity Waves in the Middle Atmosphere

6.2.1 Tides

Tides are a global phenomena driven mainly by solar heating. The diurnal heating cycle is a rich source of harmonics because it is more like a square wave function than a sinusoid. The structure of atmospheric tides has been modelled using classical tidal theory which involves the solution of linearized equations of continuity, momentum, and energy modified by time dependent solar heating [Siebert, 1961, Chapman and Lindzen, 1970]. The basic equation is Laplace's tidal equation [Lindzen, 1967], which is separable in terms of height and latitude dependence. The classical tidal theory presumes the absence of mean winds and meridional temperature gradients and neglects dissipation. The eigen function solutions are commonly called 'Hough functions'. These Hough functions \( \theta_n^m \) describe the latitudinal structure, and the corresponding eigenvalues \( h_n^m \) (also called the equivalent depths) give the vertical structure for a given background temperature distribution. They are labelled with the numbers \((m, n)\) according to the notation of Flattery [1967], where \( m \) is the zonal wave number and \( n \) is the meridional wave number. Positive or negative values of \( n \) denote gravitational or rotational modes respectively. In other words, \( m \) gives the number of cycles per day, and the number of nodes between the poles is given by \((m - n)\), with the exception of the nodes at the poles. For example a \((2, 4)\) mode has a period of 12 hours and 2 nodes \((m - n = -2)\). Solar heating is greatest at or near the equator according to the season, hence modes with an antinode at the equator, rather than a node, are preferentially excited [Hargreaves, 1979]. The most important semidiurnal Hough functions are illustrated in Figure 6.2.

Classical tidal theory is, however, only a gross approximation. In a real atmosphere the assumptions used can break down. Analytic solution of the tidal equations of motion is then no longer possible because the equations are no longer separable, and so several ‘non-classical’ models have been created to cope with this by solving the equations numerically. The consequences of the relaxation of the assumptions on the models are as follows:

(i) Above about 100 km, molecular viscosity, thermal conductivity and ion drag become increasingly important. The inclusion of dissipative terms results in height dependent Hough functions called Hough mode extensions [Forbes and Hagan, 1982]. Dissipation slows down the exponential growth of propagating
modes so that the amplitudes and phases of winds and temperatures reach constant values asymptotically at above 150 – 200 km, or even decrease [Hong and Lindzen, 1976, Forbes, 1982a, 1982b, 1987].

(ii) Lindzen and Hong [1974] showed that the inclusion of mean winds and temperature gradients results in the generation of higher order modes that have not been directly forced but achieved through 'mode coupling'. In fact, it has been suggested that this may be the dominant mechanism for the generation of higher order modes [Walterscheid et al., 1980, Aso et al., 1981, Fesen et al., 1986].

The diurnal and semidiurnal modes are the most important modes. The diurnal modes are thermally driven [Green, 1965], arising mainly from the absorption of solar radiation by water vapour in the troposphere at around 9 – 12 km [Siebert, 1961] with a further contribution from ozone in the stratosphere at around 30 km [Butler and Small, 1963, Leovy, 1964]. At the equator the upward flux of energy due to the diurnal modes has been calculated by Lindzen [1967] to be 2.5 times the downward solar flux for the spectral bands absorbed in the thermosphere. However, the latitudinal variation of the diurnal tide energy flux (Figure 6.3) is such that an average over all latitudes shows that the upward energy flux is less than half the downward flux of solar radiation. The semidiurnal modes are also thermally driven by absorption of solar flux, though the dominant contribution due to absorption is from stratospheric ozone with a smaller contribution from water vapour in the troposphere.

The diurnal modes are more strongly driven than the semidiurnal modes, yet the observed diurnal pressure oscillations are smaller and less regular. For instance, Lindzen [1967] pointed out that the semidiurnal oscillation in the Tropics is about 1 mb while the diurnal oscillation is about 0.5 mb. This was explained qualitatively by Butler and Small [1963] who argued that the diurnal tides are associated with small eigenvalues, and therefore short wavelengths. This would result in their annihilation through self-interference. As a result the stratosphere and mesosphere are dominated by the semidiurnal tide and the influence of the diurnal tides tend to be confined to the region of excitation [Lindzen, 1966].

Most studies ignore the diurnal modes because they are difficult to model for two particular reasons: the complexities of turbulent diffusion cannot be overlooked for this mode and the shorter vertical wavelengths require a higher resolution than most contemporary models can give [Forbes, 1987]. Increas-
ing the grid resolution can be uneconomical and so Forbes and Hagan [1979] overcame the latter problem by using an ‘Equivalent Gravity Wave f-Plane Formalism’ to simulate the (1,1) mode. It was found that the (1,1) mode appeared to be dominant below a height of 100 km at low latitudes where the meridional temperature gradients are weak and the mean wind flow is small compared with the phase velocity. Under these conditions phase coupling is small and the assumptions of classical tidal theory hold well. Further, the (1,1) mode was found to have an exponential growth in amplitude and a phase progression, achieving a peak amplitude between 105 km and 110 km. Then above 110 km the diurnal tide is strongly damped due to the dominance of molecular diffusion in the upper thermosphere and the absence of tidal heat and momentum sources at these heights [Forbes, 1982a,b, 1987, Vial, 1986].

In the Forbes model the diurnal tidal oscillations between 90 km and 150 km are generated equally by upwards propagating waves and by in-situ solar EUV excitation. In propagating modes the phase varies with height so that a vertical wavelength, \( \lambda_z \), exists, defined by the distance over which the phase changes by 2\( \pi \). If \( \lambda_z \) is less than or equal to the depth of the atmosphere where most of the heating occurs, then the wave suffers from self-interference and is therefore weak, thus confirming the argument of Butler and Small [1963]. For the diurnal tide \( \lambda_z \leq 30 \text{ km} \) [Forbes, 1987].

The vertical wavelength, \( \lambda_z \), may be calculated in terms of the scale height of the atmosphere, \( H \), at a height, \( z \), and the adiabatic index, \( \gamma \), which is the ratio of the specific heats at constant pressure and constant volume, and the equivalent depth of the tidal mode, \( h_{mn} \), as given in equation 6.1 [Wilkes, 1949, Chapman and Lindzen, 1970].

\[
\lambda_{z,mn} = \frac{2\pi H}{\sqrt{\left( \frac{H}{h_{mn}} \left( \frac{\gamma-1}{\gamma} + \frac{dH}{dz} - \frac{1}{4} \right) \right)}} \quad (6.1)
\]

As an example Table 6.1 gives the modes and the associated equivalent depths and vertical wavelengths calculated for semidiurnal tides at a height of 110 km in terms of the MSIS83 model by Virdi et al. [1990]. The modes used in this example are the only propagating modes to reach the thermosphere before being dispersed.

The semidiurnal (2,2) and (2,3) modes have longer vertical wavelengths than the other modes and are therefore the most penetrating, capable of affecting
the thermosphere above 150 km. However, despite the importance of these two modes they are not sufficient to model the lower thermosphere. Theoretical calculations by Vial [1986] suggested that the seasonal variation in the latitudinal structure of the wind amplitudes and phases is due to the antisymmetric modes (2,3) and (2,5). This was challenged by Forbes [1987] who doubted that the (2,5) mode would be so dominant in the real atmosphere. Instead the latitudinal variation in the vertical structure of the diurnal tidal winds and temperatures could result from superposition of the diurnal (1,1) mode and an evanescent or 'trapped' mode of the form (1,-2). An evanescent mode is a tidal oscillation that maintains the same phase at all heights and does not propagate. An evanescent mode is apparent at high latitudes in the Forbes model [Forbes, 1987] at heights below 100 km, where there is no amplitude growth or phase progression with height.

As waves propagate vertically upwards, experiencing no damping, the amplitudes become larger in order to conserve the energy density as the particle density decreases i.e. the amplitude of the air velocity, \( U \), varies proportionately with \( \sqrt{\rho} \), where \( \rho \) is the air density, in order to maintain a constant energy density of \( (\rho U^2)/2 \). Eventually the amplitudes become too large and unstable and the waves break. Instability results when the wave amplitudes are approximately equal to the local speed of sound or when the wave phase velocity approximates to the background wind speed [Lindzen, 1970]. This is the general behaviour of waves generated in the troposphere that propagate upwards to break in the mesosphere and lower thermosphere releasing their energy as heat. The horizontal structure of semidiurnal modes with a vertical wavelength greater than 40 km is not significantly affected by dissipation below about 100—110 km [Forbes and Hagan, 1982]. However, damping due to eddy or molecular dissipation may occur between 90—100 km, yet without changing the horizontal shape [Lindzen, 1971].

In summary, the influence of the semidiurnal tides on the middle atmosphere is characterised by at least four semidiurnal modes: (2,2), (2,3), (2,4) and (2,5), whose relative phases and amplitudes vary with height and latitude. Below 60 km the semidiurnal tide is dominated by the (2,2) mode, which is preferentially excited by \( \text{H}_2\text{O} \) and \( \text{O}_3 \) [Forbes and Garrett, 1978]. Between 50 km and 70 km the exponential growth of the (2,2) mode is checked because the variation in the background temperature causes this tide to become evanescent and also because of mode coupling into the higher order modes: (2,3), (2,4) and
(2,5), and thereby losing energy [Lindzen and Hong, 1974]. However, the higher order modes maintain an exponential growth with height so that between 70 km and 90 km the (2,2) mode is joined in equal quantities by the (2,3) and (2,4) modes to make up the semidiurnal tide [Forbes, 1982b]. Then in the region 90 – 120 km the semidiurnal tide is taken over by the (2,4) mode with some contributions from the (2,2) and (2,5) modes. Above 120 km, owing to their short wavelengths, the (2,4) and (2,5) modes are preferentially damped due to the exponential increase in molecular dissipation. From 140 km upwards the (2,2) mode is once again dominant, with additional contributions from the (2,4) and (2,5) modes. At equinox the asymmetric (2,3) and (2,5) modes are absent.

**Table 6.1**

<table>
<thead>
<tr>
<th>mode</th>
<th>equivalent depth</th>
<th>vertical wavelength</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$h_{mn}$ / km</td>
<td>$\lambda_z$ / km</td>
</tr>
<tr>
<td>(2,2)</td>
<td>7.85</td>
<td>96</td>
</tr>
<tr>
<td>(2,4)</td>
<td>2.11</td>
<td>38</td>
</tr>
<tr>
<td>(2,3)</td>
<td>3.67</td>
<td>53</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(2,5)</td>
<td>1.37</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
6.2.2 Acoustic-Gravity Waves

The propagation of acoustic-gravity waves in a dissipationless atmosphere can be described by a combination of the first law of thermodynamics, namely the conservation of energy, together with the continuity equation (equation 1.7b) and a modified version of the momentum balance equation (equation 1.6). The two main modifications to the momentum balance equation are the absences of the ion-neutral collision term and the Coriolis term. Ion-neutral interactions are neglected because the charged component of the mesosphere is negligible. Coriolis effects are neglected because the horizontal scale of the wave is small enough to make this term negligible with respect to the other terms.

The momentum balance equation contains both a term for compressional forces \(-\nabla P\) and a term for buoyancy forces \(g\). The difference between the pure wave forms is that the restoring force of the pure acoustic wave is due to the change in pressure when a fluid is compressed while the restoring force for a pure gravity wave is the force of gravity on a displaced body of fluid due to buoyancy. At high frequencies and short scales the compressional term dominates, consequently the waves generated are acoustic. At low frequencies buoyancy becomes important in balancing the vertical pressure gradient and gravity waves are generated. The corresponding characteristic frequencies are the acoustic resonance frequency, \(\omega_a\), and \(\omega_b\) which is the buoyancy or Brunt-Vaisala resonance frequency. For an isothermal atmosphere the acoustic resonance frequency and Brunt-Vaisala resonance frequency are given by equations 6.2 and 6.3 respectively [Houghton, 1977]:

\[
\omega_a = \frac{1}{2} \left( \frac{\gamma g}{H} \right)^{\frac{1}{2}} = \frac{c}{2H}
\]

\[
\omega_b^2 = \frac{(\gamma - 1)g}{\gamma H}
\]

Acoustic-gravity waves travel at speeds close to the speed of sound, with wavelengths that range from tens to hundreds of kilometres long and periods ranging from minutes to tens of minutes. They therefore have smaller periods than the tidal waves.

Acoustic-gravity waves are localised, set up by a boundary or discontinuity in the atmosphere. They can be caused by disturbances set up by, for example,
thunderstorms, volcanic eruptions, wind flowing over mountains and even the passage of the solar terminator over the surface of the Earth which thereby alters the quantity of solar radiation incident on the Earth. There is also a correlation between large-scale waves and the occurrence of geomagnetic storms [Francis, 1975]. A nuclear explosion is an artificial source capable of setting up an acoustic-gravity wave. As an example, the largest of the nuclear test explosions above ground was the explosion of a 58 megaton Russian nuclear bomb 4 km above Novaya Zemlya in 1961. This generated an acoustic pulse that circled the Earth more than three times and was observable at ground level for several days [Donn and Ewing, 1961, Wexler and Hass, 1962].

Gravity wave stress is a drag force occurring between 90 — 100 km which contributes to the reduction of diurnal mean winds to small values [Lindzen, 1981, Holton, 1982]. Observation of the waves is difficult and is complicated by the atmospheric temperature structure since the large scale of the disturbances makes it hard to monitor the entire wind field [Francis, 1975]. The main evidence for their presence comes indirectly from observations of the effects on the ionosphere, with little direct evidence from the neutral atmosphere. Gravity waves in the neutral atmosphere are observed by the distortion of meteor trails and vapour trails [Manson et al., 1985]. Wave-like structures appearing in noctilucent clouds at 80 — 90 km are also attributed to gravity waves disturbing the neutral air [Hargreaves, 1979].

Gravity waves passing through the ionosphere can be observed as waves of varying electron density called Travelling Ionospheric Disturbances (TID). The ionospheric observations are made by ionosondes, HF Doppler radar and incoherent scatter radar. TIDs have been observed for many years but were identified as caused by gravity waves only in 1959 by C.O.Hines [Hines, 1959]. Acceptance of this source was a long time coming, since the ubiquity of gravity waves was much doubted at that time, especially by meteorologists whose numerical models used the Boussinesq approximation which precluded the existence of waves with an exponential growth with height, which is one of the characteristics of gravity waves [Hines, 1989].

TIDs were separated by Georges [1968] into large-scale and medium-scale TIDs according to the scale of the disturbance. Large-scale TIDs have a wavelength of several hundred kilometres, with periods of 30 minutes to 3 hours. Their speeds can be much greater than the speed of sound at 300 m/s, ranging from 400 — 1000 m/s. Meanwhile medium-scale TIDs have a wavelength of sev-
eral hundred kilometres, periods of 15 minutes to 1 hour and travel at speeds of 100 – 250 m/s, which is much less than the speed of sound. Both modes have been observed to travel for thousands of kilometres without attenuation [Martyn, 1950, Munro, 1950], but the large scale modes are guided while the medium-scale are freely propagating.

The large-scale TIDs are known as guided modes because they appear to be channeled by the steep temperature gradients at the base of the thermosphere and the Earth’s surface. There is a strong correlation with severe magnetic storms [Valverde, 1958, Wright, 1961, Chan and Villard, 1962] and they travel from polar regions towards the equator. One to one relationships of mid-latitude TIDs with individual polar substorms have been identified, an example from Davis [1971] is shown in Figure 6.4. Here magnetometer data from four auroral stations: Sodankyla, Leirvogur, Great Whale and College, show the onset of a geomagnetic storm at around 17:50 PST while electron content data collected by mid-latitude stations at Edmonton, Fort Collins, Stanford and Clark Lake, show two pulses interpreted by Davis as large-scale TIDs generated by the polar substorm.

Medium-scale TIDs are not so easily attributable to any particular source. They occur more frequently than large-scale TIDs as they appear to be excited by any source at any altitude, including auroral sources. There is a strong seasonal dependence of their spectrum [Georges, 1968].
6.2.3 Influence of Tidal and Gravity Waves on the Upper Atmosphere

An understanding of tidal behaviour has relevance beyond the frontiers of the middle atmosphere. The energy propagated upwards from the troposphere and stratosphere by tides is an important contribution to the energy budget of the lower thermosphere [Hines, 1965]. Further, gravity waves can greatly modify the ionized and neutral structures in the upper thermosphere, as evidenced by phenomena such as the collapse of the ionosphere over Arecibo [Crary and Forbes, 1986] and the midnight temperature anomaly in the equatorial region [Herrero et al., 1983]. The significance of the upward propagating tides on the upper thermosphere has been demonstrated theoretically using the UCL Thermospheric General Circulation Model (TGCM) [Forbes, 1987a]. Figure 6.4a shows a simulation of equinox solar minimum conditions for an altitude range of between 80 km and 330 km without semidiurnal forcing at the lower boundary. This produces a fairly simple pattern of zonal wind contours, with little variation in the wind pattern below around 180 km. However, Figure 6.4b shows that the zonal wind contours become severely distorted on addition of the two main tidal modes: (2,2) and (2,4).

Since experimental observations of the upper atmosphere are far more abundant, the consequences of tides at these altitudes, as predicted by the models, can be tested more extensively than at mesospheric altitudes. For example Parish et al. [1990] were able to demonstrate that appropriate addition of the (2,2), (2,3) and (2,4) semidiurnal tidal modes to the lower boundary of the UCL-Sheffield ionosphere-thermosphere model improved the fit to EISCAT radar data from the lower thermosphere. However, the comparison of data with the model also revealed that the assumptions of classical tidal theory were too simplistic. The middle atmosphere models at present contain relatively large uncertainties with regard to levels of heating, the specification of mean winds and the dissipation of energy into the upper atmosphere. They are not capable of modelling case studies of specific geophysical and seasonal conditions as are the upper and lower atmosphere models. It is therefore evident that new observational data is required for further progress.
6.3 Initial Results from the Utah FPI

In August 1988 a new FPI, developed and built at UCL, was installed at the Centre for Atmospheric and Space Science (CASS), part of Utah State University (USU) in America. The main aim of the UCL collaboration with CASS was to make continuous, long-term observations of the mesosphere on an automatic basis. This has been achieved by using the FPI to measure the Doppler shifts on the infra-red emission at a wavelength of 8430Å, from which the winds of the upper mesosphere may be derived. This emission is part of the Meinel band which is composed of the rotational-vibrational bands of the excited hydroxyl radical, OH* [Meinel, 1950]. The combined emissions from the Meinel band are dominant features of the night airglow [Krasovskij and Šefov, 1965]. The peak emission height is at a height of around 88 km, which is just below the mesopause at 90—95 km [von Zahn and Kurzawa, 1989]. The FWHM of the intensity profile is about 6 km [Thomas and Young, 1981, Baker and Stair, 1988].

Generally the hydroxyl (6,2) emission band has a weak intensity of 400 — 600 Rayleighs, all in the infra-red. This has made them very difficult to observe with a standard FPI [Smith et al., 1988b]. In particular, photomultipliers and photon detectors, such as the S-25 detector, have had very poor Detective Quantum Efficiencies (DQE) in the infra-red spectral region due to the low energies involved. The etalon gap was fixed at 20.49 mm in order to overlap this line with its doublet pair at 8450Å and hence boost the observed intensity. But it is the introduction of a Gallium Arsenide (GaAs) photocathode which has revolutionized observations of the hydroxyl emission.

The FPI had been operated over the winter of 1988-1989 using a standard S-25 detector, in anticipation of delivery of the new GaAs detector. At the end of August 1989, after much preliminary testing, a GaAs detector was eventually installed in the Fabry-Perot Interferometer at Utah. Until then it was only possible to make measurements of the 6300Å emission with the S-25 detector. At middle latitudes the intensity of the airglow emission at 6300Å is very weak and so the quantity and quality of the data collected with a S-25 detector at CASS was poor compared with data from the other, high-latitude, sites such as Kiruna. Airglow at high-latitude regions achieves high intensities which are boosted by the activity of the auroral oval. The profound improvement in the data effected by the new GaAs detector was quite a sensation. Figure 6.5a is a sample of the quality of the fringes observed using the S-25 photocathode and Figure 6.5b shows the high quality obtained by the GaAs photocathode. The narrowness
of the fringes and high signal to noise ratio promise excellent determinations of the Doppler shifts and therefore accurate wind measurements. The error on the position of the peaks is the equivalent of around 1 m/s for data integrated over 240 seconds. Only a handful of peak positions each night were found to pass the quality controls imposed during data analysis on the observations made using the S-25 detector. In contrast, with the GaAs detector the number of peak positions successfully measured each night jumped to around 150, averaging 30 data points per observing direction. A comparison of the $K_p$ indices for the nights of the 27-28th and 29-30th August show that both nights were similarly fairly active with $K_p$ values around $K_p = 4$. Thus the improvement in the quality of the peaks may be attributed solely to the new GaAs detector rather than being a consequence of greater intensity in the airglow.

For the following night the filter was changed and observations of the hydroxyl emission commenced on the night of the 30th August 1989.

6.3.1 Standard Behaviour of Mid-Latitude Mesospheric Winds

The first few nights of observation, using the new GaAs detector, yielded spectacular results owing to almost perfect observing conditions of cloudless nights and high signal to noise ratios for the emissions. These wind data, from the 30th August 1989 to the 5th September 1989, are displayed in Figures 6.6 and 6.7 as both meridional and zonal components. The two figures show the wind components for six successive nights. Each arrow represents an interpolation at 15 minute intervals of the magnitude of the zonal or meridional component of a horizontal wind vector.

The most obvious point is the large variability of the winds, with no two nights exhibiting the same pattern. The peak speeds achieved were around 60 m/s, occurring at times that varied by up to 2 hours. Overall the variation in the wind components did show an obviously cyclic behaviour, despite the period of observation being limited to around eight hours, which is the length of a mid-latitude summer night. The variation appears to be nominally semidiurnal, with an underlying offset, in both the phase and the amplitude of the velocity variation, which varied from night to night. It is planned to analyse these results more extensively by fitting sinusoidal functions to find out the periodicity more exactly.

The extreme variability of the mesospheric winds is well known. Figure 6.8 is an excellent illustration of this, showing measurements made in November-
December and May-June of zonal winds by 36 rockets carrying chemical trails launched from sites near 30°S [COSPAR, 1972]. The average zonal wind speed, which is drawn as a dotted line, is a poor reflection of the instantaneous behaviour. The observed zonal winds reach speeds as high as 150 m/s yet the average speeds are no more than 50 m/s.

Further, Fesen et al. [1986] found the winds of the mesosphere to be particularly variable at middle latitudes. This was as a result of attempting to fit the semidiurnal amplitudes and phases of winds and temperatures at various altitudes predicted by the NCAR TGCM [Dickinson et al., 1981] to incoherent scatter observations from Millstone Hill (42.6°N), Saint Santin (44.6°N) and Arecibo (18.3°N).

The winds observed by the FPI can reach over 50 m/s, and are therefore in agreement with the rocket data. In contrast, the amplitude of the semidiurnal tides predicted by theory is between 5 — 10 m/s [Forbes, 1982b, 1985, Parish et al., 1990]. Even allowing for the night by night variability, an average over several nights would still produce winds much larger than the predicted values. For example, the average nighttime meridional wind for the whole of September 1989 reaches a maximum southwards value of 24 m/s at around 05UT, while the zonal wind reaches a maximum westward value of 17 m/s at around 07:45UT. It should be noted that the average zonal winds at the beginning of the night were larger than this, but have not been discussed in detail because it is not possible to tell whether they were at the peak of the variation or on the downward leg. The spread of the average wind speeds was between ±10 — 20 m/s, reflecting the lack of consistency of the nighttime behaviour.

The Forbes model [1982b, 1987a] predicts a maximum westward wind at around 08:30UT for a solar semidiurnal tide at equinox for a latitude of 42° (Figure 6.9). Between 70 km and 90 km the semidiurnal tide is composed of equal contributions from the (2,2), (2,3) and (2,4) modes [Forbes, 1982b]. At equinox, however, the asymmetric modes are absent, thus the mesospheric winds observed in early September at Utah (Figures 6.6 and 6.7) should be representative of a combination of mainly (2,2) and (2,4) tides. The model predictions for the phase of the zonal winds compares favourably with the phases observed with the FPI at Utah (Figure 6.7). The model data are calculated in terms of the Local Time, LT, and phase, $\phi$, at which a zonal or meridional wind achieves its maximum amplitude i.e. LT = ±12 — $\phi$ [Fesen et al., 1986]. This is then converted to Universal Time, UT, by calculating that the longitude of Utah
causes the Local Time to be 7.4 hours behind Greenwich Mean Time. However, the model predicts that the meridional wind should reach a maximum southward speed at around 11:30UT, which is 6.5 hours later than observed at Utah. As for the predicted phase of a diurnal tide, the vertical wavelength is so small that over the height region 80 – 90 km all phase values are possible (Figure 6.10).

A semidiurnal tide is expected to exhibit phase quadrature, which means that the wind vector rotates in a clockwise direction as the wave propagates upwards [Rees et al., 1990]. It is easily apparent from Figures 6.6 and 6.7 that a sinusoidal variation exists in the amplitude of the zonal and meridional winds, superimposed on a constant value offset. The required clockwise rotation of the wind vector can then be deduced from the relative phases of the zonal and meridional components. This is illustrated by the winds observed on the night of the 5th September, which hardly requires an offset. It can be seen that at the start of observation, at around 03:30UT, the wind blew to the south-east, and rotated to the south at 05:00UT, then by 07:30UT the wind was blowing to the west and had rotated round to the north by 10:30UT.

The zonal components appear to require a larger offset to their amplitudes than the meridional components. This could possibly be an artifact of the data analysis procedure caused by a poor choice of baseline or a real difference in the meridional and zonal interactions between tidal modes. It is not possible to check the correctness of the baseline except by a comparison of the derived winds with data from other sources, such as an incoherent or coherent scatter radar.

The eight hours of night available for FPI observations has been a limiting factor when attempting to see semidiurnal variations. Periodic variations cannot be easily resolved when the span of the observations covers less than half the period of the tide. Theoretically the error in fitting the tidal components increases at an exponential rate as the span decreases according to Crary and Forbes [1983]. They calculated that the span of hourly observations should cover 15.5 hours and 13.5 hours for the diurnal and semidiurnal components respectively in order to obtain fitting errors with a magnitude of the same size as the standard deviation due to the noise component in the observations. Practically it is not easy to determine the times when the meridional wind component reached a maximum northward or southward wind speed from the available data, shown in Figure 6.6, because the peak values often occur near the edges of the observing time-window.

235
On the 30th August 1989 a maximum northward wind occurred at around 11UT while on the 5th September a maximum southward wind occurred at around 05UT. These two nights are consistent with each other as evidence for a semidiurnal tide with a period of 12 hours, since the maximum northward and southward winds occurred with a separation of 6 hours. The other night with a clear maximum northward wind was the 3rd September. Here the maximum northward wind occurred at around 10:30UT, but the maximum southward component appeared to occur just after 06UT, which gives an apparent period of around 10 hours. Since modes other than the diurnal and semidiurnal ones are not significant a period of 10 hours is likely to indicate some mixing of the dominant modes. The night of the 1st September was more complicated, with the ebbing and flowing of the southward wind between 04UT and 08UT suggesting a jumble of competing modes.

The zonal wind components plotted in Figure 6.7 are less easily analysed than the meridional components due to the large offset in the nightly behaviour. One of the simplest nighttime variations in wind behaviour occurred on the 30th August 1989. Here a sinusoidal variation with a 6 hour period and an offset of around 25 m/s is clearly evident. Such a high order mode has never been modelled theoretically, since modes above a triurnal have not been considered fundamental. Therefore, as with the 10 hour period observed in the meridional wind variation on the 3rd September the 6 hour period variation in the zonal component was probably a fortuitous combination of other more likely modes, to give an appearance of a quaternal mode.

The interpretation of the wind data is complicated by the fact that the layer of hydroxyl emission has a finite width. This is shown in Figure 6.11 which plots the variation of the intensity of the hydroxyl emissions at various wavelengths with altitude, as measured by rockets [Krasovskij and Šefov, 1965]. Therefore the FPI will see a vertical integration of wind speeds over several kilometres. The average vertical wavelengths of the (2,2) and (2,4) modes are 96 km and 38 km respectively (Table 6.1). This means that, for instance, the (2,2) mode will rotate in phase by $2\pi$ over a height distance of 96 km. The FWHM of the hydroxyl emission layer is about 6 km so the phase change of the (2,2) mode over this distance is very small and so not important. However, with the (2,4) mode the phase change is very important since the phase will change by nearly one sixth. A full analysis of the data would need to take into account this factor. For
example, Figure 6.12 is a simplified diagram representing the emission profile of the hydroxyl layer as a tophat function, with half width half maximum (HWHM) of $\Delta H$. Let $h_o$ be the mean height of the emission and $V_o$ the wind speed at that height. If $\lambda_z$ is the vertical wavelength of a tidal mode then the average speed registered by the FPI will be given by equation 6.4.

$$V(h) = \frac{1}{2\Delta h} \int_{h_o-\Delta h}^{h_o+\Delta h} V_o \cos \left( \frac{2\pi}{\lambda_z} (h - h_o) \right) dh$$

(6.4)

Table 6.2 shows how radically the observed wind speed can vary with different values of the HWHM of the emission layer for the two main tidal modes (2,2) and (2,4). For a typical HWHM of $\Delta h = 3$ km there is not much difference between the observed wind and the mean wind at the height of the peak emission for the (2,2) mode. However, the shorter average vertical wavelength of the (2,4) mode means that the observed wind can be significantly different.

This is an over-simplified evaluation of the effect of an emission profile with a finite width on the winds observed with an FPI. The emission profile used in this example is a tophat function, while the actual shape is more like a Gaussian profile, which will modify the results in Table 6.2 accordingly. Further, no account has been taken of the possible exponential increase in wave amplitude with height nor of the effect of diurnal or other tidal modes. Most importantly, the FPI will actually see an average of the observed winds from all modes, while the calculations of the height integrated observed wind in Table 6.2 consider only one mode at a time. Thus altogether the interpretation of the FPI data may be very complex.

Table 6.2

<p>| Prediction for the Observed Wind Speed $V_{obs}$ for Various Values of the Half Width Half Maximum $\Delta h$ |
|---|---|---|
| where $V_o = 50$ m/s | $\lambda_z = 38$ km (2,4) mode | $\lambda_z = 96$ km (2,2) mode |</p>
<table>
<thead>
<tr>
<th>$\Delta h$ (km)</th>
<th>$V_{obs}$ (m/s)</th>
<th>$V_{obs}$ (m/s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>49.8</td>
<td>50.0</td>
</tr>
<tr>
<td>3</td>
<td>48.0</td>
<td>49.7</td>
</tr>
<tr>
<td>5</td>
<td>44.5</td>
<td>49.1</td>
</tr>
<tr>
<td>10</td>
<td>30.1</td>
<td>46.5</td>
</tr>
</tbody>
</table>
6.3.2 Possible Evidence of Gravity Waves

There are several instances of transient disturbances that are possible examples of gravity waves passing through the field of view of the FPI. Of these, two that occurred on the 1st September 1989 and the 9th September 1989 are illustrated in Figures 6.13 and 6.14 respectively.

On the 1st September a sudden fluctuation in the wind speed was observed at around 06 UT. This was observed only to the south, south-west, east and north-east thus implying the existence of a swathe of disturbance covering only part of the field of view. The diameter of the field of view is calculated to be around 300 km for a zenith angle of observation of 60°, and on assuming that the emission layer peaks at 85 km. The disturbance was observed at 05:54 UT in the east and at 06:42 UT in the south-west. This could indicate a propagating wave with a component travelling from east to south-west with a speed of around 280 m/s.

A large change in intensity accompanied the speed fluctuations in these two directions. The intensity increased by 25% in the east and decreased by 30% in the south-west. By the time of the next observation for each direction the intensities had returned to the previous values. There was no such intensity variation accompanying the speed fluctuations observed to the south and north-east. At 07:30 UT the speed of the wind component in the north-east direction increased sharply and the intensity increased by 20%. There was no obvious change noticeable in the other directions. These variations in intensity may be important in validating whether a gravity wave has passed by rather than a false Doppler shift caused by auroral contamination of the emission peak, as described in Chapter 2, subsection 2.4.4.3.

Disturbances were observed in all directions except the north on the night of the 9th September 1989. Large intensity fluctuations of around 25% accompanied the changes in wind speed. In the south-west it was noted that the intensity leapt by 50% and stayed at this high value between 07:30 UT and 08:30 UT. Since the Kp indices for that night ranged between 1+ and 2°, thus indicating a geomagnetically quiet night, it was unlikely that the data were contaminated by thermospheric emissions associated with an extended auroral oval and high auroral activity.

The timings of the disturbances in each direction are not consistent with a simple disturbance travelling from one end of the field of view to the other. The increase in intensity lasted for up to an hour in the south-west: from 07:30 UT to
08:30UT, while to the north-west the intensity dropped by 25% at 06:42UT and stays low until 09:12UT, when it jumped up by 25% before falling back to the previous level. The rapid fluctuations in the wind speeds are indicative of the presence of a gravity wave, but the comparatively long-term changes in intensity may also indicate the passage of a complex combination of gravity waves and tidal modes across the field of view.
Figure 6.1

Hough functions for the two antisymmetric and symmetric modes of lowest order for the semidiurnal solar tide. [Hargreaves, 1979]

Figure 6.2

Net upward flux of diurnal tide energy. The net downward flux of solar radiation in bands absorbable in the thermosphere is shown for comparison. [Lindzen, 1967]
Figure 6.3

Comparison of geomagnetic records from the auroral zone with electron content data from mid-latitudes, showing the onset of a magnetic substorm and the subsequent large-scale TIDs [Davies, 1971]
Level 12 is at an altitude of approximately 280 km. [Forbes, 1987] Interpolates are one pressure scale height. For the conditions of this model, pressure is assumed at an altitude of 80 km and pressure of 1 Pascal. The vertical pressure minimum (P10 = 75) conditions. Pressure level 0 is the model lower boundary.

University College London, UCL. The simulations correspond to equinox solar winter winds at 60°N with and without simultaneous lower boundary forcing in the

Figure 6.4
Comparison of the reduced to radius-squared fringe profile obtained for the emission at 6300Å with (a) an S-25 detector and (b) a state-of-the-art Gallium Arsenide (GaAs) detector to show the vast improvement in the signal to noise ratio.

Figure 6.5
Figure 6.6

244
Figure 6.7
245
Figure 6.8

Profiles of the W-E component of the wind as measured from 36 rockets which carried chemical trail experiments all near 30°N latitude between the 15th November and the 14th December or near 30°S latitude between the 15th May and the 14th June. (COSPAR 1972)
Figure 6.9

Solar semidiurnal westerly winds for December solstice from the Forbes model. [Forbes, 1987]
Figure 6.10

Solar diurnal westerly winds for December solstice from the Forbes model. [Forbes, 1987]
Figure 6.11

Rocket measurements of the altitude distribution of the hydroxyl emissions with wavelengths: (a) 7280Å, (b) 7600-10,400Å and (c) 9100-10,700Å. [Krasovskij and Šefov, 1965]

Figure 6.12

A simplified hydroxyl emission altitude distribution used to determine the average integrated winds in the upper mesosphere as observed with an FPI at Utah.
Figure 6.13
Figure 6.14
7

Conclusion

7.1 Review

Thermospheric and mesospheric winds have been presented in this thesis, measured from the Doppler shifts of emissions from the airglow and aurora by Fabry-Perot Interferometers situated at three high-latitude sites in Scandinavia, which are: Kiruna in Sweden, Kilpisjarvi in Finland, Longyearbyen on the island of Svalbard; and a mid-latitude site at Utah in North America. The thermospheric winds presented from the Scandinavian instruments were calculated from the Doppler shifts on the emission at 6300Å from the atomic oxygen transition: OI($^1D-^1S$). The mesospheric winds observed at Utah were calculated from the infra-red emission of the hydroxyl radical OH*, at 8430Å.

The main body of data came from Kiruna, which is a high-latitude site that borders the auroral oval. The FPI at Kiruna has been run routinely every winter since November 1981, building up a database of over 1200 nights of thermospheric wind observations that covers almost a complete solar cycle. The database has been sorted according to both geomagnetic activity in terms of the $K_p$ index, and solar flux in terms of the $F_{10.7}$ index. Several significant geomagnetic, seasonal and solar cycle effects, not previously observed nor expected, have been found.

The results of an analysis of the seasonal variation of the thermosphere in Chapter 3 have revealed a very large difference in the winds observed at Kiruna during spring and autumn for similar geomagnetic activity levels, despite both seasons experiencing the same levels of solar insolation. The average thermospheric winds observed during the autumn appear to be more 'winter-like' than the winds observed during the spring. The greatest asymmetry appears during the period around midnight, when, in particular, the average spring meridional winds are up to 70% larger than the average autumn winds. On dividing the data further to account for solar activity it was found that the asymmetry in the magnitude of the winds around midnight is greater at solar maximum than at solar minimum. In addition the average zonal winds in the dusk sector are
more westward at solar maximum than at solar minimum for all three seasons studied due to increased ion drag. The largest solar cycle effect is observed at around 18UT in the average zonal winds for winter, which are up to six times more westward at solar maximum than at solar minimum. For example, the average zonal wind for winter at solar maximum at 18UT is 61 m/s westward, while at solar minimum it is only 15 m/s westward.

A subsequent study of the seasonal variation of plasma drift, measured by the EISCAT radar, also showed a similarly large asymmetry in the magnitudes of the ion velocities in spring and autumn [Farmer and Jarvis, 1991]. However, it was found that at solar maximum the average plasma velocities in spring were larger than in autumn at solar maximum, in agreement with the thermospheric wind observations, yet at solar minimum they were smaller. For example, in the hour around midnight the average magnitude of the plasma flow vector in spring was twice as large as the average vector in autumn at solar maximum, but at solar minimum the average midnight vector in autumn was three times larger than the vector in spring.

The existence of an equinoctial asymmetry has not been predicted by either thermospheric nor ionospheric model simulations. In fact there has been comparatively little investigation into the behaviour of the thermosphere and ionosphere during the two equinox seasons compared with the two solstice seasons. Of those papers that deal with seasonal variations there is usually contained an implicit assumption that the spring and autumn thermosphere/ionosphere are fundamentally the same, so that simulations from theoretical models tend not to differentiate between spring and autumn, using forcing functions that are symmetric about the winter/summer solstice, while empirical studies often amalgamate the two data-sets.

The solar flux densities in spring and autumn are effectively the same, thus the difference in the spring and autumn thermospheric winds must be due to the momentum transferred to the neutral gas from ions. This idea is supported by the larger plasma velocities observed during the spring at solar maximum. It is suggested, therefore, that the source of the equinoctial asymmetry lies in a seasonal variation in the high-latitude electric field that drives ionospheric convection at the poles.

The strength of the high-latitude electric field is dependent on the orientation of the Interplanetary Magnetic Field (IMF), \( \mathbf{B} \), with respect to the Earth's magnetosphere, in particular the magnitude of the \( z \)-component, \( B_{mz} \).
which is the projection of the IMF onto the magnetic polar axis in terms of the GSM coordinate system [Dungey, 1961, Arnoldy, 1971, Akasofu and Ahn, 1980, Fedder et al., 1991]. The average IMF vector assumes an orientation of approximately 45° to the sun-Earth line at the radial distance of the Earth due to the ‘garden-hose effect’. However, the Earth’s rotational axis is tilted at an angle of 23.5° with respect to the ecliptic plane which results in a semi-annual variation in $B_mz$ on taking into account the ‘towards’ and ‘away’ sectors of the IMF, with the maximum southward $B_mz$ occurring at the two equinoxes [Russell and McPherron, 1973]. In addition the diurnal rotation of the geomagnetic pole about the Earth’s rotational axis results in a diurnal variation in $B_mz$ [Russell and McPherron, 1973]. An equinoctial asymmetry then arises because the magnetosphere in spring is tilted in the opposite direction to the magnetosphere in autumn resulting in a 12 hour phase difference in the diurnal variation of $B_mz$. Thus the maximum southward $B_mz$ occurs at 22UT at the spring equinox and at 10UT at the autumn equinox.

As a consequence of the variation in $B_mz$ it is to be expected that there should be both a diurnal and semiannual variation in the high-latitude electric field. Thus as Kiruna passes through the Harang discontinuity at 21UT the cross polar cap potential is near its maximum value in spring and minimum value in autumn; in which case the maximum amount of momentum is transferred from ions to the neutral gas flowing over polar cap in spring, and the minimum amount of momentum is transferred in autumn. The ions respond within a few minutes to changes in the convection field [de la Beaujardière et al., 1987], and are diverted into the dawn or dusk cells of the auroral oval as they reach lower latitudes. This may account for why the EISCAT radar does not see the predicted asymmetry (where the plasma flow is larger in spring than in autumn) at solar minimum when the site is possibly outside the auroral oval. However, the neutral gas is carried by inertia further equatorward, thus allowing the FPI at Kiruna to observe the large difference in the spring and autumn meridional winds at around midnight, at both solar maximum and solar minimum.

There is an additional equinoctial asymmetry caused by the difference in geometry when the product of the $B_{my}$ and $B_{mz}$ components is considered. In spring the average orientation of the IMF with respect to the magnetosphere is such that $B_{my}B_{mz} > 0$ while in autumn $B_{my}B_{mz} < 0$. The high-latitude electric fields associated with the four permutations and combinations of $B_{my}$ and $B_{mz}$ are very different [Heppner and Maynard, 1987], and as a consequence
should produce a further significant difference between the spring and autumn ionosphere and thermosphere.

The results described in Chapter 4 reveal a solar cycle dependence in the response of the thermospheric neutral winds to geomagnetic activity as measured by the $K_p$ index. There is a difference in time and scale of the response of the solar minimum winds compared with the solar maximum winds. As the $K_p$ value increases the neutral winds assume velocities typical of the level of geomagnetic activity. However, when the solar flux is brought in as another variable it is found that the velocities of the neutral winds change with geomagnetic activity as before but at different values of the $K_p$ index. At solar maximum the neutral winds advance rapidly through their sequence of response to geomagnetic activity, achieving each pattern of behaviour at lower values of $K_p$ than at solar minimum. Thus the nighttime thermospheric winds observed for $2 < K_p < 3$ at solar minimum are similar to the winds observed for $1 < K_p < 2$ at solar maximum. This is a solar cycle difference of one unit of $K_p$ during geomagnetically quiet conditions, but at a higher level of activity a similarity exists between the nighttime winds for $4 < K_p < 5$ at solar minimum and for $2 < K_p < 3$ at solar maximum, which is a difference of two units of $K_p$. In addition, the absolute magnitudes of the neutral winds increase significantly with solar activity for all levels of geomagnetic activity. For example the average meridional winds around midnight for moderate geomagnetic activity ($2° < K_p < 5°$) increase by nearly 70% from around 88 m/s at solar minimum to around 147 m/s at solar maximum.

It was noted also that a solar cycle dependence of thermospheric winds is not apparent at low-latitudes [Babcock and Evans, 1979, Sipler et al., 1982, Burns and Tepley, 1989] and therefore the behaviour of the thermospheric winds observed at Kiruna appears to be associated with the presence of the auroral oval. At low-latitudes the increase in solar heating at solar maximum is offset by increased plasma densities and consequently increased ion drag, resulting in no significant net change in the thermospheric winds. However, at high-latitudes ion drag acts as a driving force on the neutrals due to ionospheric convection within the auroral oval, which then introduces the possibility of a solar cycle dependence.

Altogether the results presented in Chapter 4 indicate that the $K_p$ index is not a sufficient index for the analysis of the response of high-latitude ther-
mospheric winds to geomagnetic activity. The inadequacy of the $Kp$ index as a measure of the input of energy from the magnetosphere into the high-latitude ionosphere has already been observed and it has sponsored the creation of new and exotic indices which are extremely specific to high-latitude geomagnetic activity, the main ones are the $AE$, $AU$, $AL$ and $PC$ indices. Of these the $AE$, $AU$ and $AL$ indices, are derived from measurements of the magnetic perturbations caused by auroral electrojet activity at twelve observatories in the region of the auroral oval surrounding the north pole [Davis and Sugiura, 1966]. Meanwhile, the $PC$ index is derived from the horizontal magnetic perturbations measured at one observatory close to the magnetic pole. [Fairfield, 1967, Troshichev, 1979, Troshichev et al., 1988]. Some of these indices may prove to be a better measure of the effect of geomagnetic activity at the height of the thermosphere and the latitude of Kiruna. However, they have several drawbacks, in particular that the $AE$, $AU$ and $AL$ indices are available only after a delay of several years [Vennerstrøm et al., 1991].

Chapter 5 presented case studies of high-latitude thermospheric winds observed simultaneously at the three Scandinavian sites. Four nights of data selected from December 1987 and January 1988 were displayed to represent geomagnetically quiet, moderate and active periods. The geomagnetically quiet night (11th January 1988) was used to demonstrate that solar heating is the dominant driving force for thermospheric winds under these geomagnetic conditions, driving mainly antisunward winds. Meanwhile the two moderately active nights (15th December 1987 and 21st December 1987) revealed a very different wind flow, where ionospheric convection had become strong enough to divert the antisunward flow driven by solar heating. Under these conditions the evening winds were no longer eastward, as with the geomagnetically quiet night, but were quite strongly westward, following the ion flow in the dusk cell of the auroral oval. The two nights classified as moderately active were also used to demonstrate how very different the winds can be despite both nights being under similar conditions. Short-lived structures were noticed in the thermospheric wind flow which appeared at all three sites despite a latitudinal separation of over 1100 km between Kiruna and Longyearbyen. The geomagnetically active night showed a rare occurrence of very high geomagnetic activity when all three sites were within the polar cap, as is shown by the strong south-eastward winds observed throughout the evening hours at all three sites.
The case studies were compared with simulations of steady state conditions from the UCL-Sheffield thermosphere/ionosphere model. The constraints on the choice of simulation were more severe since the simulation had to match thermospheric wind observations from three independent sites covering a latitude range of nearly 20°. It was possible to find reasonably good comparisons with the observations made on the geomagnetically quiet and moderate nights, but there was no model simulation available to compare with the unusually active conditions observed on the 14th January 1988, which would require a greatly extended electric field model.

Mesospheric winds observed from Utah, which is a mid-latitude site, were presented in Chapter 6. This is the newest of the UCL ventures. The main source of measurements at this altitude have been radar observations of the plasma density through coherent and incoherent scatter. From this neutral winds are derived. Consequently data are sporadic and not too abundant. Observations of the mesosphere by interferometers have been very difficult in the past due to the weakness of the emissions at this altitude and due to the wavelengths being in the infra-red region. The use of a Gallium Arsenide detector in the FPI has greatly extended the spectral range of response of the photon-detecting device, and consequently revolutionized the observations that can be made. For the first time continuous observations have been made of the mesosphere. Tidal and gravity waves have been seen, but much more analysis is needed of the data.
7.3 Future Work

There are many questions still to be answered about the seasonal and solar cycle variation of high-latitude thermospheric winds presented in this thesis. The seasonal variation at solar minimum is not the same as at solar maximum, although the equinoctial asymmetry is still strongly apparent, with the thermospheric winds significantly larger in spring than in autumn for both solar maximum and solar minimum conditions. Interestingly EISCAT data measured at solar maximum show larger spring plasma velocities while at solar minimum the plasma velocities are smaller in spring than in autumn. As yet this has been attributed to the average size of the auroral oval being smaller at solar minimum, but no formal investigation of this proposal has been done.

The model presented in explanation of the equinoctial asymmetry is based on the observations made at Kiruna and Tromsø, which are two sites in close proximity. It will be necessary to test the predictions from the model with observations from high-latitude sites at other longitudes. For example, the FPI located at College, Alaska (65°N, 148°W) passes through the Harang discontinuity at around 12UT. As a consequence observations of nighttime thermospheric winds would be expected to show an equinoctial asymmetry, but with larger meridional winds during the autumn, which is the opposite of the observations at Kiruna.

There are also many refinements that can be made to the model since it explains only the average behaviour. The instantaneous values of the IMF vectors can be greatly different from the average value. For example, it has been assumed that the average number of ‘away’ sectors equals the number of ‘towards’ sectors but the ratio can vary from solar cycle to solar cycle and only the average numbers over several solar cycles appears to be equal [Rosenberg and Coleman, 1969, Russell and McPherron, 1973]. In addition the average magnitude of B in the ‘towards’ and ‘away’ sectors is not necessarily the same, as has been shown by the averaged IMF data over the years 1963-1990 [Wild, private communication, 1991]. The data analysed here cover slightly less than one solar cycle and run from one solar maximum to the next solar maximum. A new solar cycle starts at solar minimum thus the data cover the second half of one cycle and the first half of the next cycle. Thus the effect on this database of different proportions of ‘towards’ and ‘away’ sectors on the solar wind-magnetospheric coupling warrants investigation.

The unequal numbers of ‘towards’ and ‘away’ sectors of the IMF within a
solar cycle would also have an effect on the average high-latitude electric field owing to the dependence on the $y-$component of the IMF. Theoretically, with no $B_y-$dependence, the two ionospheric convection cells are symmetric, but introduction of this parameter leads to the dominance of one cell over the other [Heelis, 1984, Sojka et al., 1986, Heppner and Maynard, 1987]. This bias in the orientation of the IMF will need to be considered in the interpretation of the seasonally averaged thermospheric winds.

We are now attempting to simulate the equinoctial asymmetry observed at Kiruna using the UCL-Sheffield thermosphere/ionosphere model. Since it is a time-dependent model it is possible to introduce a diurnally varying high-latitude electric field. Theoretically derived functions relating the orientation of the IMF to the cross polar cap potential will be tried, such as the solar rectifier model [Arnoldy, 1971] and those suggested by Fedder et al. [1991]. The asymmetry in the high-latitude convection pattern due to the product of $B_{my}B_{mz}$ will also be tested.

A further project to be undertaken is the simulation of the conditions observed on the night of the 14th January 1988, presented in Chapter 5. As yet there are no ionospheric models that simulate such a large auroral oval as was apparent on this night. Geomagnetic activity as high as this occurs very rarely, less than 1% of the time [Foster et al., 1986], and so there are no standard, average electric fields, or particle precipitation patterns on which to base a model simulation.

Waves, tidal and gravity, dominate the physics of the middle atmosphere. The analysis of the upper mesospheric data from Utah is still in its early stages and it is planned to analyse the mesospheric winds observed more extensively by, for example, fitting sinesoidal functions to find out the periodicity more exactly and also to continue the search for gravity waves.

The upper atmosphere is a highly complex system of many separate layers which, historically, have been monitored and analysed quite independently. Now that the gross behaviour of most of the layers has been understood and modelled, the more recent modelling work has attempted to couple the different layers as evidence mounts against the inviolability of the separate layers. For example, the UCL thermospheric model has been coupled with the Sheffield ionospheric model [Fuller-Rowell et al., 1984, 1987] and the boundary of the model has
been modified to permit tidal forcing of the thermosphere by propagating tides from the lower atmosphere [Parish, 1990, Parish et al., 1990]. Now results such as the equinoctial asymmetry and the solar cycle dependence of the $Kp$ variation presented in this thesis show that there are still major variations and anomalies yet to be investigated and understood, and particularly indicate a need to couple with a magnetospheric model. In other words it is not yet time for upper atmospheric physicists to rest on their laurels!
Acknowledgements

My time in the Atmospheric Physics Laboratory of University College London has been a pleasure due to all my colleagues, past and present, who have made this group so cohesive and lively. This facet of my working life has been very important to me.

I have been especially fortunate to have worked with such a large and long-term database of thermospheric winds which has proved to be particularly fruitful in new results. Long-term monitoring programs are unglamorous and therefore vulnerable to short-term monetary policies. The survival of the UCL network of FPIs, particularly during this period of crisis in the funding of science, owes a great deal to David Rees for his persistence and determination to widen the scope of the FPI observations while maintaining the instruments at the existing sites.

My thanks are extended to a number of people: to Nick Lloyd for his initial guidance; to Tim Fuller-Rowell for his kind encouragement, clear explanations and gentle re-routing of my more wayward ideas; to Alan Farmer for his sharp insight and challenging arguments, and also for his generous help with providing the EISCAT data used in Chapter 3. Thanks also to Matthew Wild for compiling the IMF data from the massive database at the World Data Centre at the Rutherford Appleton Laboratory.

During my time in this group the FPIs were designed, built and set up by Ian McWhirter, Jim Percival and Paul Hammond. The sophisticated programming of the instruments evolved through several people, and latterly through Daniel Wade. The number and variety of sites for installing the FPIs would not have been possible without the help and cooperation of a number of people such as Åke Steen and his assistants at the Swedish Institute of Space Physics in Kiruna; Liisa Harrinkoski, who runs the FPI in her house in Kilpisjarvi, Finland; Risto Pellinen at the Finnish Meteorological Institute; Esko Kyro at Sodankyla Ionospheric Observatory, Finland; Kari Kaila at the University of Oulu, Finland; Kjell Henricksen at Tromsø University, Norway; Roger Smith and his assistants from the University of Alaska, who lately have provided our only link with the FPI at Longyearbyen due to the constraints on the funding of field trips; and Vince Wickwar at Utah State University, USA.

Finally I would like to thank my husband Seán McQuaid, who nagged at me to finish this thesis when I deserved it and encouraged me when I needed it. My love to him especially. (There, are you satisfied now?)
References,


262


263


264


265


Fuller-Rowell T.J., D.S.Evans, Height Integrated Pedersen and Hall Conductivity Patterns Inferred from the TIROS-NOAA Satellite Data,


Gussenhowen M.S., D.A.Hardy, N.Heinemenn, R.K.Burkhardt, Morphology of


Hedin A.E., Correlations Between Thermospheric Density and Temperature, So-


270


272


Murayama T., T.Aoki, H.Nakai, K.Hakamada,  Empirical Formula to Relate the Auroral Electrojet Intensity With Interplanetary Parameters,


Oliver W.L., J.M.Holt, R.H.Wand, J.V.Evans, Millstone Hill Incoherent Scatter Observations of Auroral Convection Over $60^\circ < \lambda < 75^\circ$. 3. Average Patterns Versus $K_p$, *J.Geophys.Res.*, 88, pp 5505-5516, 1983


Instrum., 14, pp 1320-1325, 1981


Rich F.J., N.C. Maynard, Consequences of Using Simple Analytical Functions


278
Smith R.W., G.J.Romick, R.Viereck, G.Hernandez, Measurements of Temperature and Wind at the Mesopause by High-Resolution Spectroscopy of
the OH (6-2) and (8-3) Bands, *EOS*, abstract, 1988b


Torr M.R., D.G.Torr, H.E.Hinteregger,  Solar Flux Variability in the Schumann-


281


