Gravity Waves and Small-Scale Structure of the High Latitude Upper Atmosphere

Elaina Anna Katrina Ford

Atmospheric Physics Laboratory,
Department of Physics and Astronomy

University College London

Thesis submitted for the degree of Doctor of Philosophy at the University of London

June 2006

I, Elaina Anna Katrina Ford, confirm that the work presented in this thesis is my own.
Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

This thesis is dedicated to Anna and Robert Ford,

my dearest parents to whom I owe so much.

ABSTRACT

Small-scale structure of the thermosphere is studied at high-latitudes for its important role in ion-neutral coupling. Four Fabry-Perot Interferometers (FPIs) in Scandinavia are primarily used. These are supplemented by a range of other instruments, including the Spectrograph Imaging Facility, radars, magnetometers, all-sky cameras, and satellite data. The FPIs measure the atomic oxygen emission line at 6300Å, from 240km altitude. Emission intensities, thermospheric line of sight wind speeds, and neutral temperatures are obtained.

Comparisons of electron densities from tomography data and EISCAT (European Incoherent SCATter) radar with FPI intensities allow the investigation whether dissociative recombination is the dominant production mechanisms of the nighttime 6300Å oxygen line. Cross correlations indicate that the thermosphere varies on short temporal scales. Altitude variations have less effect due to the broad (~50km) emission height band.

Atmospheric gravity waves in the thermosphere have been detected for the first time in ground-based FPI data using Lomb-Scargle analysis. Joule heating from electrojet currents and particle precipitation in the auroral oval have been identified as the primary source mechanisms using two case study nights. High time resolution data shows a limit to the variability of the thermosphere to be approximately 1-minute.

Statistical studies of the gravity waves from 567 nights of FPI data show that the length of the night and time resolution are the most important influences on the number and periods of waves detected. Greater numbers of short period waves are detected in the rapidly responding intensities, than in the winds and temperatures. Little variation with geomagnetic activity or solar cycle is observed. Periods at particular harmonics of the length of the night are preferred between October and February. Comparisons of mainland and Svalbard data show that the shorter period waves that are formed equatorward in the auroral oval mostly dissipate before reaching Svalbard.

CONTENTS

ABSTRACT **4**

TABLE OF FIGURES **7**

1 INTRODUCTION **16**

1.1 AIMS **16**
1.2 THE EARTH'S ATMOSPHERE **18**
1.2.1 ATMOSPHERIC STRUCTURE 18
1.2.2 THE IONOSPHERE AND MAGNETOSPHERE 22
1.2.3 THE THERMOSPHERE 28
1.2.4 ATOMIC OXYGEN EMISSIONS 31
1.3 SMALL SCALE STRUCTURE **33**
1.4 ATMOSPHERIC GRAVITY WAVES **35**
1.4.1 INTRODUCTION 35
1.4.2 THEORY 40
1.4.3 CLASSIFICATIONS AND PROPERTIES 48
1.4.4 OBSERVATIONS 50

2 INSTRUMENTATION **56**

2.1 INTRODUCTION **56**
2.2 FABRY-PEROT INTERFEROMETERS (FPIS) **57**
2.2.1 INTRODUCTION 57
2.2.2 THE ETALON 58
2.2.3 THE INTERFEROMETER SETUP 63
2.2.4 DETECTORS 66
2.2.5 CALIBRATION 69
2.2.6 DATA PROCESSING 72
2.3 SPECTROGRAPH IMAGING FACILITY (SIF) **74**
2.3.1 INTRODUCTION 74
2.3.2 INSTRUMENTS 75
2.3.3 ALIGNMENT 80
2.4 SUPPLEMENTARY INSTRUMENTS **84**
2.4.1 EISCAT 84
2.4.2 TOMOGRAPHY CHAIN 86
2.4.3 MIRACLE 90
2.4.3.1 IMAGE 90
2.4.3.2 ASCs 91
2.4.3.3 STARE 91
2.4.4 CUTLASS 93
2.4.5 IRIS 95
2.4.6 SATELLITES 95

3 ELECTRON DENSITY CORRELATIONS WITH FPI INTENSITIES 97
3.1 INTRODUCTION 97
3.2 FPI-EISCAT TRISTATIC CAMPAIGNS 99
3.2.1 INTRODUCTION 99
3.2.2 TEMPERATURES AND INTENSITIES 102
3.2.3 NEUTRAL WIND VELOCITIES 107
3.3 TOMOGRAPHY DATA CORRELATIONS 111
3.3.1 DATABASE 111
3.3.2 RESULTS AND DISCUSSION 113
3.4 RADAR DATA CORRELATIONS 120
4 ATMOSPHERIC GRAVITY WAVES: CASE STUDIES 125
4.1 INTRODUCTION 125
4.2 LOMB-SCARGLE ANALYSIS 127
4.3 SYNTHETIC DATA TEST 131
4.4 SECOND TRISTATIC CAMPAIGN 140
4.4.1 THERMOSPHERIC DATA 141
4.4.2 ELECTROJETS 157
4.4.3 ION-NEUTRAL COUPLING 161
4.4.4 SUMMARY OF RESULTS 166
4.5 HIGH-RESOLUTION KEOPS DATA 168
4.6 SVALBARD AND MAINLAND COMPARISON 181
5 ATMOSPHERIC GRAVITY WAVES: STATISTICAL ANALYSIS 194
5.1 INTRODUCTION 194
5.2 DATABASE 197
5.3 RESULTS AND DISCUSSION 201
5.3.1 KEOPS DATA 202
5.3.2 SODANKYLÄ DATA 221
5.3.3 SVALBARD DATA 227
5.3.4 SITE COMPARISONS 235
6 CONCLUSIONS & SUMMARIES 245
6.1 ELECTRON DENSITY CORRELATIONS 246
6.2 AGW CASE STUDIES 248
6.3 AGW STATISTICAL SURVEY 253
6.4 FUTURE WORK 258
6.5 ACKNOWLEDGEMENTS 262
REFERENCES 263

TABLE OF FIGURES

Figure 1.1 *The thermal, ion and neutral density structure of the Earth's atmosphere with altitude. From Rees, (1989).* 18

Figure 1.2 *Electron density variations with altitude showing mid-latitude ionosphere layers (Hargreaves, 1992)* 20

Figure 1.3 *The Earth's magnetic field configuration for a) a northward IMF, b) a southward IMF and c) the convection pattern across the polar regions for a southward IMF. Field lines labelled A are inter-planetary field lines; B are IMF lines reconnecting with a geomagnetic field line; C are open and D closed geomagnetic field lines. Points marked N are the neutral points, dayside reconnection point and the near Earth neutral line (after Hargreaves, 1979).* 24

Figure 1.4 *Convection patterns in a) the equatorial plane and b) the polar cap region with dawn, noon, dusk and midnight sectors (Hargreaves, 1992).* 25

Figure 1.5 *Plasma populations and current systems in the Earth's magnetosphere (Hargreaves, 1992)* 26

Figure 1.6 *Magnetosphere-ionosphere coupling through current systems (Hargreaves, 1992).* 27

Figure 1.7 *Energy level diagrams showing the electronic transitions and emissions of atomic oxygen (Rees, 1989)* 31

Figure 1.8 *Schematic of gravity waves formed in the lee of a mountain range (Houghton, 2002).* 37

Figure 1.9 *Relationships between wavelength, velocity and propagation angle of gravity waves (Hargreaves, 1992)* 42

Figure 1.10 *Dispersion curves for gravity waves, showing the frequency ranges of acoustic and gravity waves. N_B is the Brunt-Väisälä frequency (ω_B). The dashed line is a wave with the speed of sound. (Houghton, 2002)* 45

Figure 1.11 *Propagation in a simple gravity wave. Energy flow is perpendicular to phase propagation, (Hargreaves, 1979).* 46

Figure 1.12 *Height profiles of the speeds of sound (c), the acoustic cut-off frequency ω_a and the Brunt-Väisälä frequency, ω_B. (Hargreaves, 1979).* 49

Figure 1.13 *Gravity waves in all sky camera images of the green line and OH airglow, from Haleakala, Hawaii. (Nappo, 2002).* 51

Figure 2.1 *Constructive interference in a Fabry-Perot Etalon* 60

Figure 2.2 *Circular interference pattern of the sky, through the etalon.* 60

Figure 2.3 *A schematic of the configuration of the FPI. Light enters via the mirror at the top, passes through the etalon, the lenses, and the filter, and then is received by the detector at the bottom.* *64*

Figure 2.4 *Typical laser calibration image showing several narrow ring interference fringes.* *69*

Figure 2.5 *FPI ring-integrated calibration and observation profiles.* *73*

Figure 2.6 *The optical configuration of the SIF spectrograph showing the components and light path (McWhirter, 2002).* *76*

Figure 2.7 *The SIF filter mosaics: a) the 3-band and b) the 4-band mosaic showing emission lines observed and central wavelengths or wavelength bands* *77*

Figure 2.8 *The SIF platform in its dome in Svalbard, showing the four photometers. The spectrograph and TLC are on the other side of the back plate.* *81*

Figure 2.9 *A TLC image from 10th December 1999 showing corona aurora, and the positions of the spectrograph and photometer fields of view, the green curve and red dot respectively.* *83*

Figure 2.10 *Schematic of the spectra from a typical incoherent scattered signal from the F-region, showing the parameters obtained* *86*

Figure 2.11 *Left: track of the satellite path (solid line), receiver stations (see key, bottom right), and minimum observing angle – 25° above the horizon (dotted lines). Right: phases of signal for the four receiver stations, and key to the stations.* *88*

Figure 2.12 *Reconstructions for a pass on the 28th February 2003: top: Chapman reconstruction, bottom: IRI model reconstruction.* *89*

Figure 2.13 *MIRACLE instruments – magnetometer sites, all-sky camera field of views (circles) and STARE Finland radar field of view (from MIRACLE, Finnish Meteorological Institute).* *92*

Figure 2.14 *SuperDARN radar field of views of the 9 northern (left) and 6 southern (right) radars (image is courtesy of the Johns Hopkins University Applied Physics Laboratory).* *94*

Figure 3.1 *Circles representing fields of view of the three FPIs, at Sodankylä; KEOPS, and Skibotn. Points show the measurements taken, in the six directions: north N, east E, south S, west W, and the bistatic A and B positions. The radii of the circles are due to observing the 6300Å oxygen emission line at 240km altitude. Crosses show positions of some of the MIRACLE all sky cameras* *100*

Figure 3.2 *The clear advantage of tristatic measurements is that they observe a common volume. Black lines show the line of sight of the EISCAT radars and red for the FPIs when viewing the tristatic A position.* *101*

Figure 3.3 *Tristatic intensities for two tristatic campaign nights, the 27th and 28th February 2003, and the night before, the 26th, for comparison. Intensities are line of sight measurements from each of the three sites*..102

Figure 3.4 *All Sky Camera Keogram from Kilpisjärvi, which is between the tristatic A point and Skibotn, for the period of the first tristatic campaign night, the 27th February 2003. From the MIRACLE network: http://www.geo.fmi.fi/MIRACLE/ASC/*...103

Figure 3.5 *This All Sky Camera image is from KEOPS at 2323UT on the 28th February 2003.*..104

Figure 3.6 *Neutral temperatures from the Sodankylä north and tristatic A positions (two black lines with crosses) and EISCAT ion temperatures (red line) for the 27th February 2003.*...105

Figure 3.7 *X-component magnetic field data from the IMAGE magnetometer network, which runs from Svalbard down through Scandinavia. The east-west components at Tromsø, Masi (nearest magnetometer to the tristatic point), and Sodankylä are shown*...106

Figure 3.8 *This shows the meridional winds from three different latitude bands for the 27th February 2003 (as may be seen on Figure 3.1 if the viewing volumes are considered in horizontal bands).*...108

Figure 3.9 *Zonal winds for 27th February 2003 for the FPIs at Skibotn (top), Sodankylä (middle), and KEOPS (bottom).*...109

Figure 3.10 *Zenith winds measured vertically above each of the three FPIs for the 27th February 2003. The vertical winds vary with much smaller amplitude than the horizontal winds).*...110

Figure 3.11 *Fields of view of the three FPIs at KEOPS, Sodankylä, and Skibotn, and SGO's Tomography receiver sites (red asterisks). The red line indicates the path of the reconstruction models. Crosses are MIRACLE ASCs.*..................................111

Figure 3.12 *Data from all the usable satellite passes. IRI reconstruction N_e data are compared with FPI intensities from half way through the satellite pass time.*.....115

Figure 3.13 *Correlations between Tromsø electron densities at various altitudes, and FPI data from the three sites, for the first tristatic campaign night.*..................123

Figure 4.1 *a) Time series of test data of a period of 6 hours and noise levels 50% of wave amplitude, with its LSP (b) and amplitudes (c)*..132

Figure 4.2 *a) Time series of test data of a period of 30 minutes and equal noise levels to wave amplitude, with its LSP (b) and amplitudes (c)*................................134

Figure 4.3 *a) Time series of test data of a period of 6 hours, with its LSP (b) and amplitudes (c), with large noise levels (4 times wave amplitude).*............................135

Figure 4.4 *Test data with waves of periods of 4 hours and of 30 minutes, with noise levels equal to that of the 4 hour wave amplitude: a) time series, (b) its LSP and (c) wave amplitudes.*..139

Figure 4.5 *EISCAT data from 25th November 2003, showing a) electron density, b) electron temperature, c) ion temperature, and d) ion velocity. A wavefront of a gravity wave in electron density is shown with a dashed line.*.................................142

Figure 4.6 *6300Å intensities from KEOPS, Sodankylä, and Skibotn on the 25th November 2003 towards the tristatic point above KEOPS. As the intensities are not calibrated, the Sodankylä data are scaled to match the other sites by a factor of the ratios of the median value of Skibotn to Sodankylä data for this night.*....................143

Figure 4.7 *6300Å intensities (a), neutral winds (b), and temperatures (c) for all the look directions from KEOPS on the 25th November 2003.* 144

Figure 4.8 *Lomb-Scargle periodograms for a) KEOPS b) Sodankylä and c) Skibotn FPI intensities on the 25th November 2003. The colours represent each of the different look directions observed, and the horizontal lines are 70%, 90% and 99% confidence levels.* 146

Figure 4.9 *All Sky Camera Keogram from Kevo for the 25th November 2003. The image shows 5577Å aurora, and has been made negative for clarity.*.....................147

Figure 4.10 *Amplitudes of Lomb-Scargle periodograms for FPI intensities for a) KEOPS, b) Sodankylä, and c) Skibotn, on the 25th November 2003.* 149

Figure 4.11 *Periodogram from KEOPS wind data with a 30-point (~120 minute) smoothed value removed, for the 25th November 2003.*..150

Figure 4.12 *Neutral temperatures with a 2-hour smoothed value removed (a) and their periodogram (b) from KEOPS from the 25th November 2003. The bottom plot also shows the periodogram for the equivalent current densities (thick black line) from IMAGE data (smoothed and reduced by a factor of 2 for comparisons).*.........152

Figure 4.13 *All-Sky Camera image from KEOPS on the 25th November 2003, showing an auroral arc to the north of KEOPS. Green lines show the look directions; red dots show the 6300Å emitting volume presuming a 240km height; circles show other height distances.* 154

Figure 4.14 *Cross correlations on smoothed temperatures from KEOPS on the 25th November 2003: north and south (solid line), east and west (dotted line).* 155

Figure 4.15 *Neutral temperatures with a 2-hour smoothed value removed, then with a secondary 40 minute smoothing, from KEOPS on the 25th November 2003.* 156

Figure 4.16 *The LSP for FPI intensities from KEOPS from the 25th November 2003. The periodogram for the equivalent current densities from IMAGE data (thick* *black line) has a 200-minute smoothing removed, and the periodogram was divided by 4 for comparisons.*......158

Figure 4.17 *Positions of the electrojets (a) and periodogram (b) from IMAGE B_z data, from the 25th November 2003.*......160

Figure 4.18 *Lomb-Scargle Periodogram from EISCAT electron densities from the 25th November 2003. Colours are for different heights along the Tromsø beam, blues are low altitudes, red high altitudes.*......161

Figure 4.19 *EISCAT hmF2 values for the 25th November 2003 (grey) and estimated OI emission altitude (black). The presumed altitude of 240km is shown with a horizontal dotted line.*......162

Figure 4.20 *STARE electric field (northwards component) for the 25th November 2003 (top) and LSP (bottom), at a longitude of 18.0°E. Horizontal lines (top plot) show the latitudes of KEOPS, the tristatic A point, and the northernmost extent of the FPI data. Vertical lines in the LSP highlight some of the waves present.*......164

Figure 4.21 *ASC keogram from Muonio on the 3rd April 2004.*......169

Figure 4.22 *FPI data from KEOPS on the 3rd April 2004 (a) intensities, (b) winds, and (c) temperatures.*......170

Figure 4.23 *1 hour of high resolution FPI data from KEOPS on the 3rd April 2004, including 0.1hr (6min) interpolated tristatic A data: (a) intensities, (b) winds, and (c) temperatures.*......172

Figure 4.24 *10 minutes of high resolution FPI data from KEOPS on the 3rd April 2004, including 0.1hr interpolated tristatic A data (purple line): (a) intensities, (b) winds, and (c) temperatures.*......173

Figure 4.25 *Lomb-Scargle Periodogram of FPI data from KEOPS on the 3rd April 2004: (a) intensities, (b) winds, and (c) temperatures. The low resolution look directions are scaled by a factor of ten for comparisons.*......174

Figure 4.26 *Lomb-Scargle Periodogram up to periods of 1 hour, from KEOPS data on the 3rd April 2004: (a) intensities, (b) winds, and (c) temperatures. The low resolution look directions are scaled by a factor of 4.*......176

Figure 4.27 *Lomb-Scargle Periodogram of smoothed intensities, winds, and temperatures, from KEOPS data on the 3rd April 2004, using a 30-minute running smoothing value.*......177

Figure 4.28 *Lomb-Scargle Periodogram up to periods of 1 hour, from KEOPS data on the 5th April 2004: (a) intensities, (b) winds, and (c) temperatures. The low resolution look directions are scaled by a factor of 4.*......178

Figure 4.29 *Lomb-Scargle Periodogram up to periods of 1 hour, from KEOPS data on the 10th April 2004: (a) intensities, (b) winds, and (c) temperatures. The low resolution look directions are scaled by a factor of 4.*......179

Figure 4.30 *Lomb-Scargle Periodogram up to periods of 1 hour, from KEOPS data on the 11th April 2004: (a) intensities, (b) winds, and (c) temperatures. The low resolution look directions are scaled by a factor of 4.* 180

Figure 4.31 *KEOPS data from the 15th December 2003: a) intensities, b) winds, and c) temperatures.* 182

Figure 4.32 *Lomb – Scargle periodograms from KEOPS data: a) intensities, b) winds, and c) temperatures, from the 15th December 2003.* 184

Figure 4.33 *Svalbard data from the 15th December 2003: a) intensities, b) winds, and c) temperatures.* 185

Figure 4.34 *Lomb – Scargle periodograms from Svalbard data: a) intensities, b) winds, and c) temperatures, from the 15th December 2003.* 186

Figure 4.35 *SIF photometer data from the 15th December 2003, Hβ at 4861Å, blue line and* N_2^+ *at 4652Å, green line.* 188

Figure 4.36 *LSP of smoothed SIF photometer data from the 15th December 2003, a) Hβ at 4861Å, and b)* N_2^+ *at 4652Å* 189

Figure 4.37 *CUTLASS Pykkvibaer beam 5 power (top) and velocity (bottom), on the 15th December 2003.* 191

Figure 4.38 *LSP of IMAGE* B_x *component data from the Kiruna magnetometer, on the 15th December 2003.* 192

Figure 4.39 *60-point smoothed temperatures from KEOPS from the 15th December 2003, with a further 15-point running smooth removed to show lags between directions* 193

Figure 5.1 *Histogram of gravity wave periods for all nights and all activity levels, from KEOPS data over the winter season of 2000-2001: a) intensities, b) winds, and c) temperatures.* 203

Figure 5.2 *Gravity wave peak periods for each night over 2000-2001, from all KEOPS intensities (plus symbols), winds (squares), and temperatures (crosses). Solid line is half the night length from civil twilight times.* 205

Figure 5.3 *Ratio of gravity wave peak periods to night length for KEOPS 2000-2001, intensities (plus symbols), winds (squares), and temperatures (crosses). The season has been folded around the winter solstice to improve statistics.* 206

Figure 5.4 *Monthly counts of gravity waves per number of clear nights, with periods from 0 – 4 hours, from KEOPS during 2000-2001: a) intensities, b) winds, and c) temperatures.* 208

Figure 5.5 *Histogram of gravity wave periods for all nights and all activity levels, from KEOPS data over the winter season of 2001-2002: a) intensities, b) winds, and c) temperatures.*..........209

Figure 5.6 *Histogram of gravity wave periods for all nights and all activity levels, from KEOPS data over the winter season of 2002-2003: a) intensities, b) winds, and c) temperatures.*..........211

Figure 5.7 *Histogram of gravity wave periods for all nights and all activity levels, from KEOPS data over the winter season of 2003-2004: a) intensities, b) winds, and c) temperatures.*..........213

Figure 5.8 *Histogram of gravity wave periods for all nights and all activity levels, from KEOPS data over the winter season of 2004-2005: a) intensities, b) winds, and c) temperatures.*..........214

Figure 5.9 *Histogram of gravity wave periods for all nights for all look directions individually, from KEOPS data from the north, east, south, and west directions, over the winter season of 2004-2005: a) intensities, b) winds, and c) temperatures.*..........216

Figure 5.10 *Proportion of gravity wave periods, for the north, east, south, and west directions, compared with the number per night, from KEOPS data over the 2004-2005 season: a) intensities, b) winds, and c) temperatures.*..........217

Figure 5.11 *Histogram of gravity wave periods for all nights and all activity levels, from KEOPS data over the winter season of 2005-2006: a) intensities, b) winds, and c) temperatures.*..........219

Figure 5.12 *Histogram of gravity wave periods for all nights and all activity levels, from Sodankylä data over the winter season of 2002-2003: a) intensities, b) winds, and c) temperatures.*..........222

Figure 5.13 *Histogram of gravity wave periods for all nights and all activity levels, from Sodankylä data over the winter season of 2003-2004: a) intensities, b) winds, and c) temperatures.*..........223

Figure 5.14 *Histogram of gravity wave periods for all nights and all activity levels, from Sodankylä data over the winter season of 2004-2005: a) intensities, b) winds, and c) temperatures.*..........225

Figure 5.15 *Histogram of gravity wave periods for all nights and all activity levels, from Sodankylä data over the winter season of 2005-2006: a) intensities, b) winds, and c) temperatures.*..........226

Figure 5.16 *Histogram of gravity wave periods for all nights and all activity levels, from Svalbard data over the winter season of 2000-2001: a) intensities, b) winds, and c) temperatures.*..........228

Figure 5.17 *Histogram of gravity wave periods for all nights and all activity levels, from Svalbard data over the winter season of 2001-2002: a) intensities, b) winds, and c) temperatures..230*

Figure 5.18 *Histogram of gravity wave periods for all nights and all activity levels, from Svalbard data over the winter season of 2003-2004: a) intensities, b) winds, and c) temperatures..231*

Figure 5.19 *Histogram of gravity wave periods for all nights and all activity levels, from Svalbard data over the winter season of 2005-2006: a) intensities, b) winds, and c) temperatures..233*

Figure 5.20 *Histogram of gravity wave periods for all KEOPS data (2000 – 2006) at all activity levels: a) intensities, b) winds, and c) temperatures. 236*

Figure 5.21 *Histogram of gravity wave periods for all Sodankylä data at all activity levels: a) intensities, b) winds, and c) temperatures....................................237*

Figure 5.22 *Histogram of gravity wave periods for all Svalbard data at all activity levels: a) intensities, b) winds, and c) temperatures....................................239*

Figure 5.23 *Number of gravity waves with periods 0-4 hours detected in each month as a proportion of the number nights of data, for all KEOPS data: a) intensities, b) winds, and c) temperatures..242*

Figure 5.24 *Number of gravity waves with periods 0-4 hours detected in each month as a proportion of the number nights of data, for all Svalbard data: a) intensities, b) winds, and c) temperatures..243*

TABLES

Table 3.1 *Positions of FPI look directions and the nearest point on the tomography chain reconstruction plane.* 112

Table 3.2 *Summary of results of correlations of Chapman and IRI reconstruction model N_e values with FPI intensities from each of the sites over various periods. Greyed out cells have correlations less that 95% confidence.* 114

Table 3.3 *Correlations of Chapman and IRI reconstruction model N_e values with FPI intensities for significant results*.. 118

Table 3.4 *Correlations of FPI intensities with IRI reconstruction model N_e values at various altitudes.* 119

Table 3.5 *Correlations of FPI intensities with EISCAT radar N_e values between each combination of instruments.* 121

Table 3.6 *Correlations of FPI intensities with EISCAT radar N_e values at various altitudes.* 122

Table 4.1 *Waves obtained from a Lomb-Scargle analysis of synthetic waves of 0.5, 4 and 6 hours, with varying noise levels.* 133

Table 4.2 *Waves obtained from a Lomb-Scargle analysis of synthetic data with multiple waves of periods of 0.5, 4 and 6 hours, with varying noise levels*............ 138

Table 5.1 *FPI data availability for each site: dates data were collected; the detector; the typical cycle times of the FPI; and the numbers of clear nights for geomagnetically quiet, moderate, and active nights, and the total clear nights.* 200

Table 5.2 *Occurrence of gravity waves per night, for each site, over all seasons of data for the intensities, winds and temperatures*.. 240

1 INTRODUCTION

1.1 AIMS

The aim of this thesis research is to study small-scale structures in the high-latitude upper atmosphere with spatial scales on the order of tens of kilometres and temporal scales of tens of seconds. This is principally achieved by analysing data from four Fabry-Perot Interferometers (FPIs) (discussed in section 2.2) in northern Scandinavia operated by the Atmospheric Physics Laboratory (APL) at University College London (UCL). Data from a spectrograph platform (SIF, see section 2.3) that includes photometers and a video camera, and data from radars, magnetometers, all sky camera and other instruments are also used.

The thermosphere is monitored at auroral and polar latitudes by using Fabry-Perot Interferometers, which measure the 6300Å red line emission from atomic oxygen. Parameters obtained from the FPIs are the intensity of the red line emission, the line of sight wind neutral velocities from the Doppler shifts and the neutral gas temperatures from the broadening of the line (see section 2.2.6). The separation of the FPI look direction volumes both from a single FPI and between multiple FPIs provide plenty of scope to study meso-scale structures spatially. Small temporal scales of thermospheric variability can now be measured due to recent large improvements in detector technology and capabilities (section 2.2.4).

The Spectrograph Imaging Facility (SIF) is located in the polar cap region and is designed to study proton and electron aurorae. Electron aurorae produce features such as arcs, which are well-defined discrete events and are very dynamic, and hence vary considerably on short spatial and temporal scales. SIF is well suited to study small-scale structures of the polar atmosphere due to the high temporal resolution of both the spectrograph and the photometers. Spectral features can be identified through each of three filters, and spatial structures can be observed along the spectrograph slit (section 2.3). Photometers have a small field of view and have a very high time resolution, while the video camera has a wider field of view and provides a context for the other instruments.

The motivation for the project is to gain a better understanding of processes of the upper atmosphere, such as the energetics, momentum transfer, and dynamics. As the thermosphere comprises at least 99% of the upper atmosphere, it is an important sink for magnetospheric energy and its effect on the ionosphere is significant. However, as the thermosphere is harder to measure than the ionosphere its role is often over-simplified. Studying this will therefore lead to a better understanding of the thermosphere-ionosphere-magnetosphere system which is a key part of the Sun – Earth interaction.

This chapter covers the background to the Earth's atmosphere and, in particular, the parts and processes that are relevant to the work in this thesis. It also covers the theory behind the processes that are studied in the results chapters. Chapter 2 outlines the different instruments used to obtain the data that are later analysed. In Chapter 3 the intensity of the 6300Å emissions from Fabry-Perot interferometer data is compared with measurements of electron densities so that the processes causing the emission can be identified. Chapter 4 shows the first detections of atmospheric gravity waves (AGWs) in thermospheric FPI data, and the first observations of AGWs in neutral temperatures. These results are compared to travelling ionospheric disturbances, which are the ionospheric signature of AGWs, and to ionospheric electric currents. This allows the source and the mechanism for production of the gravity waves can be identified. This is achieved through data from radars, the Spectrograph Imaging Facility (SIF), magnetometer and other data. This chapter shows some case studies, where some particular nights are analysed in detail. Chapter 5 is an analysis of the whole FPI database, which is used to perform a statistical survey of gravity waves, to determine their typical properties and how these vary in different conditions. The key results are then summarised in Chapter 6 and additional future directions of this work are discussed.

1.2 THE EARTH'S ATMOSPHERE

1.2.1 ATMOSPHERIC STRUCTURE

The structure of the Earth's atmosphere is defined by both the temperature gradients with altitude and the ion density changes. The 'shells' of atmosphere that these produce are shown in Figure 1.1 below.

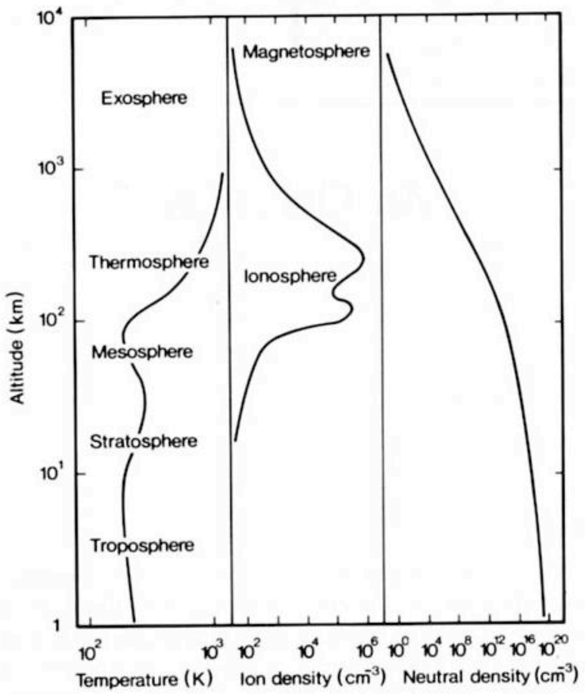

Figure 1.1 *The thermal, ion and neutral density structure of the Earth's atmosphere with altitude. From Rees, (1989).*

The troposphere is the lowest layer of the atmosphere, starting at ground level. This, therefore, contains the biosphere and weather systems and it is a turbulent layer of the atmosphere. Its top boundary is the tropopause, which is at a height that varies with season and latitude, from 4km at winter high latitudes, to 18km in summer equatorial regions. It is heated from re-radiation of solar energy from the ground, so temperature slowly decreases with altitude. Above the troposphere is the stratosphere, where the temperature gradually increases to over 200K, up to 270K, at the stratopause, which is at an altitude of 50km. Ultraviolet energy is absorbed by stratospheric ozone, therefore the temperature increases with altitude, and so convection does not take place, creating a stable layer. From the stratopause, the temperature decreases through the mesosphere, reaching a minimum at the mesopause of 180K at about 80 – 90km. The temperature decreases due to cooling by CO_2 and NO emissions. The turbopause marks the boundary between the lower and upper atmosphere from around 100 – 130km. This is where turbulent mixing becomes less important for distributing atmospheric constituents than diffusion.

The mesosphere and lower thermosphere (MLT) region, sometimes called the middle atmosphere. The top of this region is the mesopause, where meteors burn up which can be viewed by meteor radars. Noctilucent clouds also occur at mesospheric heights, which are ice particles at mesospheric altitudes that reflect sunlight, and so can be seen at ground level after the Sun has set. Above the mesopause, which is the coldest part of the planet, lies the thermosphere. The temperature rises again, due to UV and EUV solar heating, and reaches a maximum of around 2000K, and then remains isothermal above this. The thermosphere reaches altitudes of around 600km, and above this, the density gradually decreases through the exosphere, reaching its upper limit at around 1000km.

The atmosphere may also be defined in terms of electron density, through a classification based on layers or regions. This is shown in Figure 1.2 below giving the vertical profiles of the electron densities for day and night at solar maximum and minimum. The ionosphere is produced by ionisation of gases, mainly N_2, O_2, and O, principally from solar EUV (extreme ultra-violet) and X-rays. This is the reason why the daytime ionosphere has larger electron densities than the night time. As solar input is responsible, the increased EUV flux at solar maximum also increases the strength of the ionosphere at these times. Another source of ionisation is from precipitating electrons and protons and is important at high latitudes in the auroral ovals.

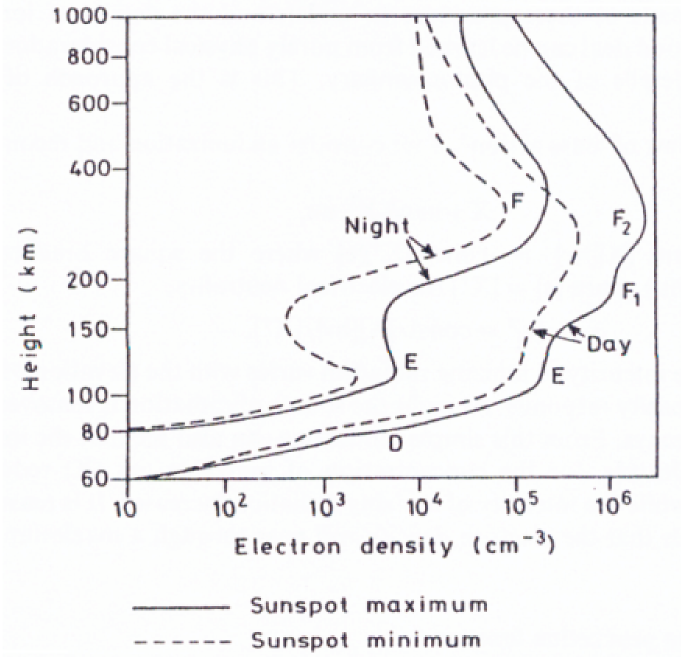

Figure 1.2 *Electron density variations with altitude showing mid-latitude ionosphere layers (Hargreaves, 1992).*

The daytime ionosphere has four distinct peaks in density with altitude. The D region, which peaks at around 90km, and usually disappears at night. The E-region consists mainly of O_2^+ and NO^+ ions and peaks at 110km. The night time E-region it is considerably weaker than during the day. The F-region is mostly O^+ ions and is split into two parts during the day, the F1 and F2 regions. The F1 region, peaking at

170km, is also greatly diminished at night, but the F2 region at 250km altitude remains through the night at a slightly reduced density. This height of the maximum electron concentration is known as hmF2. The region above the F2 peak is known as the topside, it is made of O^+ and H^+ ions, and the density gradually decreases to become the magnetosphere at about 1000km. The role of the ionosphere is discussed in the next section.

A useful parameter for describing processes within the atmosphere is the scale height, H. For example, above the turbopause, the scale height is the range where turbulence no longer mixes the constituent molecules, and their distributions depend on the balance of gravity and the vertical pressure gradient. The atmosphere can be treated as a fluid up to altitudes of around 600km (the topside of the ionosphere). Above this, the density is too low for collisions to dominate. The scale height can be defined using the equation of hydrostatic equilibrium (equation [1]):

$$\frac{\partial P}{\partial z} = -\rho g \qquad [1]$$

This defines the rate of change of pressure, P, with height, z, from the density at that height, ρ, and the gravitational acceleration, g. Using the equation of state for the atmosphere, the ideal gas law, equation [2]:

$$PV = nRT \qquad [2]$$

For a volume of gas, V, at temperature, T, and n moles of gas. R is the universal gas constant, which is defined by $N_A k$, where N_A is Avogadro's number and k is Boltzmann's constant (1.381×10^{-23}J/K). Substituting for a volume of $V = nm/\rho$ for gas of molar mass m, and rearranging, for one mole of gas (i.e. $N_A = 1$), gives:

$$\frac{\partial P}{P} = -\frac{mg}{kT}\partial z = -\frac{\partial z}{H} \qquad [3]$$

where the scale height is defined as

$$H = \frac{kT}{mg} \qquad [4]$$

This is the e-folding distance for the atmosphere as is shown below (5) which is obtained by integrating equation [3]. It is the altitude range over which parameters and processes can be assumed constant.

$$P(z) = P_0 e^{-\int \frac{mg}{kT}\partial z} = P_0 e^{-\frac{z - z_0}{H}} \qquad [5]$$

This is known as the law of atmospheres, or the barometric law, and defines how the pressure of the atmosphere varies with altitude. The scale height at the Earth's surface, presuming a temperature of 250K, is 7.4km. This is valid up to the turbopause because, due to turbulent mixing, the temperature remains approximately constant, and an average value of molecular weight can be used. Above this, the concentrations of the heavier molecules fall off more quickly with altitude, so the mean molecular mass decreases, increasing the scale height. Moreover, temperature changes significantly with altitude, further changing the scale height. At F2 region heights, the scale height has increased to around 50km.

1.2.2 The Ionosphere and Magnetosphere

Although the ionosphere accounts for less than one percent of the mass of the atmosphere, it plays an important role in the atmosphere/magnetosphere system due to the electric currents that are produced because of ions and electrons moving within the Earth's magnetic field. The electrical properties of the ionosphere cause it to interact with radio waves, and this was how it was first discovered. Radars or ionosondes are used to transmit radio waves into the ionosphere, where they are scattered and then the weaker return signal is measured by the radar. The height of the reflections of the radio waves depends on the frequency of the transmitted wave. Scattering can occur either incoherently (see also section 2.4.1), from electron density fluctuations due to the presence of the ions, when the wavelength of the radio wave is smaller than the scattering structure, or scattering from coherent, larger scale, structures (see section 2.4.4) occurs when the transmitted wavelength is larger than the structures.

Ions in the atmosphere are dragged across field lines by the bulk motion of the thermospheric winds, due to Coriolis forces (see section 1.2.3). This for example creates currents and frictional heating sources. At high latitudes however, the situation is much more complicated and the interaction with the solar wind and magnetosphere become important and play dominant roles.

The Sun is the source of the particles that enter the ionosphere at auroral latitudes, and they reach the Earth through the solar wind, the interplanetary magnetic field (IMF), and the Earth's magnetic field. Figure 1.3 shows the process of magnetic reconnection and the Dungey cycle in a simplified form, which can be used to explain particle injection into the high latitudes. The Earth's magnetic field is given for different IMF conditions and the Sun is to the left of the diagrams.

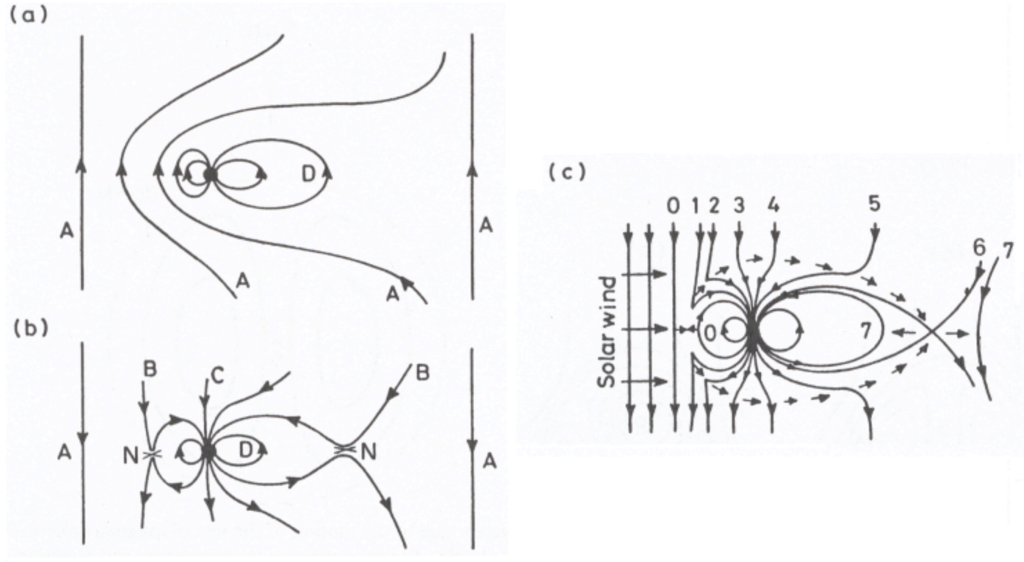

Figure 1.3 *The Earth's magnetic field configuration for a) a northward IMF, b) a southward IMF and c) the convection pattern across the polar regions for a southward IMF. Field lines labelled A are inter-planetary field lines; B are IMF lines reconnecting with a geomagnetic field line; C are open and D closed geomagnetic field lines. Points marked N are the neutral points, dayside reconnection point and the near Earth neutral line (after Hargreaves, 1979).*

For northward IMF configurations, as shown in Figure 1.3a, dayside magnetic reconnection is not possible as the IMF and Earth's magnetic fields are parallel. However, for southward IMF (Figure 1.3b), the IMF and geomagnetic field lines are anti-parallel, so field lines can reconnect on the dayside. This process creates field lines that are connected to the IMF at one end, and the Earth's magnetic field at the other (open field lines labelled C). As the solar wind brings in more field lines, they are dragged anti-sunward across the poles of the Earth and will eventually meet again in the magnetotail and reconnect at the near-Earth neutral line. This convection of field lines is known as the Dungey cycle, and is shown in Figure 1.3c with the numbered progression of field lines from day to night side in the polar plane. The night side reconnected field lines then progress back to the dayside around the sides of the planet. This produces a two-cell convection pattern across the polar cap region, which is shown in Figure 1.4.

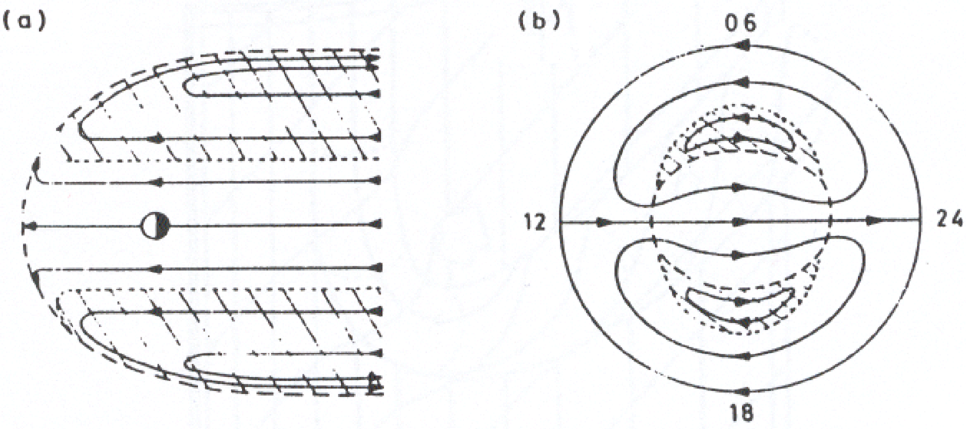

Figure 1.4 *Convection patterns in a) the equatorial plane and b) the polar cap region with dawn, noon, dusk and midnight sectors (Hargreaves, 1992).*

This broad pattern of anti-sunward flow across the pole and sunward flow at lower latitudes is present for southward IMF conditions but varies in shape with IMF B_y component and in size with solar wind speed and intensity, increasing in size for active conditions.

The solar wind consists of particles, mostly electrons and protons, which are frozen-in to the interplanetary magnetic field. When the IMF is connected to the geomagnetic field, these particles can enter the atmosphere and can be observed as aurora. The motion of plasma through the magnetic fields causes currents to flow. The current systems for the magnetosphere are shown in Figure 1.5.

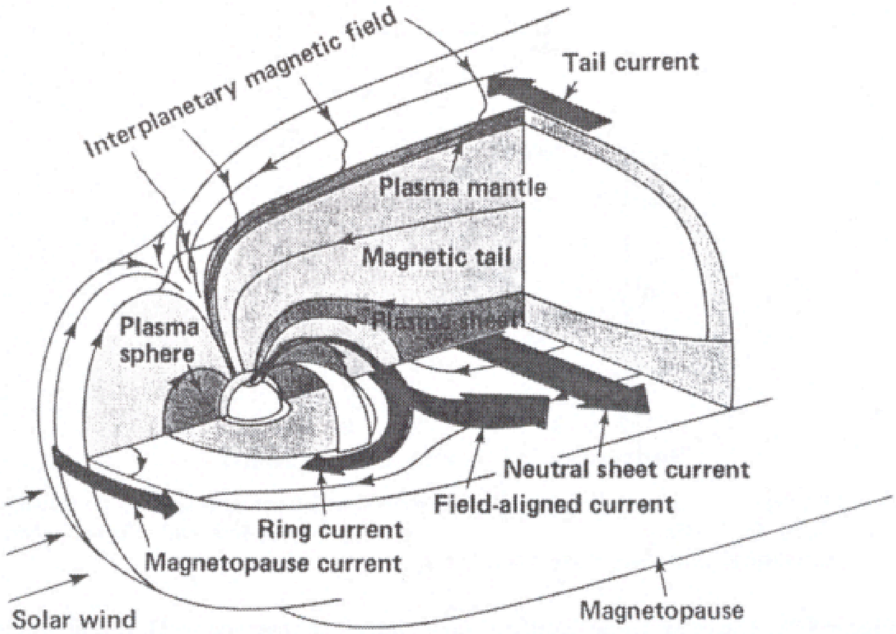

Figure 1.5 *Plasma populations and current systems in the Earth's magnetosphere (Hargreaves, 1992).*

The main parts of Figure 1.5 that are important to the high latitude atmosphere are the field aligned currents and the cusp region (not labelled). The cusp is where day and night side field lines converge, inside the open/closed field line boundary. The cusp region, therefore, contains open field lines, on the dayside, where plasma can easily flow down to the ionosphere. This often happens in bursts when reconnection takes place through what is known as flux transfer events (FTEs). The cusp footprint is on the noon side of the polar cap in the area viewed at Svalbard which is why this is an important region for instrumentation.

The field-aligned currents (FACs) follow the field lines along the tail and into the auroral regions of the ionosphere. They are important because magnetotail plasma, and therefore momentum and energy, is transported along them into the ionosphere after reconnection at the near-Earth neutral line in a substorm (see Figure 1.3b). The magnetosphere is coupled to the ionosphere where the FACs enter the ionosphere, in the auroral ovals. This leads to the formation of electrojets - east/west currents in the auroral ionosphere - shown in Figure 1.6.

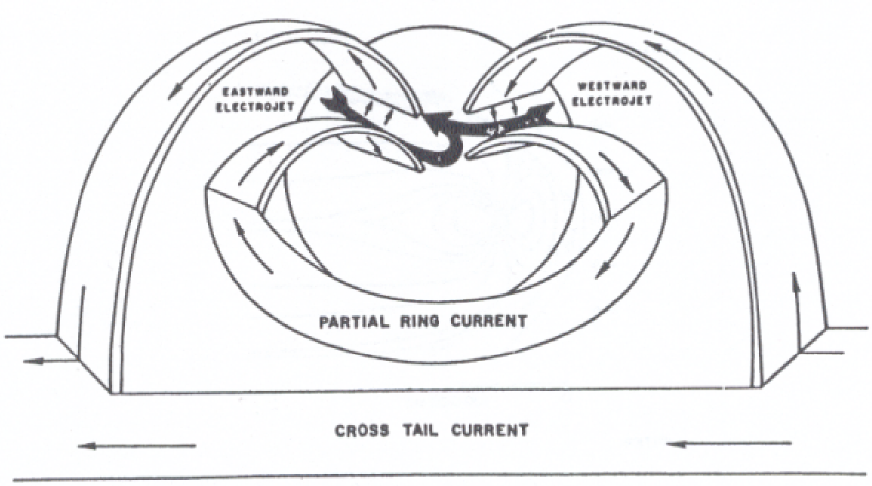

Figure 1.6 *Magnetosphere-ionosphere coupling through current systems (Hargreaves, 1992).*

The field-aligned currents enter the auroral ionosphere in the post-midnight sector, pass horizontally equatorwards through the ionosphere, and out back into the magnetosphere along lower latitude field lines, in the post-midnight sector, and into the partial ring current in the equatorial plane. The current then re-enters the ionosphere, from the equatorial ring current, along field lines to the pre-midnight sector. The currents move polewards and leave the ionosphere at higher latitudes back along field-aligned currents, which take the current, back out into the pre-midnight sector tail. The magnetospheric current systems are therefore completed with the electrojets in the ionosphere.

The ionospheric currents are in the E-region, at altitudes of about 100km, where the density is sufficient for the ion-neutral collision frequency to slow the ions, but the electrons are still free to move creating a potential and current flow. These currents are called electrojets and are shown by thick black arrows in Figure 1.6. In the post-midnight sector, the magnetospheric circuit is completed by an equatorward current through the ionosphere carried by the eastward flow of electrons in the E-region, creating westward electrojets. The reverse of this happens in the pre-midnight sector, where a poleward current creates an eastward electrojet.

There is a resistance to these current flows due to the presence of the thermosphere. The neutral density is much greater than the ion density and as the neutrals do not move with the currents, they provide a resistive friction on the ions, known as ion drag. This resistance provides a mechanism for energy to be transferred from the ionosphere to the thermosphere, and so the thermosphere and ionosphere are coupled.

1.2.3 The Thermosphere

Compared to the ionosphere the thermosphere is relatively inactive, though as over 99% of the atmosphere consists of neutrals they cannot be neglected. Unfortunately, due to their relatively stable large-scale motions and the difficulties in measuring them, the neutrals are rarely measured and coarse modelling of their trends is often deemed sufficient. This is changing, however, as detector technology is improving to allow FPIs to measure the smaller details of the neutrals' behaviour.

Motion of the thermosphere is due to a mixture of driving forces that create winds. Convection due to magnetic reconnection processes creates an anti-sunward flow across the polar cap. Heating, from several sources, causes the atmosphere to expand, leading to pressure gradients between heated and non-heated areas, which creates the thermospheric winds. The strongest source is from solar radiation on the dayside producing a temperature differential from day to night side. However, this flow is complicated from a simple day to night flow pattern by the rotation of the Earth. This creates a Coriolis force, when the winds are considered from a fixed frame of reference. The factors that create these winds are given by equation [6], a simplified Navier-Stokes equation for fluid flows (see e.g. Batchelor, (1967), Houghton, (2002)) the equation of motion for per unit mass.

$$\frac{d\boldsymbol{U}}{dt} = -\nabla P - 2\Omega \times \boldsymbol{U} - \nu_{in}(\boldsymbol{U} - \boldsymbol{V}) + \frac{\mu}{\rho}\nabla^2\boldsymbol{U} + g \qquad [6]$$

This equation is split into several terms. ∇P is the pressure gradient. The Coriolis force ($2\Omega \times \mathbf{U}$) is due to the rotation of the Earth, with angular velocity, Ω, and horizontal winds of velocity, $\boldsymbol{U}$. The term $\nu_{in}(\boldsymbol{U}-\boldsymbol{V})$ is the ion drag for an ion and neutral collision frequency of ν_{in}. The viscosity, μ, of the atmosphere also produces a force, $\mu\nabla^2\boldsymbol{U}/\rho$, where ρ is the density of the gas.

At high latitudes, solar heating produces antisunward winds over the poles. In the auroral regions the magnetosphere couples with the ionosphere. Consequently, ions are dragged, at a velocity $\boldsymbol{V}$, with the magnetic field lines in a sunward direction due to magnetospheric convection. This produces the two-cell convection pattern in Figure 1.4b. Therefore, in the auroral ovals an eastward zonal wind occurs in the morning sectors (i.e. local magnetic midnight to midday) and westward winds in the afternoon and evening sectors (magnetic midday to midnight). These opposing flows converge at the cusp around local magnetic midday and at the Harang discontinuity at midnight. Through collisions and transfer of momentum between ions and neutrals, the antisunward flow of neutral winds becomes distorted and tends to follow the two-cell convection pattern of the plasma, according to the strength of the ion-neutral coupling.

Particle precipitation in the auroral ovals can also produce vertical thermospheric winds. The precipitation causes heating of the ionosphere, mainly at E-region altitudes, which in turn heats the thermosphere. This warmer gas therefore expands, and pushes gas above it upwards, in what is known as upwelling. When the precipitation ceases and the lower atmosphere cools, the atmosphere returns to its original state via downwelling. As a large mass of gas needs to be moved in this process, only small movements are possible, and the vertical winds are an order of magnitude smaller than the horizontal winds, which can reach speeds of 200-300ms^{-1}. Vertical motions of the atmosphere can also be caused by horizontal advection (see for example Hines et al., (1965)) where collisional interactions cause increases in ionisation.

Joule heating is an important mechanism for transferring energy into the upper thermosphere. It is calculated from the Pedersen conductivity σ_p at height z in the presence of an electric field $\boldsymbol{E}$, in the frame of reference of the Earth:

$$JH = \int \sigma_P \boldsymbol{E} dz = \int \sigma_P \left(\boldsymbol{V} \times \boldsymbol{B} + \boldsymbol{U} \times \boldsymbol{B} \right) dz \qquad [7]$$

The electric field $\boldsymbol{E}$ is the sum of the magnetospheric dynamo (***VxB***) and neutral wind dynamo (***UxB***) for ion velocity ***V***, neutral velocity ***U*** and magnetic field ***B***. As the neutral wind, ***U***, cannot easily be measured, it is difficult to calculate ***UxB*** for all heights. Thus, the neutral wind dynamo is often assumed to be zero to simplify the calculation. The Joule heating can then be calculated simply from the electric field in the frame of reference of the neutral atmosphere ($\boldsymbol{E}'=\boldsymbol{VxB}$) and the height integrated conductivity, $\Sigma_P = \int \sigma_P \, dz$. The electric field can be taken outside the integral as it does not vary with height as it can be assumed constant along a magnetic field line. Both of these values can be obtained directly from radars (Aruliah & Griffin, 2001).

1.2.4 ATOMIC OXYGEN EMISSIONS

Emissions are produced from many atmospheric constituents in the infrared and ultraviolet as well as the visible. Photochemical reactions, such as radiative recombination and ionisation reactions, produce emissions constantly, at all latitudes, which are called airglow. For the dayglow, these reactions are obscured by solar radiation, which is many orders of magnitude stronger. At high latitudes, nightglow additionally includes the emissions produced by particle precipitation, known as the aurora.

There are two main auroral neutral emission lines that are both from atomic oxygen (OI). These are the $^1S_0$ – $^1D_2$ transition which is at 5577Å, and the $^1D_2$ – $^3P_{2,1,0}$ transition at 6300Å, which have lifetimes of 0.91s and 110s respectively. This is shown in Figure 1.7 below.

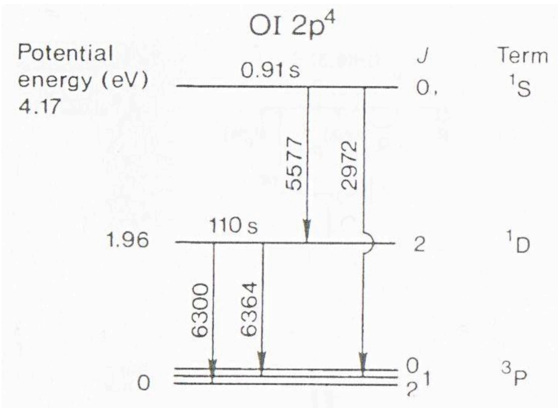

Figure 1.7 *Energy level diagrams showing the electronic transitions and emissions of atomic oxygen (Rees, 1989).*

The 5577Å line is the strongest auroral emission and is most commonly observed. However, the height of the emission is not well defined, with peaks at around 100km and 300km. Therefore, 6300Å observations are mostly used in this thesis as they have a well defined emission height at 240km (see below).

Several production methods of 6300Å emissions (Rees & Roble, 1986) are possible:

i) Dissociative recombination of O_2^+ ions:

$$O_2^+ + e^- \rightarrow O + O(^1D)$$

ii) Electron impact by energetic auroral electrons

$$O + e^* \rightarrow O(^1D) + e^-$$

iii) Electron impact excitation by electrons in the high energy tail of the ambient thermal population

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

iv) Electron impact dissociation of O_2

$$O_2 + e^* \rightarrow O + O(^1D)$$

v) Cascading from higher electronic states of the atom

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

vi) Reaction with nitrogen – production of nitric oxide

$$N(^2D) + O_2 \rightarrow NO + O(^1D)$$

Dissociative recombination (i) is dominant in the production of 6300Å emissions above 280km, (vi) is dominant below 200km, and all others are minor contributors. (v) is also a production mechanism for other lines, notably the green line at 5577Å and in the infrared. The main production method of the 6300Å emission viewed by the FPIs should therefore be from atomic oxygen created by the dissociative recombination of O_2^+. This is investigated in Chapter 3.

The FPIs measure the 6300Å emissions in one degree along a line of sight. They therefore measure the total line of sight integrated intensities. It is therefore important to know from what height the emissions originate. The rule of thumb (Makela et al., (2001)) is that they occur at one scale height below the F2 peak at 290 – 300km (hmF2). The scale height at these altitudes is around 40 – 50km, and the OI peak is taken to be 240km. The width of the emission band in altitude is also one scale height, 50km. The hmF2 altitude however changes with long-term factors such as seasons and solar variability, and geomagnetic activity will push the peak down by 10 – 20km (e.g. Sica et al., 1986). See also discussions in the results sections.

1.3 SMALL SCALE STRUCTURE

Small-scale structure can easily be seen in electron aurorae, through arcs that can be less than 1km across (Kivelson & Russell, 1989) and which move rapidly. SIF has the spatial and temporal resolution to detect this while also providing spectral information. Proton aurorae are also observed, though they do not vary on particularly small scales. This is because can emission sources can spread far from the original collision point, creating diffuse aurora. The precipitating protons can combine with atmospheric electrons several times to produce hydrogen atoms that are not confined to magnetic field lines and so can move horizontally. However, successive collisions of the protons with atmospheric particles produces secondary electrons, and hence electron aurora. Variations in the emission intensities of electron aurorae along the spectrograph slit (8°) are detectable to show small spatial variations in the aurora. The photometers and video complement this by providing high time resolution measurements of auroral features.

Studying the small-scale structure of the thermosphere is important both to better understand the processes, such as the energetics and dynamics of the atmosphere, and to improve models of the atmosphere so that they can better predict physical quantities under different conditions. The energetics of ion – neutral coupling is not fully understood, and this is because small-scale variations of order of tens of kilometres and tens of minutes are not usually considered. The thermosphere is often considered on scales of thousands of kilometres, i.e. assumed constant over the fields of view of most ground-based instruments (of hundreds of kilometres or less). It is therefore important to measure the thermosphere on small scales so that it can be understood over the same ranges as the ionosphere, and the structure within an instrument's field of view can be observed.

The thermosphere is usually assumed to have little or no small-scale structure due to its large inertia. This would make it slower to respond to changes than the much more dynamic ionosphere, which is true over large scales, both spatially and temporally, but not over smaller scales. Thermospheric winds are assumed constant up to periods over at least an hour with atmospheric tides occurring with periods of a few hours (see section 1.4). Variations also occur over longer time scales due to seasonal and solar cycle dependences (e.g. Aruliah & Rees, 1995). There may be little variation for quiet geomagnetic conditions, but this is not the case for active conditions where the ionosphere does influence the thermosphere. Measurements with FPIs do show variations on scales of tens of kilometres and tens of minutes (e.g. Aruliah & Griffin, 2001). As well as the differences that can be seen between the look directions of the FPIs, variations can be seen within a night in an individual set of data. Variations in the neutral intensities, temperatures, and winds have been detected that have been caused by acoustic and atmospheric gravity waves. The improved detectors now available for the FPIs have allowed measurements to be made at very high time resolutions. This will allow the physical limit of variability of the thermosphere to be calculated if this is larger than the resolution of the data, see Chapter 4.

Joule heating is an important energy transfer mechanism for the upper atmosphere. It is a frictional heating from the interaction of the ionosphere with the thermosphere, but often the neutral wind dynamo is ignored (see equation [7]). Aruliah et al., (2004) have shown that the neutral wind dynamo can contribute 30% to the total Joule heating at F-region heights. In this case, the amount of energy transferred from the ions to the neutrals will often be overestimated and hence the neutral temperature will be underestimated. This also results in more of the initial energy input from the magnetosphere being transferred to potential energy and kinetic energy, and so the momentum transfer and ion velocities will be overestimated.

Small-scale variations in the electric fields are also extremely important because they contribute as much to Joule heating as the average electric fields (Codrescu et al., 2000). It is therefore important to understand the small-scale variability so that models, such as MSIS (Mass Spectrometer and Incoherent Scatter empirical model, (Hedin, 1983)), with which it is not currently possible to address this structure, can be improved to better predict atmospheric values especially under geomagnetically active conditions.

1.4 ATMOSPHERIC GRAVITY WAVES

1.4.1 INTRODUCTION

Atmospheric gravity waves (AGWs) are perturbations in the motion, density, and temperature of the atmosphere, where the restoring forces are pressure and gravity. The first theoretical work on AGWs was by Hines, (1960). For a parcel of fluid that has been displaced from equilibrium, oscillations will occur in the atmosphere where gravity pulling the parcel downwards is balanced by the buoyancy of the atmosphere restoring the parcel back up to the equilibrium position. The majority of measurements in the upper atmosphere of AGWs have not been directly of the AGWs, but of the ionospheric response to gravity waves, Travelling Ionospheric Disturbances (TIDs). As an AGW passes, the movement of the neutral atmosphere drives the ionosphere plasma along the magnetic field lines, so that the wave can also be monitored in the plasma measurements, and are observed as small perturbations in ion and electron densities, temperatures and drifts. These are seen in ionospheric measurements made by a variety of instruments.

Gravity waves are an important mechanism for the transfer of energy and momentum between the upper and lower atmosphere. They can transfer energy from low to higher altitudes and from polar regions to low latitudes. They also affect the zonal wind flow patterns.

There are several mechanisms by which gravity waves can be created, and these vary with location. In low altitudes, they have tropospheric sources and then propagate up to mesospheric heights. There are a great variety of possible mechanisms, which include air passing over topographical features; air disturbances from thunderstorms, jet streams, volcanoes, wind shears, and weather fronts; and air temperature changes from solar eclipses. An example of gravity waves formed from air rising over mountain ranges is shown in Figure 1.8. Air forced to rise as it passes over mountain ranges will set up gravity waves in the lee of the mountain, under stable conditions. Waves are stationary with respect to the mountain and have zero horizontal phase velocity. As their energy source is near the ground, their vertical phase velocity and energy propagate upwards into the mesosphere.

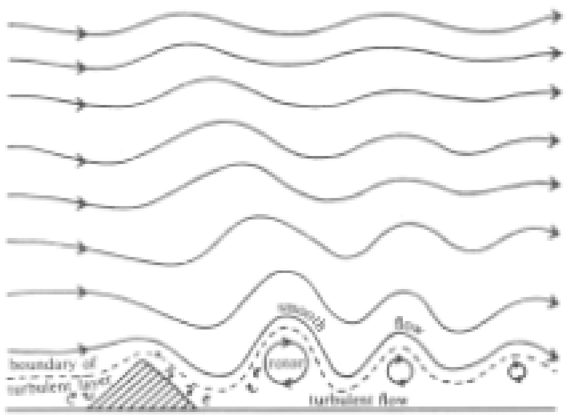

Figure 1.8 *Schematic of gravity waves formed in the lee of a mountain range (Houghton, 2002).*

The gravity waves will not propagate upwards indefinitely, at some height they will dissipate (see section 1.4.2). Tropospheric gravity waves will not be able to propagate up to thermospheric heights. High altitude AGWs therefore have to have sources at higher altitudes, and there is much evidence that they are auroral in origin (e.g. Williams et al., 1993, de Deuge et al., 1994, Bristow et al., 1994). These can propagate from high latitudes as far as equatorial regions.

There are two main source mechanisms of aurorally generated gravity waves. Large-scale gravity waves can be created from the effects of geomagnetic disturbances such as heating due to particle precipitation, or from Lorentz forces and Joule heating from electrojet currents (de Deuge et al., 1994) or rapid motions of the aurora (Oyama et al, 2001). Particle precipitation deposits energy directly into the thermosphere, so the change in pressure this causes leads to the formation of gravity waves. Alternatively, electrojet currents in the E-region ionosphere create Lorentz forces which produce Joule heating and hence gravity waves (de Deuge et al., 1994). An initial surge in the electrojet will generate a single pulse and this will lead to an oscillation through the effects of interference, filtering, and dispersion. This however would only lead to a few cycles of the wave, and cannot explain the twelve or more cycles that have been observed in TIDs (Lewis et al, 1996). This suggests a periodicity in the source itself. Observations of waves from these two sources are described in section 1.4.4.

Gravity waves have been studied extensively in the mesosphere and lower thermosphere (MLT) regions, for example Williams et al., (1993). AGWs have been observed in the upper thermosphere too, over the southern polar cap, for example by de Deuge et al., (1994) and Innis et al., (2001) in photometer observations of the 6300Å oxygen emissions. The theory of gravity waves is discussed in the next section, followed by their observed features and how they are classified. The application of this to the FPI data, and data from other instruments, is covered in Chapters 4 and 5.

Atmospheric tides are also present in the thermosphere, and lower atmosphere, and these are global in nature. Diurnal variations are excited directly in the thermosphere by solar heating. This produces tides with periods of 24 hours, the diurnal tide, and the semi-diurnal of 12 hours, the terdiurnal tide of 8 hours and the harmonics of 6, 4 and 3 hours. Whereas the diurnal tide is predominantly formed from direct heating, the semi-diurnal tides are excited by reflections of solar heating from the surface and propagate upwards (Hargreaves, 1979; Rees, 1989). The predominant tides are migrating – i.e. they vary with local time and migrate westwards with the Sun. Non-migrating tides also exist and have fixed phases with respect to the surface of the Earth. These can be formed, for example, by Joule heating in the thermosphere.

There are also a variety of waves with periods longer than a day, and these are called planetary waves. The most common is the quasi 2-day wave (‘q2dw’) with a period of around 2.3 days, and others occur at approximately 5 days and 15 days. These are also global waves, and they are caused by changes in the Coriolis force with latitude. Studies of tides and planetary waves are not made in this thesis as the instruments used here are only in a relatively small part of the globe. All waves with periods between 6 and 24 hours are neglected, as it cannot be distinguished whether these are tides or gravity waves. In addition, only night-time measurements are made, so data sets long enough to study planetary waves are only available when there are several successive clear nights, which is not a common occurrence.

As gravity waves are a mechanism for transporting energy and momentum from one part of the atmosphere to another, they are an important part of modelling work. However, for gravity waves of tropospheric sources, individual waves are often not modelled, as the waves are much smaller than the grid spacing of the model. For this reason, gravity waves are parameterised, which means a spectrum is input at the source region, then the transmission of the wave is calculated from effects such as turbulence and molecular diffusion. Gravity waves dissipate where the phase velocity equals the background winds (see below). For tropospheric sourced gravity waves with typical velocities, this height is in the mesosphere, at a maximum of around 85km. Therefore, only lower atmosphere (stratospheric and mesospheric) gravity waves are included in parameterisations, as waves of tropospheric sources are assumed to have dissipated before reaching higher altitudes. An example of the effects of gravity waves is that they can increase drag and so decrease eastward winds at equatorial latitudes and mesospheric altitudes. Without the inclusion of gravity waves in the models, large eastward equatorial winds are created, which do not agree with the measurements of these mesospheric winds (England et al., 2006).

The other type of gravity wave that can be included in atmospheric modelling is large-scale waves. Waves of sufficiently large horizontal wavelengths will cover several grid points, and their propagation can be tracked around the globe. Using a global thermosphere – ionosphere model Millward et al., (1993) tracked the propagation of a gravity wave from an auroral source down to mid latitudes, and analysed its variations with altitude. Balthazor & Moffett, (1997) found, using a global coupled thermosphere-ionosphere-plasmasphere model, that atmospheric gravity waves that have conjugate sources in the northern and southern auroral ovals propagate to the magnetic equator, where they constructively interfere, and pass through to the other hemisphere with an unchanged velocity. In the F region, where the amplitudes of the waves are larger, the TIDs increase in amplitude at the equatorial regions.

1.4.2 THEORY

Much work has been done on the theory of acoustic-gravity waves. Books that cover the mathematics and properties in detail include for example Hargreaves, (1979, 1992), Kato, (1980), Gill, (1982), Lindzen, (1990), Houghton, (2002), Nappo, (2002). There are a wide range of papers on theory and observations of gravity waves in many forms. This chapter provides an overview of the main points that are relevant to this thesis. Further work not covered here has been done on, for example, gravity wave parameterisations used to include gravity wave effects in atmospheric models. Due to the scale sizes of gravity waves, the effects of the Earth's rotation and curvature can be neglected. These however are important for atmospheric tides and planetary waves.

Gravity waves propagate through the atmosphere, which can be treated as a fluid, so general wave equations will apply. For a gravity wave with wavelength ℓ, the wavenumber is

$$k = \frac{2\pi}{\ell} \qquad [8]$$

In the x, y (horizontal component) and z (vertical) directions the wavenumbers are k, l and m, respectively:

$$k = \frac{2\pi}{\lambda_x} \qquad l = \frac{2\pi}{\lambda_y} \qquad m = \frac{2\pi}{\lambda_z} \tag{9}$$

The wave frequency for a wave with period τ is

$$\omega = \frac{2\pi}{\tau} \tag{10}$$

For a wave described by $A\cos(kx-\omega t)$ with amplitude A, the phase of the wave is given by the phase angle, ϕ, which for the two-dimensional case is:

$$\phi = \mathbf{k}.\mathbf{r} - \omega t = kx + mz - \omega t \tag{11}$$

The wave then has two velocities, the phase speed, c_x and then group velocity, v_g. The phase speed is the velocity of a point of constant phase in the direction of propagation of the wave. The group velocity is the more important and is the velocity at which energy is transported away from the source of the wave.

$$c_x = \frac{\omega}{k} \qquad v_g = \frac{\partial \omega}{\partial k} \tag{12}$$

These values can be seen in Figure 1.9 below.

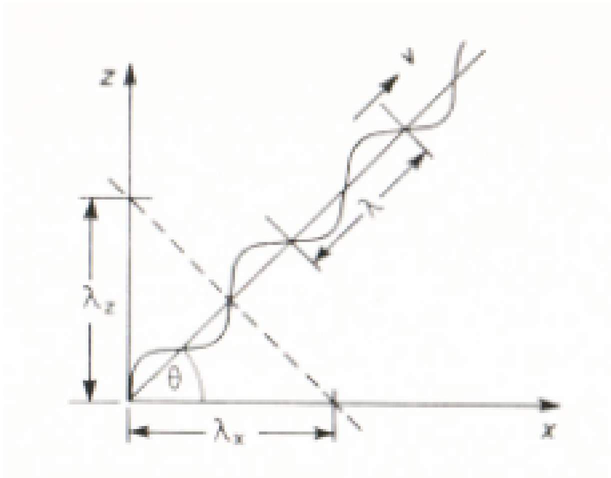

Figure 1.9 *Relationships between wavelength, velocity and propagation angle of gravity waves (Hargreaves, 1992).*

To determine the constraints on gravity waves, they need to be considered as a balance of pressure and gravity on a perturbation away from equilibrium. The equation for hydrostatic equilibrium, $\partial P/\partial z = -\rho g$ (equation [1]) shows that the rate of change of pressure with altitude is a balance of the density of the atmosphere and gravity. Assuming an ideal gas, the equation of state is $PV = nRT$ (equation [2]). The ratio of specific heat capacities at constant pressure, C_P, and constant volume, C_V, is $\gamma = C_P/C_V$. The speed of sound is $c = (\partial P/\partial \rho)^{1/2} = (\gamma RT/M)^{1/2}$. This can lead to the scale height, H, (equation [4]) being re-written (Hargreaves, 1992) as equation [13].

$$H = \frac{c_s^2}{g\gamma} \qquad [13]$$

Here c_s is the speed of sound, and equation [13] shows that this is another quantity that varies with scale height. The dispersion relation for the waves, equation [14] (Hargreaves, 1992), relates the wavenumber to the frequency.

$$\omega^4 - \omega^2 c_s^2\left(k_x^2 + k_z^2\right) + (\gamma - 1)g^2 k_x^2 + i\gamma g\omega^2 k_z = 0 \qquad [14]$$

This is the two-dimensional dispersion relation, which is sufficient as there is no horizontal asymmetry in the wave. There are two solutions to this equation as k_z is complex. For a real value of k_x and an imaginary value of k_z, there is no vertical propagation and these are external or surface waves, for example ocean waves. They only occur where there is a discontinuity in density and the wave will propagate horizontally. If k_z is complex, $k_z = i/2H$, and the dispersion relation becomes

$$\omega^4 - \omega^2 c_s^2\left(k_x^2 + k_z^2\right) + (\gamma - 1)g^2 k_x^2 + \frac{\gamma^2 g^2 \omega^2}{4c_s^2} = 0 \qquad [15]$$

Real k_x and real k_z values therefore create a quadratic solution in ω. These are internal waves, and the two roots of ω^2 give two limits.

$$\omega_a \geq \frac{\gamma g}{2c_s} \qquad \omega_B \leq \frac{(\gamma - 1)^{1/2} g}{c_s} \qquad [16]$$

ω_a is the acoustic cut-off frequency and ω_B is known as the Brunt-Väisälä frequency. These limits are shown schematically in Figure 1.10. Acoustic waves are longitudinal waves, where the restoring force is pressure. Their phase propagation is in the direction of the group velocity and energy flow and they have periods of a few minutes at most. The Brunt-Väisälä frequency, given by $\omega_B = (\gamma-1)^{1/2}$ g/c, or in terms of scale height $\omega_B^2 = (\gamma-1)g / \gamma H$, is the natural resonance of a displaced parcel of air with a restoring force of buoyancy. This is the maximum frequency possible for a gravity wave. Their propagation properties are described below.

The Brunt-Väisälä frequency in equation [16] can be understood from considering (Hargreaves, 1992) a parcel of air, initially in equilibrium at a pressure P_0 and density ρ_0, that is adiabatically raised by a height δz to an area of surrounding pressure of $P = P_0 + \delta P$ and density $\rho = \rho_0 + \delta\rho$. The parcel, of mass m has new values of pressure P' and density ρ' giving an upward force of buoyancy of $mg(\rho - \rho')/\rho' =$

$mg((\rho - \rho')/\rho_0$ as $(\rho - \rho') << \rho_0$. As the surrounding air is in hydrostatic equilibrium over a scale height (H):

$$\frac{\rho}{\rho_0} = 1 + \frac{1}{\rho_0}\frac{\partial \rho}{\partial z}\delta z = 1 - \frac{\delta z}{H} = 1 - \frac{g\gamma}{c^2}\delta z \qquad [17]$$

As the parcel of air expands adiabatically,

$$\frac{\rho'}{\rho_0} = 1 + \frac{1}{\rho_0}\frac{\partial \rho}{\partial P}\delta P = 1 + \frac{1}{\rho_0}\frac{\delta P}{c^2} \qquad [18]$$

Therefore, the upward acceleration of the parcel of air, due to buoyancy, and using $\partial P/\partial z = -g\rho$ is:

$$\frac{d^2 z}{dt^2} = -g\left(\frac{g\gamma}{c^2}\delta z + \frac{1}{\rho_0}\frac{\delta P}{c^2}\right) = -g\left(\frac{g\gamma}{c^2} - \frac{g}{c^2}\right)\delta z = -\frac{(\gamma - 1)g^2}{c^2}\delta z \qquad [19]$$

Since the acceleration is proportional to $-\delta z$, the motion of the parcel is simple harmonic and it therefore has an angular frequency of $(\gamma - 1)^{1/2}\, g/c$, which is the Brunt-Väisälä frequency of equation [16]. The range of this frequency for the atmosphere is shown in Figure 1.10 below.

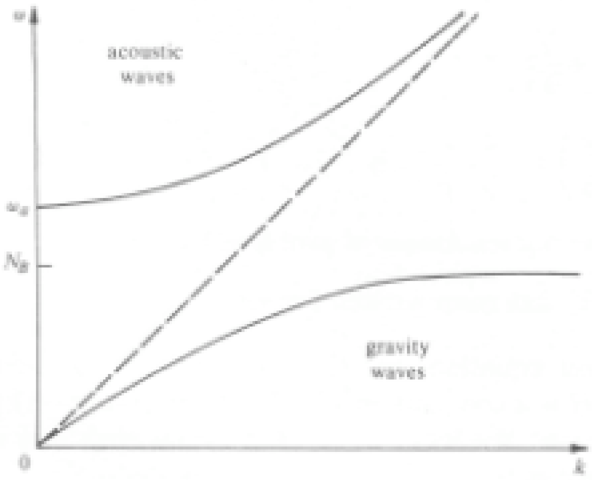

Figure 1.10 *Dispersion curves for gravity waves, showing the frequency ranges of acoustic and gravity waves. N_B is the Brunt-Väisälä frequency (ω_B). The dashed line is a wave with the speed of sound. (Houghton, 2002).*

The dashed line in Figure 1.10 is the velocity of sound; acoustic waves are at frequencies greater than ω_a, and gravity waves are below ω_B. The regions are well separated as $\omega_B < \omega_a$. Waves between the two solid lines are evanescent, and so only if k_x and k_z are real can the waves propagate, i.e. above or below the solid lines in Figure 1.10. At F2-region heights, ω_B is around 0.008rad/s, and the speed of sound, c_s is 0.014rad/s. this means waves with periods greater than 10-15 minutes will be gravity waves, those less than that will be acoustic (Hargreaves, (1979), and see section 1.4.3).

Gravity waves will propagate away from their source region into layers of atmosphere with different density. This will cause their properties to change. Gravity wave propagation is shown in Figure 1.11.

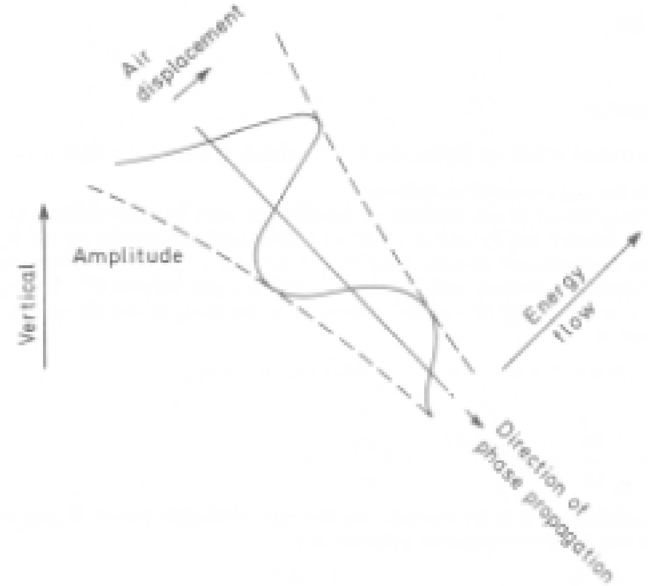

Figure 1.11 *Propagation in a simple gravity wave. Energy flow is perpendicular to phase propagation, (Hargreaves, 1979).*

Air particles move perpendicular to the direction of phase propagation and energy travels perpendicular to the phase velocity. To conserve energy, the amplitude of the wave increases with altitude. The wave's kinetic energy per unit volume is $1/2\ \rho v^2$, where v is the wave velocity amplitude, and the density, ρ, is given by $\rho_0\ e^{-z/2H}$. Therefore, as density decreases with increasing height, v^2 must decrease to conserve energy and so $v \approx v_0\ e^{-z/2H}$. Fritts & Alexander (2003) calculated the vertical wave number, equation [20], for a wave propagating vertically, in three dimensions.

$$m^2 = \frac{\left(k^2 + l^2\right)\left(N^2 - \hat{\omega}^2\right)}{\left(\hat{\omega}^2 - f^2\right)} - \frac{1}{4H^2} \qquad [20]$$

The frequency of the gravity wave is ω, and $\hat{\omega}$ is the intrinsic frequency, i.e. the frequency in the frame of reference moving with the background wind. Here f is the Coriolis effect, $f = 2\Omega\ \sin\theta$, where θ is latitude. For a vertically propagating gravity wave, k, l, and m are real values. Also, the intrinsic frequency is limited to $N > \hat{\omega} > |f|$. Vertical propagation is dependant on the horizontal wavelength and the scale height at the altitude of the wave. The intrinsic frequency and therefore the vertical wavenumber, *m*, can vary significantly as the wave propagates.

Gravity waves will not propagate indefinitely in a real atmosphere. They will dissipate over time, due to several reasons that are complications to the theory outlined above. The most significant of these is energy dissipation due to the viscosity of the atmosphere. This limits the distances over which gravity waves can transport energy. Energy dissipation is more important when the wave reaches higher altitudes because the amplitude of the waves is larger so the waves are no longer small variations. Reflections can occur where the refractive index of the atmosphere changes due to the vertical temperature structure. This leads to ducted waves, described further below. Other factors that affect the propagation of gravity waves include changes in velocity due to background winds, which can change the direction of the waves and lead to ducting.

Waves propagate vertically upwards, conserving momentum. They will dissipate at the altitude where the phase velocity equals the background wind velocity. However, they will break at an altitude below this if the perturbation from the wave results in the parcel of air having a higher temperature than the background atmosphere. This will result in a warm parcel of air in colder surroundings, so this will be convectively unstable, as the parcel will expand to match the surroundings. The wave amplitude cannot increase above the altitude where this occurs, due to turbulent diffusion. Momentum is therefore deposited, and the wave will dissipate.

1.4.3 CLASSIFICATIONS AND PROPERTIES

As gravity waves of tropospheric origin do not propagate up to thermospheric heights and they are not measured by the instrumentation used for this thesis, they are not considered further here. Atmospheric gravity waves at thermospheric heights are often classified into medium and large-scale waves (e.g. de Deuge et al., 1994, Hocke & Schlegel, 1996) from their velocity and period. Medium scale waves have velocities between 100 and 250m/s, wavelengths of several hundred kilometres and periods between 15 and 60 minutes (Hunsucker, 1982). Large-scale waves have horizontal velocities of 400 – 1000m/s, horizontal wavelengths of over 1000km and periods from 30 minutes to a few hours. Typical amplitudes of wave velocities are a few cm/s in the troposphere, a few m/s in the stratosphere and a few tens of m/s in the mesosphere and thermosphere. Waves moving with the background winds and at the same velocity will dissipate, decreasing the number of waves detected in this direction. Observations show that propagation from the source region is either uniformly spread or in zonal directions. Long period waves can travel further before being dissipated, and so can cover many thousands of kilometres. Shorter period waves generated by auroral sources mostly will not survive to low latitudes.

Modelling by Mayr et al., (1990) identified how gravity wave propagation can be complicated by effects in the middle atmosphere. Variations in winds with height lead to refraction, reflection, and ducting of the waves, and local wind variations will change the phase speed of the waves. Some waves with auroral sources can travel large distances from the source, and for example are seen at mid-latitudes as TIDs, propagating equatorwards. Waves with a short vertical wavelength are confined to the high latitudes. Ducting of the waves can occur because they are reflected by layers in the atmosphere created by refractive index changes with height.

The velocities and periods of possible gravity waves are dependant on the speed of sound of the atmosphere they pass through. This varies with altitude through the atmosphere. At ionospheric heights, waves with periods greater than 10-15 minutes will be gravity waves, those less than that will be acoustic, as can be seen in Figure 1.12. This provides a limiting value on the periods of gravity waves in the data. At 240km, the Brunt-Väisälä period is taken to be 12 minutes and the speed of sound as 720m/s (de Deuge et al., 1994). For the purposes of this thesis, the upper limit on the period is defined by the presence of tides, as these are not easily distinguishable from gravity waves using only a localised data source with the small proportion of the globe that is viewed by the Scandinavian instruments. This limit is therefore taken as the 6-hour wave.

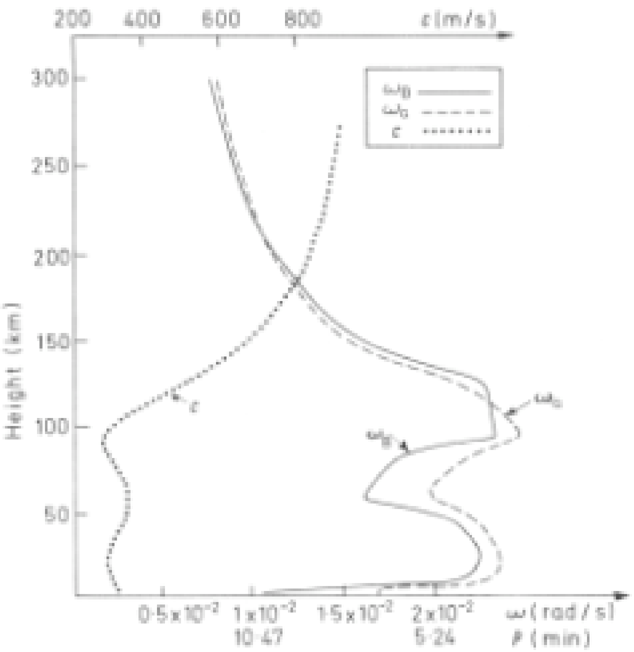

Figure 1.12 *Height profiles of the speeds of sound (c), the acoustic cut-off frequency ω_a and the Brunt-Väisälä frequency, ω_B. (Hargreaves, 1979).*

As the local speed of sound varies with altitude, properties of the gravity waves and TIDs also vary with altitude. For example, shorter horizontal wavelengths in electron density variations (350km from models, Hock & Schlegel, 1996) are favoured at altitudes of approximately 220km, but by around 350km longer wavelengths have nearly as large amplitudes. Relative phases between waves in different ionospheric parameters also change with height. The change in phase of ion temperature follows that of ion velocity, as these are closely related to the phases of the neutral temperature and velocity. Below 250km, the phases of ion and electron temperatures are very similar, showing strong thermal coupling between neutrals, ions and electrons. Above 300km, the electron temperature phase is in anti-phase with the electron density, because the electron temperature fluctuation is induced by the electron density fluctuations, which is consistent with the cooling of the electron gas by collisions with ions and neutrals in the F2 region.

1.4.4 Observations

Gravity waves have been seen in many data sets. See, for example, the clear gravity waves in Figure 1.13 in all sky camera data. This shows atomic oxygen and OH airglow emissions over Hawaii observed during the ALOHA-93 campaign (Taylor et al, 1995). Assuming an emission altitude of 100km, speeds of the structures were calculated to be 20 and 43m/s (Nappo, 2002). Waves have been seen and modelled (Fritts & Alexander, 2003) in noctilucent clouds at mesospheric altitudes. More often however, TIDs are measured in ionospheric data. TIDs can often be seen in radar electron densities, in both incoherent scatter radars such as EISCAT and coherent scatter radars such as CUTLASS and the SuperDARN network (see section 2.4). Other instruments that measure electron densities, such as ionosondes, also can observe gravity waves.

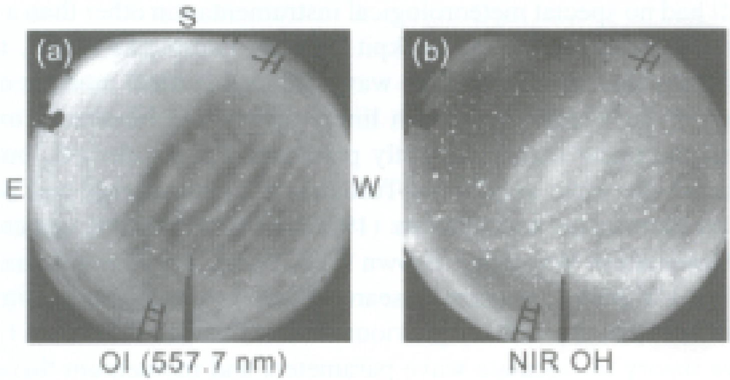

Figure 1.13 *Gravity waves in all sky camera images of the green line and OH airglow, from Haleakala, Hawaii. (Nappo, 2002).*

Typical properties of gravity waves were obtained from the World-wide Atmospheric Gravity-wave Study (WAGS) campaigns in the 1980s (e.g. Williams et al., 1993). This was a series of four campaigns to investigate the global propagation of AGWs and TIDs by co-ordinating world-wide measurements. Locations of Joule Heating and Lorentz forcing were found as clearly defined events, with a quasi-periodic structure of variations. The events were followed by TIDs with similar periodicities by various instruments. It was found that the relative TID amplitudes were proportional to that of the auroral electric field. Most measurements were from AGWs from the westward electrojet, as these are easier to measure as they propagate during night-time and so are less attenuated by ion drag. During the first WAGS campaign, the Dynamics Explorer satellite showed the source of the AGWs as two substorms 75 minutes apart. Radars across the northern polar region measured periodic variations in various parameters with periods from 60 to 90 minutes.

From these various measurements, typical values for gravity waves at thermospheric heights can be estimated. Periods range from the Brunt-Väisälä period (around 12 minutes) to several hours (tides are also present at 6 hours). Horizontal wavelengths range from hundreds to thousands of kilometres. Vertical wavelengths are of the order of hundreds of kilometres, and vertical wave speeds are of the order of hundreds of metres per second. Medium scale TIDs have been found to tend to propagate zonally; large-scale waves mostly propagate south-westerly.

Most of the observational work on gravity waves has been on middle atmosphere waves. In addition, most observations are of TIDs rather than the true thermospheric gravity waves, as the ionosphere is easier to measure so there are many more instruments monitoring the ionospheric parameters. There is however evidence that gravity waves can be seen in the upper thermosphere. For example, AGWs have been observed in the upper thermosphere over the southern polar cap in photometer observations of the 6300Å oxygen emissions (de Deuge et al., 1994 (see below) and Innis et al., 2001). It is reasonable to expect, therefore, that gravity waves should be seen in the FPI data presented here. These will increase the currently limited number of observations of gravity waves in the upper thermosphere, and they will provide a view of the gravity waves in the northern hemisphere to complement the current measurements in the south.

Optical observations were made of gravity waves in the atomic oxygen red (6300Å) and green (5577Å) line emissions over Mawson, Antarctica (67.6°S, 62.9°E) by de Deuge et al., (1994). They used a photometer with a 3-channel filter wheel (also including the N_2^+ blue line at 4286Å) and with three fields of view, at 4.7° from zenith, giving separations of 31km for the 6300Å and 17km for the 5577Å, with 13km and 7km field diameters respectively. This limited their observable gravity waves to horizontal wavelengths, λ_h, of twice this at most, i.e. λ_h = 62km. APL's FPIs have fields of view that will allow horizontal wavelengths of several hundred kilometres, or if Svalbard and mainland data are compared, up to 3000km. However, Deuge et al., (1994) had a high time resolution, 18s, so the shortest gravity waves could be measured; the upper limit is the length of their data sets at 4 hours. The FPI data sets do not have this limitation as they record all dark hours of the night, so longer period waves can be observed. To find gravity waves formed from structures in auroral precipitation as suggested by Lui et al., (1987), they only considered periods of data without intense auroral activity. They did not de-trend their data but pointed out that this will not change properties such as frequency and phase of the waves, and may only affect their relative power. They only measured intensities of the emission lines and the main limitation of this is that intensities only show the progression of the phase fronts. For gravity waves, this is not the direction of energy propagation. A major advantage of the FPIs is that they measure the thermospheric temperatures and winds concurrently with the emission intensities.

De Deuge et al., (1994) performed Fourier transforms of their data, which allowed calculations of wave speeds and horizontal direction and wavelengths. Their results confirmed theoretical predictions that upper thermospheric gravity waves (6300Å observations) are large-scale gravity waves, and 5577Å results at 120km are medium-scale gravity waves. They found that waves tended to propagate to the northwest and southeast, which at the location of Mawson is equatorward and poleward from the generation in the auroral electrojet. Northward and southward propagation of gravity waves in Scandinavia would therefore be expected, if the same source mechanism were responsible for the gravity wave generation.

De Deuge et al., (1994) looked at a few individual case studies. In contrast, five years of southern polar cap photometric data were searched for gravity waves by Innis et al., (2001). Case studies of individual nights of FPI data are discussed in Chapter 4 and Chapter 5 shows a statistical survey of many years of the FPI data. De Deuge et al., (1994) find gravity waves in 42% of their valid nights. These are chosen as those that have clear skies and where the auroral oval is northward of their site (Davis, Antarctica) as they are only interested in polar cap gravity waves. They use the same 3-field photometer as de Deuge et al., (1994), but Davis is nearer the southern magnetic pole, at co-ordinates of 68.6°S, 78.0°E. Observations are again of the atomic oxygen red line at 6300Å. Gravity waves would be expected in roughly half of the FPI data sets, at least in Svalbard, which is also in the polar cap region. However, for the FPI analysis of many nights (see Chapter 5), all the clear nights are used, including those where the auroral oval is poleward or overhead Svalbard, so a greater percentage of data sets showing gravity waves would be expected.

Innis et al., (2001) also find the biased direction of propagation that de Deuge et al., (1994) find. The majority of the waves were found to travel towards 130°, which at Davis is towards the southern magnetic pole. As a point source would be expected to produce circular wave fronts, they attribute this effect to planar waves from an elongated source region. This could be achieved from a source along the electrojet, or a point source spread in local time. Innis et al., (2001) also do not consider the effects of background winds in their calculations. Gravity waves can be formed from a moving source region, and the phase velocity can be an indication of the speed of the wave source.

All previous ground based measurements of upper thermospheric gravity waves have been photometric and so have only measured the emission line intensity. Velocities and temperatures have only been measured from satellites. Johnson et al., (1995) analysed Dynamics Explorer 2 (DE2) satellite data and found 50 F-region polar cap gravity waves. These had vertical velocities of 100m/s and temperature changes of about 100K. Innis and Conde, (2001) studied the vertical winds from the WATS (Wind and Temperature Spectrometer) on the DE2 satellite. They find enhanced vertical wind velocities in the polar cap regions, and attributed them to AGWs. They also find that the source region is mainly the post-midnight to dawn sectors of the auroral oval. Gravity waves propagating poleward were found to travel further than those propagating equatorward, which would be due to either the fact that they are travelling approximately parallel to the background winds, or to ion drag.

A further complication to gravity waves in wind data is discussed by Lilensten & Amblard, (2001). They compared tristatic EISCAT data with WINDII (Wind Imaging Interferometer) wind data from the UARS (Upper Atmosphere Research Satellite) (described by Sheperd et al., (1993)). Their WINDII data showed that the oscillations they found were a neutral and not an ionospheric phenomena, so they suggest gravity waves could be formed from eddies in the neutral winds. To test this, they suggest using interferometer data in the same set up as the tristatic EISCAT data. This setup is described in section 3.2. If eddies are the source of the waves, the data would show the oscillations with a $\pi/2$ phase difference between the meridional and zonal components of the neutral wind. Gravity waves on this type of data (i.e. tristatic FPI data) are shown in sections 4.4 and 5.3.

In Chapter 4, FPI data will be analysed for gravity waves. This is the first detection of atmospheric gravity waves in thermospheric FPI data, and the first observations of gravity waves in neutral temperatures. It is also the first time gravity waves have been observed in the northern polar cap and auroral oval thermosphere. The thermospheric gravity waves are compared to TIDs and to electric currents. This allows the source and the mechanism for production of the gravity waves to be identified. In Chapter 5, the typical properties of gravity waves are examined by comparing statistical samples of nights, and their variations with parameters such as season, solar cycle and geomagnetic activity are compared.

2 INSTRUMENTATION

2.1 INTRODUCTION

The main data used in this thesis are from Fabry-Perot Interferometers (section 2.2), which measure meso-scale structures in the thermosphere. The Atmospheric Physics Laboratory (APL) has four Fabry-Perot Interferometers, located at Sodankylä, Finland (67.4°N, 26.6°E); KEOPS (Kiruna Esrange Optical Site) (one for the red line at 6300Å and one for the green 5577Å line); Sweden (67.8°N, 20.4°E) and at the Auroral Station in Adventdalen on Svalbard in the Arctic Ocean (78.2°N, 15.6°E) (Aruliah & Griffin, 2001). The University of Lancaster also has an FPI in Skibotn in northern Norway (69.3°N, 20.4°E), which allows us to take tristatic measurements with the three FPIs in northern Scandinavia (see section 3.2).

As the auroral oval and polar cap regions are complex, dynamic, and important regions, many instruments have been installed to observe the various aspects of the coupled magnetosphere, ionosphere, and thermosphere at these latitudes. This is especially true of the Northern Scandinavian sector, which is easily accessible (compared with, say, Antarctica), and has good facilities for installing these instruments.

The other instruments run by APL are on the Spectrograph Imaging Facility (SIF) in collaboration with the University of Southampton, see section 2.3. This is co-located with the Svalbard FPI in the Adventdalen Auroral Station. The main other data set used is from the EISCAT (European Incoherent SCATter) radars, described in section 2.4.1. The tomography chain which calculates electron densities is described in section 2.4.2 and is correlated with FPI data and compared with EISCAT data in chapter 3. Section 2.4.3 discusses the MIRACLE (Magnetometers, Ionospheric Radars, All-sky Cameras Large Experiment) suite of instruments. As well as the instruments in northern Scandinavia, there are several satellites that have data coverage of this region. These are discussed in section 2.4.6.

2.2 Fabry-Perot Interferometers (FPIs)

2.2.1 Introduction

Fabry-Perot Interferometers (FPIs) were developed by Fabry & Perot, (1901). They have been used for astronomical purposes for many years. They are pertinent for measuring the upper thermosphere for two reasons. The 6300Å airglow and aurora are weak sources, so a high light gathering power is needed to detect them. Maximum intensities of the 6300Å emission are of the order of 50kRayleighs. This is also the reason why observations are restricted to night times, as these low intensities are easily swamped by daylight. FPIs have a high sensitivity as the interference is due to many reflections of the light, compare with, for example, a Michelson interferometer that only uses two-beam interference. Also, a high spectral resolution is needed to be able to measure the thermospheric winds – winds of 200m/s produce a Doppler shift of 4.2mÅ at 6300Å emissions (Aruliah, 1991). The first measurements of the Earth's atmosphere using an FPI were made by Babcock, (1923).

The instrument consists of a Fabry-Perot etalon, imaging optics, a filter, and a detector. These are discussed in the following section, along with the calibration used and the analysis of the data to obtain emission intensities, wind velocities, and temperatures. For a more thorough description, see e.g. Hernandez (1986), Born and Wolf (1987), Hecht (2001).

APL currently owns and runs four FPIs. Two of these are located at KEOPS near Kiruna in northern Sweden. One measures the auroral red line at 6300Å, the other the green line at 5577Å. They have been running since 1999 at KEOPS, but before that an earlier version ran at the Institute for Space Physics (IRF) in Kiruna for nearly 20 years. The FPI in Sodankylä, Finland, also measures the red line and was installed at the Sodankylä Geophysical Observatory (SGO) in November 2002. These are situated in the auroral oval area, and are co-located with the EISCAT receiver dishes (see section 2.4.1). The EISCAT transmitters are at Tromsø in Norway (69.66°N, 18.94°E) and APL has a collaboration with the University of Lancaster who runs an FPI near Tromsø, at Skibotn (e.g. Kosch et al., 1997). These sites are discussed further in the tristatic section 3.2. The final FPI is located in the polar cap region at Svalbard, Spitzbergen, and is near the EISCAT Svalbard Radar (ESR).

The FPIs can be controlled remotely. The instruments can be monitored and data can be collected automatically throughout the dark season (September to April) where the Sun is below the horizon for significant periods. Data have been collected from Kiruna every winter from 1981. An FPI has been running in Svalbard over the winters of 1982-1984 and since 1997. Green line tests have been made at various times, and data were collected in KEOPS over the winters of 2000-2001 and 2004-2005. The Sodankylä FPI was installed at the beginning of the 2002 season.

2.2.2 THE ETALON

Incident light enters the etalon by way of a scanning mirror (see section 2.2.3). The etalon consists of two circular mirrors made of fused silica, of a diameter of 15cm, that are partially reflecting on the inside surfaces. This material must be of a high quality to have a uniform refractive index and be thermally stable. The surfaces are very flat, to within approximately a 100th of the wavelength of the intended observations. The mirrors have a slight wedge shape to reduce interference from the outer edges by removing this interference away from the optical axis of the instrument. This stops Fabry-Perot interference occurring within the glass. The two plates are parallel, at a fixed distance, *d*. This distance is kept constant by three spacers, made of a glass called Zerodur. This has very low thermal expansion, which is important to keep the distance constant, as any variations will change the path difference and the position of the interference rings on the detector. These spacers are adjusted to obtain parallelism of the plates.

The whole etalon is kept in a vacuum, so there is a constant refractive index, μ, in the space between the plates, and pressure variations due to ambient temperature changes are minimised. This means the plates need to be contained in a sealed housing, which is made of an anodised aluminium alloy with glass windows and must have tight seals to prevent leaks and consequent changes in pressure.

The inner surfaces of the plates are coated with a dielectric, which allows light entering normal to this surface to be partially reflected, to a very high degree (typically around 90%). Constructive and destructive interference of light, of a wavelength λ, as determined by Bragg's Law, equation [21], occurs in the etalon gap due to the multiple reflections the light undergoes.

$$2d\cos\theta_n = n\lambda \qquad [21]$$

Where *n* is the order of reflection and is an integer, and θ_n is the angle of interference, at an order *n*. The optical path difference (OPD) is the difference in path length between successive reflections. Constructive interference occurs for path differences *n*□ and destructive interference at (*n* + 1/2)□. This can be seen in Figure 2.1 below.

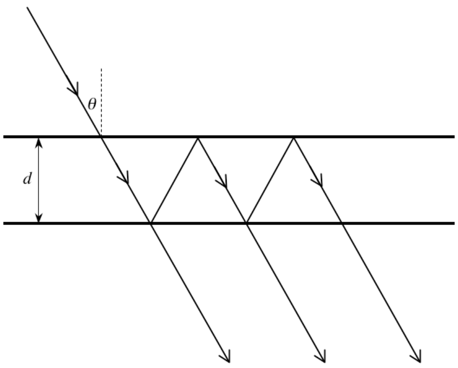

Figure 2.1 *Constructive interference in a Fabry-Perot Etalon.*

The interference fringes produced are circular as constructive interference occurs at constant angles (□), i.e. at equal inclinations. Typical fringes are shown in Figure 2.2.

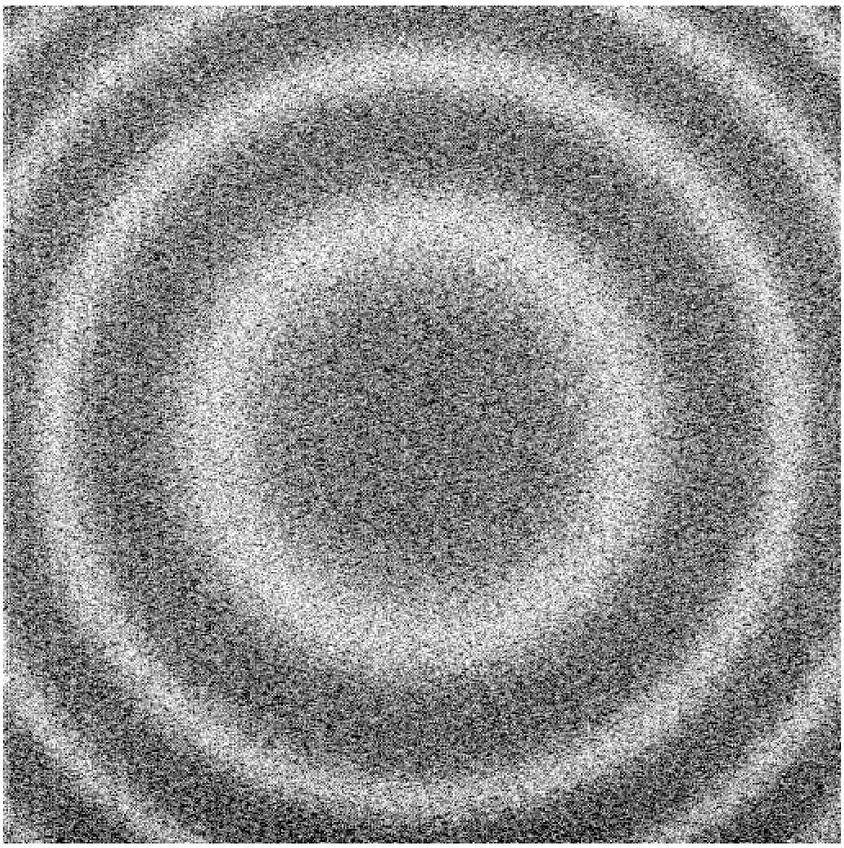

Figure 2.2 *Circular interference pattern of the sky, through the etalon.*

The angular diameters of the rings are given by θ_n as given by equation [21]. To obtain this image in Figure 2.2, a detector is needed, and further optics are required to focus the image onto the detector. These are discussed in the following sections.

For a wavelength interval of $\Delta\lambda$ between transmission peaks for a set gap, the free spectral range (FSR) is where the interference pattern repeats. The diameter of the rings is a function of the wavelength, from equation [21] and Doppler shifts of the winds cause a change in wavelength and so the angular diameter of the rings will change. Interference repeats at $\lambda+\Delta\lambda$ for the next order, therefore

$$2d\cos\theta_n = (n-1)(\lambda + \Delta\lambda) \qquad [22]$$

At the point of repetition, this equation is equal to equation [21], and therefore

$$n\Delta\lambda = (\lambda + \Delta\lambda) \qquad [23]$$

Substituting equation [21] gives the free spectral range:

$$\Delta\lambda = \frac{\lambda(\lambda + \Delta\lambda)}{2d\cos\theta_n} \qquad [24]$$

This may be simplified however for light incident near the normal, where cos $\theta \approx 0$, and also for small shifts so that $\Delta\lambda << \lambda$. These approximations lead to:

$$\Delta\lambda \approx \frac{\lambda^2}{2d} \qquad [25]$$

Therefore, the distance between the plates, *d*, defines the free spectral range for the wavelength observed. The gap size chosen is a trade-off, as a smaller free spectral range will give greater accuracy but it also needs to be large enough to fit the emission line peak profile on the detector. The free spectral range is chosen so that there is no aliasing, so the free spectral range is greater than the largest possible Doppler shift. A gap distance of 14mm or 18.5mm is used for the UCL FPIs that measure the 6300Å (and green line 5577Å) emissions, 14mm gives a value of 14pm for the free spectral range.

The light gathering power of an optical instrument is determined by a factor called etendue, which is the product $A\Omega$. A is the light collecting area of the instrument, determined by the diameter of the etalon; and Ω is the solid angle of the field of view. This angle is determined by the gap between the plates (equation [21]), which is itself chosen to give a free spectral range appropriate to observe the broadened emission line studied. The product $A\Omega$ is large for the FPI, which is the reason for it being used in observing the weak auroral emissions, where as many photons as possible need to be collected.

The transmitted intensity, I, is given by the Airy function, equation [26]:

$$I = \frac{1}{1 + F\,\mathbf{sin}^2 \frac{\Delta\phi}{2}} \qquad [26]$$

Where $\Delta\phi$ is the phase change between reflections in the etalon, given by:

$$\Delta\phi = \frac{\pi\mu d}{\lambda_0}\cos\theta \qquad [27]$$

For surface reflectivity R, the coefficient of finesse, F, is defined by:

$$F = \frac{4R}{(1 - R)^2} \qquad [28]$$

The finesse of the interferometer is a useful quantity as it is the measure of the sharpness of the fringes. It is the ratio of the separation and the half width of the fringes (Born & Wolf, 1987). The finesse has several contributing sources. At the half width, the intensity, of equation [26], is half that at the maximum. $\Delta\delta/2$ is sufficiently small that the sine term may be approximated by the angle. The separation of fringes corresponds to a phase change of 2π, so that the ideal reflectivity finesse, N_R is given by equation [29].

$$N_R = \frac{\pi}{2}\sqrt{F} = \frac{\pi\sqrt{R}}{1-R} \qquad [29]$$

The total, or effective, finesse (N_{eff}) is also dependant on the flatness, aperture, and parallelism finesses. These need to be made as large as possible, i.e. by using highly flat glass, etc., so that the reflective finesse is near to the effective finesse. The effective finesse is therefore in practice not quite as high as is given by equation [29] due to, for example, surface defects or warps on the etalon and the two plates not being completely parallel. Vaughan, (1989) describes the effects of different types of defects in detail.

2.2.3 The Interferometer Setup

Figure 2.3 below shows the setup of the FPI, as used by APL. An optical bench, consisting of four parallel rods held together with fixed brackets, holds the components of the FPIs in place. These are suspended in shafts, of varying accessibility, or straight under the dome, depending on the site. The tops of the shafts are covered with clear domes, made of Plexiglas, which are heated to prevent condensation and to melt away snow that lands on them. At the bottoms of the shafts are small rooms that contain the electronics and computers to control the FPI and the cooling system for the detector.

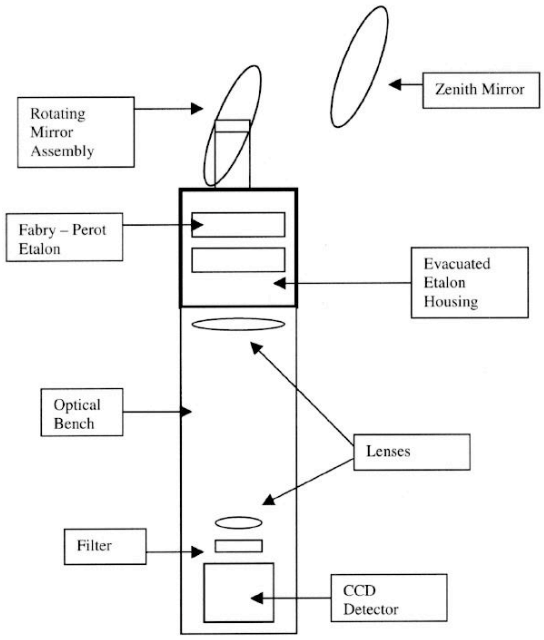

Figure 2.3 *A schematic of the configuration of the FPI. Light enters via the mirror at the top, passes through the etalon, the lenses, and the filter, and then is received by the detector at the bottom.*

Light enters the etalon by way of a tilted mirror that can be rotated to view different portions of the sky. This mirror is mounted on a platform on top of the optical bench (which is vertically mounted). The mirror is rotated by a motor that moves it to a preset position. The position is detected by an opto-sensor that reads coded reflective strips located on the underside of the mirror mount (see below). Measurements of the zenith position are taken either by rotating the mirror to point at an extra, fixed mirror angled to view the vertical, or by setting the mirror vertically to allow direct observations of the zenith (used in Svalbard). For the calibration (see section 2.2.5), the source lamp illuminates a screen that scatters the light evenly across the field of view.

The mirror is angled to view at 45° above the horizon for the mainland, and 30° for Svalbard. The FPIs measure aurora and airglow emissions from the 6300Å OI red line, which emits in the F2 layer, typically at about 240km altitude, see section 1.2.4. Therefore, for measurements at an elevation angle of 45^o the observations are approximately 240km from the observation site. The etalon has a 1° field of view, so at 240km altitude, an area of approximately 10x10km is observed. The Skibotn FPI has a more complicated mirror adjustment mechanism that allows the elevation angle to be controlled as well as the azimuth direction. The elevation angle is currently set to 51.5° to coincide with the KEOPS and Sodankylä fields of view for simultaneous tristatic measurements of a common volume (see section 3.2).

Each cycle of the mirror consists of several look directions: north, east, south, west and the zenith, as well as an exposure for a neon calibration lamp (see section 2.2.5) that measures the stability of the instrument. Depending on the site and the science needs, exposures are also taken at the tristatic A and bistatic B positions (see Figure 3.1) for the mainland FPIs for the most recent winters. North-East and North-West are sometimes included for mainland FPIs (KEOPS and Sodankylä) to increase the resolution in the auroral oval, which is to the north of these sites. Similarly, South East and South West are included in Svalbard measurements.

The diameter of the etalon (150mm for APL FPIs) defines the aperture of the instrument. The light from the etalon has to be imaged onto the detector. In older FPIs this was achieved by mirrors in a Cassegrain telescope arrangement as they allow a shorter optical bench to be used. Current FPIs use two lenses to focus the image onto the detector in the form of a Galilean telescope. These are a positive and negative lens and have a combined focal length of 1.2m. The two lenses allow the bench to be a metre long.

The last component of the FPI before the detector is the filter. This is placed between the secondary lens and the detector. The filter is required to limit the bandwidth of the incoming light and to thereby reduce the continuum levels so that the spectral line required can be observed and to suppress other unwanted orders from outside the passband. The bandwidth of the filters is 10Å, which is wide enough for the Doppler shifts of the line to be observed. For red line emissions, a filter at 6304Å is used. The FPIs have a 1-degree light cone and the filter response moves to a lower wavelength for off-axis light. A filter with a higher wavelength range, of 6302-6304Å, compensates for this.

2.2.4 DETECTORS

Various detectors have been used for the FPIs over the years. The first FPIs used photographic plates, then photomultiplier tubes, before moving on to the imaging capabilities of CCDs (Charged Coupled Devices) used today. Atmospheric observations using FPIs were transformed by the use of Imaging Photon Detectors (IPDs). A full description of IPDs is given by McWhirter, (1993). These can detect individual photons and can therefore make use of the good light gathering power that the etalon provides. Detection of individual photons is particularly important in atmospheric physics due to the extremely low light levels of the aurorae.

In a summary of IPDs, the incident photons hit a semiconductor photocathode, and each photon has the potential to cause one electron to be emitted. In practice, only a few percent do so. This percentage defines the quantum efficiency (QE) of the detector and the amount of incident photons that are detected. The detector needs to be cooled to prevent thermal electrons being emitted as well. Cooling is achieved by a Peltier cooler from which heat is removed by a water flow, which is pumped and cooled to around -5°C. Each photoelectron then passes through a micro channel plate (MCP). This is a grid of small (20μm diameter) lead-doped glass tubes, which creates a cascade of secondary electrons. Around three or more of these are used to increase the gain to around 10^7. These tubes keep the photoelectrons in the same relative position to where the photon initially struck. These secondary electrons are then collected by an anode, which is also position sensitive, and by this point there are sufficient numbers of electrons to be measured electronically. The x-y co-ordinates are determined by a ratio-determining circuit and the signal is digitised and sent to a computer for processing, as discussed in the next sections.

CCDs are very sensitive detectors made of silicon, which also generates electrons from incident photons. These are collected for the length of the exposure, held in the lattice structure of the silicon until the pattern is read out by electronics, and then sent to a computer. The main advantage of IPDs over CCDs is that they have no read-out noise and so CCDs may need longer integration times to obtain a signal above the read-out noise. CCDs are better detectors for point sources because their intensity response is linear; they have a much higher quantum efficiency than IPDs; and they have exact positional determination. IPDs are used here for detection of aurora and airglow, as these are very weak signals. IPDs are preferable to CCDs when it is necessary to keep time resolution to a minimum yet accumulate as many photons as possible.

CCDs have however been developed to improve this, in the form of Intensified CCDs (ICCDs). These are a merging of the technologies of IPDs and CCDs. As for an IPD, incident photons hit a photocathode, which produces a photoelectron. This then enters a micro channel plate to produce a cascade of secondary electrons. Rather than being collected by an anode, the electrons hit a phosphor, screen to emit photons, which are then detected by the CCD. The advantages of this are that the number of electrons hitting the phosphor is proportional to the number of incident photons on the photocathode. As the number of electrons reaching the CCD is much larger, there is a larger signal to noise ratio, and so shorter integration times can be used. The readout noise is not important, as only the position of the scintillation of the phosphor needs to be read. However, the Skibotn FPI has an ICCD made by EEV Ltd. (e2v Technologies) that uses a slow phosphor. Due to the persistence of the phosphor on the cathode, a pause time is necessary between successive images to prevent an afterglow effect from contaminating images. This, as well as a more complicated mirror mechanism, creates a lower time resolution for data from this site. The APL FPIs have used a variety of detectors over the years. The KEOPS FPI originally had an IPD. When this failed in 2001, the red line filter was moved to the green line (5577Å) instrument that had an Astrocam detector. These are bare CCDs from Astrocam with sensors from e2v Technologies. They are not intensified so reasonably long integration times were used, and hence the cycle times were slow and time resolution of the data was not high. Astrocam CCDs were also used on the Sodankylä and Svalbard FPIs.

The latest cameras are from Andor Technologies, and they use EMCCDs – Electron Multiplying CCDs, which are also back illuminated. An Andor camera was installed on the KEOPS FPI for the beginning of the 2004 season, and the Svalbard FPI in the 2005 season. EMCCDs have a solid-state electron multiplier built into the chip, rather than a separate intensifier. This consists of an additional serial readout register operated at a voltage high enough to cause impact ionisation, which creates a high gain when summed over the entire row. With a high gain, readout noise is not a problem so high time resolution can be maintained. The advantage of back-illumination is that it has a larger quantum efficiency. The front faces of the CCDs absorb some of the incident photons, which therefore do not produce photoelectrons, and are therefore not registered by the detector. For back-illuminated CCDs, the back of the silicon wafer (the substrate) is thinned, down to around 15μm, and the CCD is turned around. This thinned back is more transparent, and so the quantum efficiency can increase from 30-40% for front illuminated to 80 or 90% for back illuminated

CCDs. A list of the detectors on the instruments, and the data available, is given in Chapter 5.

2.2.5 Calibration

The temperature of the atmosphere may be calculated from the Doppler broadening of the emission. However, the line profile of the interference fringes, without any Doppler broadening, must first be determined. This is the instrument function. The shape of the fringe profile is measured by taking an image after each pressure change, using an effectively infinitely narrow source, i.e. one that has an unbroadened line width, much less than the FPI instrument function. A HeNe laser, with a stabilised power supply, is used for this purpose (see Figure 2.4 below).

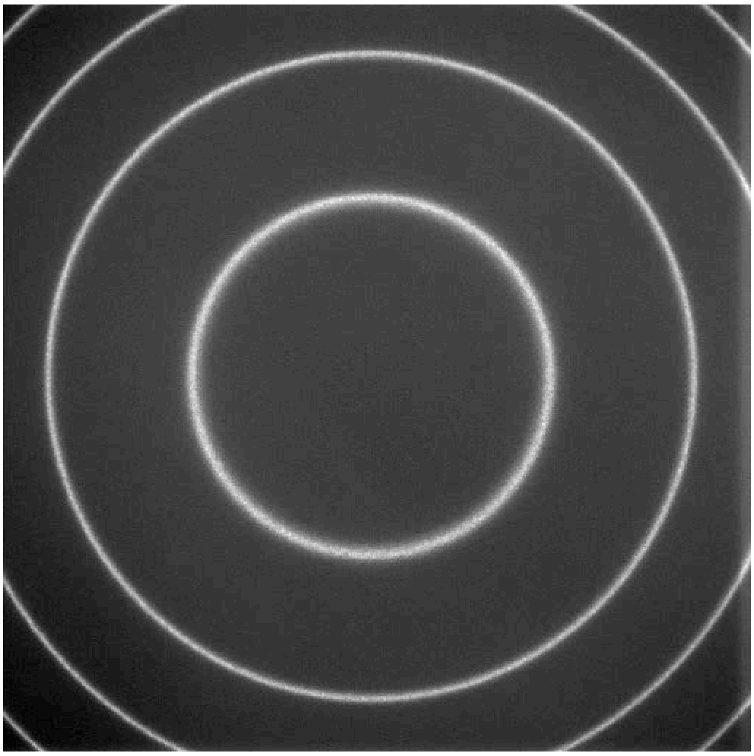

Figure 2.4 *Typical laser calibration image showing several narrow ring interference fringes.*

Comparisons of the laser calibration image (Figure 2.4) with the sky images (e.g. Figure 2.2) show that the calibration laser produces much narrower, well-defined fringes due to an intense coherent source. This can also be seen from the integrated fringes in Figure 2.5 (see below). As the HeNe laser emits at a different wavelength to the observations at 6300Å, the filter needs to be changed to take these calibration images. Laser calibration images are also used when the instrument is being set up and each season on maintenance and campaign trips to centre the image (i.e. make sure all the components are in line with the optical path) and to make sure that the rings are circular.

There are two methods for scanning the fringes to measure the instrument function. The method used for APL's FPIs, is to change the optical path difference by changing the refractive index of the air inside the gap. This is most easily done by changing the pressure of the air in a method known as pressure scanning. The distance between the plates must be kept stable. The pressure is incrementally changed to determine the instrument function at every possible fringe diameter. The pressure of the air in the gap is then adjusted for taking observations so that a suitable diameter of fringe is obtained where a full fringe is observed, and if possible two complete fringes.

The second method for determining the instrument function is used by the Skibotn FPI by the University of Lancaster and the Scanning Doppler Imager (SCANDI), an all-sky imaging version of the FPI. As the wavelength is a function of the distance d between plates (equation [21]), if this distance is varied, the fringe diameter will vary. The physical distance between the plates can be adjusted by small amounts by changing the length of the spacers. This can be done by making the spacers out of a piezoelectric material so that changing the voltage across the spacer changes its length, in a controlled manner. However, this is harder and more expensive to achieve than changing the air pressure, as the changes must be very carefully controlled and monitored to keep the parallelism of the plates. Capacitance micrometers are used to control the gap stability. These produce an output voltage proportional to the displacement, i.e. the spacer length, so that the gap can be accurately monitored. These etalons are therefore called capacitance stabilised etalons (CSEs) and are discussed further by McWhirter (1993). They are used in the next stage of instrument for APL; SCANDI. This is an all-sky FPI and is discussed in the future work section (6.4).

Wind speeds are determined using a calibration lamp image, taken after each cycle of observations. The 6300Å atomic oxygen line is a forbidden transition, with a long lifetime, and therefore needs low densities to allow that transition to occur. Therefore, lamps of exactly the right wavelength are difficult to produce and operate, and so are not viable options. A neon lamp is used to calibrate the atomic oxygen red line at 6300Å, since it has a line close to this, at 6304Å. As the calibration lamp is not at exactly the same wavelength as the emission line, this cannot be used to find the zero Doppler shift position, which is necessary to calculate the wind velocity. So to obtain the line of sight wind velocity, the offset of the calibration lamp from the vertical wind velocity is found. This assumes a zero vertical wind, which is generally acceptable as it is an order of magnitude smaller than horizontal winds (of order of tens of m/s rather than a couple of hundred m/s). The vertical wind over a 24-hour period should theoretically average to zero due to conservation of mass over this long timescale. Often however, 24 hours of data are not available, so the vertical wind is averaged over the night's observations. It may not average to zero, particularly in active conditions (Aruliah & Rees, (1995)). Avoiding this assumption is one of the main advantages of the tristatic campaign, where true vectors can be calculated (section 3.2). As well as calibrating the data, the measurements allow faults to be detected, such as leaks in the vacuum system or temperature variations that cause changes in pressure, which will change the optical path length.

Intensities of the emission line are not yet calibrated; as for this, the intensity of the calibration lamp must be accurately known. The neon lamps do not have a well-defined output so the intensity would have to be measured by an instrument that can be accurately calibrated, such as a photometer. Plans for this for the future are underway.

2.2.6 DATA PROCESSING

The image received from the detector is a series of concentric light and dark fringe interference rings, as in Figure 2.2. The intensity is the number of counts at the peak of a ring cross section, the temperature is calculated from the width of the peak, and the diameter of the ring is dependant on the Doppler shift and therefore gives the line of sight velocity. The FPI is scanned through different directions, to provide the different components of the wind vectors.

However, as the image is all of the same 1° portion of the sky, there is very little positional information in the rings, so nothing is lost by integrating around each ring and this also greatly improves the signal to noise ratio. The centre of the ring pattern is determined, and integrating around the circles produces a spectrum that has clear, sharp peaks, see Figure 2.5. This figure also shows the Doppler broadening due to temperature.

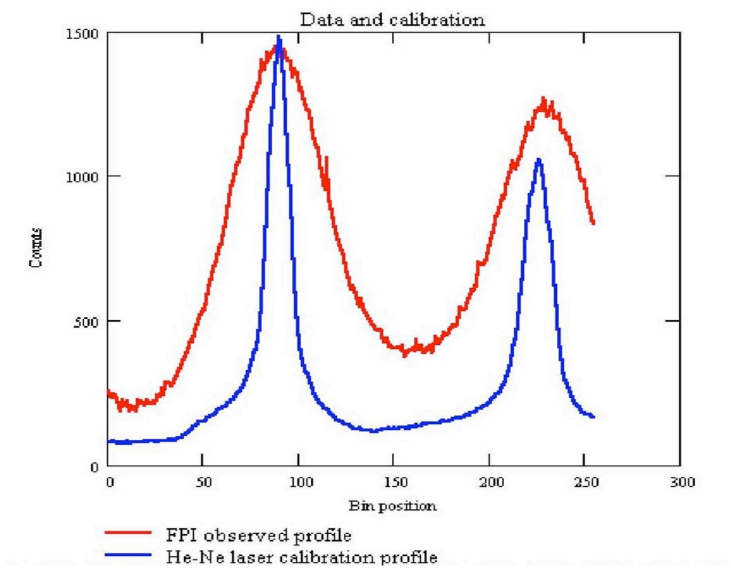

Figure 2.5 *FPI ring-integrated calibration and observation profiles.*

For small angles, (□) analysis of the Airy function shows that the wavelength is proportional to the squared radius of the ring peaks. Therefore, plots of peaks in radius squared space show equal distances between successive rings. A series of IDL programs have been developed to process and plot the data. Firstly, a peak-fitting program uses a Gaussian distribution to fit to the integrated ring profiles and is used to calculate the Doppler shift. Data with very poor signal to noise ratios are removed. As each of the look directions are slightly displaced in time, as measurements are taken around the cycle, for them to be compared each look direction is plotted then interpolated to 15-minute intervals. Wind speeds are calculated for the original times of observations as well as the 15-minute intervals. These data are stored as binary data, but can be converted to text files for analysis.

A final program exists that flags possible cloudy data, assumed to be those periods where the intensity peaks coincide for all look directions. Another indication of cloud in the data is where time series plots of opposite directions (i.e. north and south or east and west) have the same shape curves, but are mirror images of each other. A large gradient in intensity is common in the auroral oval region, as only one look direction is likely to be observing the enhanced precipitation. With a large intensity gradient and scattering from clouds, the Doppler shifts will scatter to the other look directions.

The exposures are typically 60 seconds, and with the time taken for the mirror to change positions, this leaves up to 15 minutes between measurements in the same direction. The 15-minute averaging is now too long for latest KEOPS data, which has exposures as short as 10s with the new improved EMCCD, so this needs to be revised.

2.3 SPECTROGRAPH IMAGING FACILITY (SIF)

2.3.1 INTRODUCTION

The Spectrograph Imaging Facility (SIF) is located in Adventdalen, near Longyearbyen, on Svalbard (78.2°N, 15.8°E). The primary reason for SIF being located at this site is to view the cusp regions, which were discussed in 1.2.2. It also provides optical support for ESR, the EISCAT Svalbard Radar (see section 2.4.1), which is nearby at the end of the Adventdalen valley. SIF is co-located with several other instruments at the Auroral Station on Svalbard, such as one of UCL’s Fabry-Perot Interferometers, all-sky cameras and a magnetometer run by the University of Tromsø in Norway, a meridian scanning photometer (MSP) and many others. Svalbard is also in the CUTLASS radar field of view. Both electron and proton aurora over Svalbard can be measured with good spatial and temporal resolutions through SIF and complementary instruments. It is therefore used to study small-scale structures such as auroral arcs, as well as for the more diffuse proton aurora studies through the hydrogen Balmer series □ line.

SIF consists of a platform with, currently, six instruments mounted. The instruments on the platform are co-aligned and the platform itself is aligned along the magnetic meridian. The primary instrument is a spectrograph, which has an 8° slit. Field-aligned photometers each have a 1° field of view, and a video camera (TLC) covers 12° by 16° around the magnetic zenith. From installation in November 1999 there were two photometers, then an additional two were installed in November 2004. The platform is housed in a dome similar to that of the FPIs. This dome is also heated with a fan heater to keep the dome clear from moisture and ice. Cables for the instruments are fed to computers in an adjacent control room.

2.3.2 INSTRUMENTS

The spectrograph itself is a High Throughput Imaging Echelle Spectrograph (HiTIES) (manufactured in Boston by Baumgardner et al., 1993 and see McWhirter et al., 2002) with a wavelength resolution of about 1.3Å. The optical configuration is shown in Figure 2.6 below. Light enters through a fixed objective lens with a 300mm focal length, then passes through the slit. The length of the slit is fixed at 45mm, which corresponds to 8° in the sky with this optical arrangement. The width of the slit is adjustable, from 0.14mm (0.0249° or 1.49') to 1mm (10.7'). The usual operating width is 0.27mm, corresponding to 2.88 arc minutes or 2'53". This width is chosen to get a balance between resolution and throughput of photons. The beam is then collimated with lens L1 so that parallel light hits the grating. An Echelle grating is used in high orders producing a high spectral resolution. The angle of the grating can be adjusted to select the required diffracted wavelength.

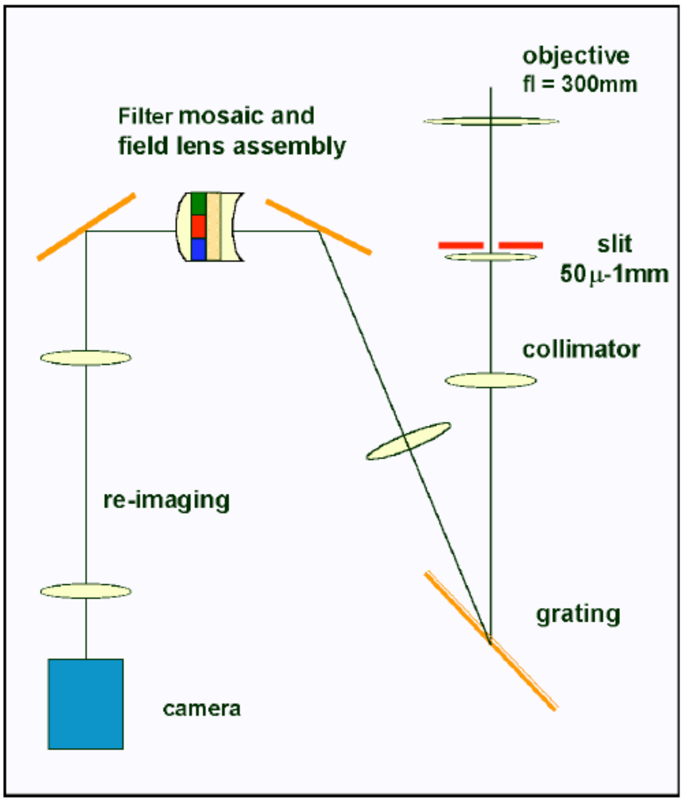

Figure 2.6 *The optical configuration of the SIF spectrograph showing the components and light path (McWhirter, 2002).*

The beam is then focused with lens L2 to image on the filters via a mirror. The spectrograph uses a 50mm square filter mosaic so that multiple emission lines can be studied simultaneously. The filter panels select different orders of diffraction. The filter assembly contains a concave, field flattening lens, a field lens, and the filter mosaic. The spectrograph has two filter mosaics, a 3-band and a 4-band (Figure 2.7) that must be manually interchanged, so this can only be done on campaigns. The 3-band has 3 strips of filter, one for the Hβ line at 4861Å covering a range of 4844Å – 4876Å, and one each for two N_2^+ rotational lines. These are the first negative band lines, the (1,3) line at 4651.8Å (a range of 4635Å – 4660Å) and the (0,2) 4709Å line from 4690Å – 4715Å (Lanchester et al., 2001 and Rees, 1989). These emissions were chosen because the Hβ is a signature of proton aurora, while the nitrogen lines are due to electron aurora, but are near in wavelength to the Hβ line. The nitrogen bandpasses also contain rotational lines from the molecules, which could be used to calculate the temperature by comparing the relative intensities of the different rotational bands, and comparing to theoretical models.

The mosaic is usually orientated so that the length of each band is along the slit direction, so the full spatial scale of the slit is used, but this can also be rotated 90° to get a larger spectral range which can be useful to observe more of the background tails of the Hβ line, at the expense of spatial resolution. This is also beneficial when studying proton aurora, as this does not have as much spatial structure across the field of view, so in the rotated orientation the maximum spectral structure can be seen. However, with the original orientation, all the spectral lines cover all the spatial scale of the slit, so that the lines are observed simultaneously from the same patch of sky. When the filter mosaic is rotated, this property is lost, and the lines will originate from spatially separate sky regions. While studying electron aurora (for instance when EISCAT radar runs in a high resolution mode) where fine structure along the spatial length of the slit is needed, the filter block is rotated back to its original position.

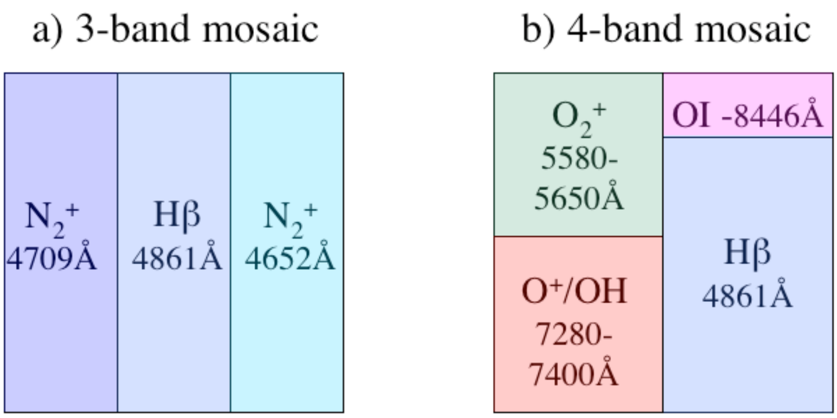

Figure 2.7 *The SIF filter mosaics: a) the 3-band and b) the 4-band mosaic showing emission lines observed and central wavelengths or wavelength bands.*

The 4-band mosaic has the same H filter as the 3-band mosaic, and three oxygen lines. The H panel is larger than the other three because the line is wide due to proton aurora production processes. The smallest panel at the end of the H is a near-infrared atomic oxygen line at 8446Å. A third panel is for a molecular oxygen ion (O_2^+) from 5580Å – 5650Å. The final panel is at 7280Å – 7400Å and is for the O^+ line at 7327Å but also includes several fainter OH lines. These are prominent emissions after the 5577Å and 6300Å and they are used to study the ionospheric species.

After passing through the filters, the beam is then focused with lens L3 (Figure 2.6) onto the detector, a CCD camera. Several detectors have been used at various stages of operation of SIF. The main detector is a water-cooled CCD camera (WCCCD), a PixelVision camera cooled to 225K by using a Peltier cooler, itself cooled by flowing water. Until the winter of 2002 the camera was run at 230K as the water had to be kept above zero so that ice did not form in the water tank. Adding glycol, an antifreeze, to the water allowed the temperature to be reduced to -10°C. This meant that the Peltier cooler could reach a temperature of 220K. As these were roughly the limits of the system, the water is now kept at -8°C and the CCD at 225K. This decrease in temperature has the advantage of improving the signal to noise ratio as dark current is reduced. The camera pixels (of an array of 1100 x 1050) are usually binned to increase sensitivity (McWhirter et al., 2002), at the expense of resolution. Other detectors were used at times, mainly due to failures with the WCCCD. An intensified CCD was used for 3 months at the end of 2000. A microchannel plate intensified CCD (the MIC, Fordham et al., 1990; 1991) was used over the winter of 2001 – 2002 and a month at the end of 2003. The MIC is an improved ICCD, and is more sensitive than the WCCCD, but can only be used on campaigns, as the controlling computer cannot be run remotely.

Two photometers were originally installed to support the spectrograph with high time resolution data of 0.2 seconds (adjustable, down to 50ms) (McWhirter et al., 2002). Lenses image a 1-degree field of view onto a photomultiplier tube (PMT). There is no spatial resolution across the field of view but the photomultiplier has a high light gathering power so very high time resolutions are achievable. The 1° field of view of the photometers is field-aligned and at the centre of the spectrograph slit. They have filters for the $H\beta$ and 4652Å N_2^+ lines, the same as two of the 3-band spectrograph mosaic filter. In November 2004, two additional photometers were installed. One of these has a filter for just off the $H\beta$ at 4880Å ($H\beta$ is at 4861Å), which gives a background level for the $H\beta$. The fourth photometer has a different type of PMT to the other three, a gallium arsenide detector, which is more sensitive to red light. This PMT has to be cooled, which is achieved with a Peltier cooler. For the first campaign where this was installed it had a N_2^+ (1,4) filter at 5149Å, but after the campaign this was changed to an OI filter at 7774Å.

The final instrument on the platform is TLC; an intensified video camera that provides a wider context for the other instruments, as for example the passage of auroral features across the spectrograph field of view can be seen. The field of view is 12° by 16° and is aligned with the slit north – south. North and south, and east and west are reversed though due to the optics and the mounting. The video has a cut-off filter at 6450Å which has the purpose of eliminating the red line emissions as these have a half life longer than the frame time (25 frames per second) and so would blur over several frames and saturate the image so that structured aurora would not be seen. Data are recorded on SVHS videos at six times slower than the speed of real time, allowing 18 hours of images to be recorded on a tape, allowing over-night recordings. On a campaign basis, a digital recorder can record onto 3-hour digital tapes, though due to their expense at this time these are only used for radar runs and periods of auroral activity. Tapes also need to be changed manually so the lack of remote control prevents them being used continuously at this time. A frame-grabbing program on a PC also allows individual images to be saved digitally.

SIF also has a flat-field lamp (light from a bulb is scattered onto a screen so it is uniformly illuminated) and a hydrogen lamp that emits at the Hα wavelength. Dark exposures can be taken by covering the fronts of the instruments with lens caps. This allows the thermal noise of the detector to be measured and subtracted from data. The flat field lamp evenly illuminates the CCD so that variations in sensitivity over different pixels can be measured. Data images are then divided by averages of flat field images. Flat fields and dark exposures are also taken, usually daily, on campaigns (the flat field lamp needs to be turned on, and the lamp or lens caps must be placed over the instruments manually, so these cannot be done remotely). These are then used for the rest of the season. The hydrogen lamp image is used to centre the CCD position onto the mosaic filter, to focus the camera, and to calibrate the wavelength of the emission lines.

SIF is designed to be run remotely via the internet, and the spectrograph software can be set to take exposures, of varying time periods as determined by light levels. Exposures are mostly taken at 30 seconds, but when more dynamic features are being studied this can be reduced (to as short as around 10s before the signal is too small for a sufficient signal to noise ratio).

2.3.3 Alignment

To get the most out of SIF, all the instruments on the platform need to be accurately co-aligned, and the platform needs to be well aligned onto the magnetic field line, which at Svalbard is at an azimuth of 183.0° and an elevation of 82.1°. This was done when the platform was first installed in Svalbard, but it also had to be re-done when the extra two photometers were added in November 2004. This opportunity was used to improve the accuracy of the alignment. This process is described in Ford et al., (2005) and is summarised in this section.

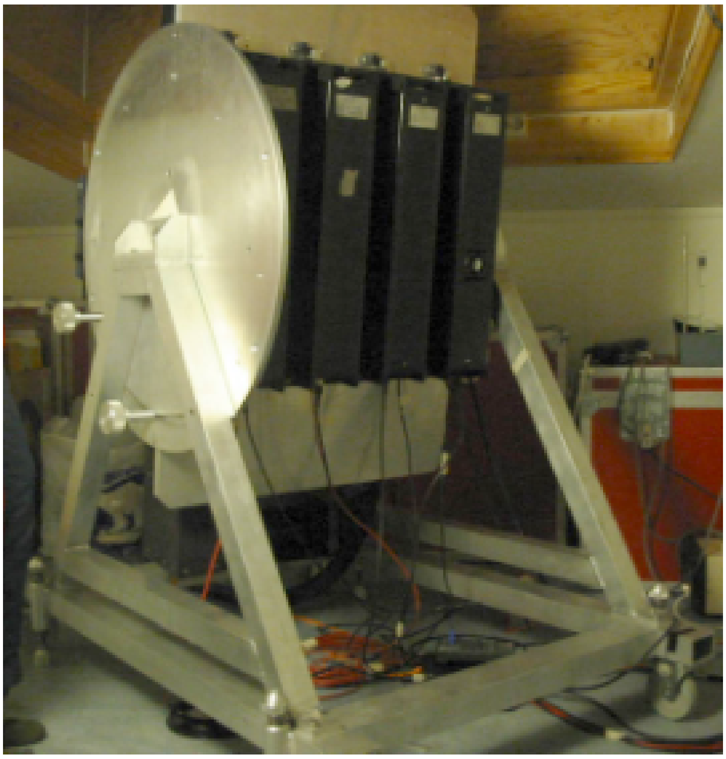

Figure 2.8 *The SIF platform in its dome in Svalbard, showing the four photometers. The spectrograph and TLC are on the other side of the back plate.*

All the SIF instruments are mounted on a platform consisting of a flat plate on a frame. Four feet support the platform (see Figure 2.8) with wheels for movement and feet that can be screwed down to lock the platform in position (i.e. in azimuth). The elevation angle of the instruments can be changed by adjusting the angle of the plate with respect to the platform frame. Accuracy is only possible in changes in elevation to within 1°, due to the weight of the platform and the locking mechanism. To move the platform in azimuth the feet are unlocked and the platform moves on its wheels. Small movements are possible, so the accuracy achievable in azimuth is slightly better than that of elevation. However, for alignment of the platform large movements are generally needed to make even small changes in azimuth, as Polaris is at a very high elevation, at 78° north.

The method used to co-align the instruments is to align each one onto Polaris. As the position of the spectrograph relative to the platform cannot be adjusted, the platform needs to be aligned onto Polaris using the spectrograph, and then the other instruments are adjusted to match. Once the platform is roughly in the position to see Polaris, it needs to be placed in the centre of the spectrograph slit (by moving the platform). There is a finderscope (an 8x50mm, 5º f.o.v. Meade sighting telescope) that should be co-aligned with the spectrograph. Removing the filter mosaic increases the light throughput to the detector so shorter integration times can be used. The slit width can be opened out, as this will increase the field of view (the normal operation only has a field of view of 2'53"), which is needed in poor seeing conditions. Polaris is aligned to the centre of the region of the CCD that is illuminated.

With Polaris in the centre of the spectrograph slit, the other instruments can then be aligned. The finderscope and TLC can be simply adjusted in their mountings. The photometers need to be co-aligned onto Polaris in turn. As these have a 1° field of view, it is not sufficient just to point the photometers onto Polaris as they could still be up to a degree off co-alignment. Polaris therefore needs to be centred onto each photometer, by moving incrementally in azimuth and elevation separately until a peak in signal is recorded. If the photometer position is moved so the star passes right across its field of view, the central position can be determined from where the star moves off the edges of the photometer field of view.

Once all the instruments are co-aligned, the platform needs to be returned to the field-aligned position. The nominal field line position can be calculated and the azimuth and elevation can be matched to star charts. This calculated position is then checked by tracing the rays of coronal aurora to find the centre of the field line, and moving the platform to centre this on TLC. An example image from TLC with the spectrograph and photometer fields of view is shown in Figure 2.9.

Figure 2.9 *A TLC image from 10th December 1999 showing corona aurora, and the positions of the spectrograph and photometer fields of view, the green curve and red dot respectively.*

SIF has been collecting data in the winters from 1999 and much work has been done on these data. The Southampton group has particularly studied an event on the 26th November 2000 (e.g. Lanchester et al., 2001). Much of the data are still to be analysed however, and to identify periods of interest, the spectrograph videos (TLC) are being logged. Video data mostly exists only for campaign times, due to the need to manually change tapes. The spectrograph and photometers however can be controlled remotely so data can be collected continuously. Studies of proton aurora from SIF data have been made for example by Ivchenko et al., (2003), and of electron aurora from SIF by Ivchenko et al., (2004).

2.4 SUPPLEMENTARY INSTRUMENTS

A large range of other instrumentation is available covering the region of the FPIs in northern Scandinavia and Svalbard. These are described in the following sections. Comparisons of the FPI and SIF data with other instruments is important for putting the data in context.

2.4.1 EISCAT

EISCAT is the European Incoherent SCATter radar facility, which comprises of four incoherent radar transmitters and two extra receivers. Incoherent scatter radars transmit radio waves into the ionosphere at high frequencies and measure the signal returned from scatter due to electron density fluctuations. Coupling between electrons and the ions in the atmosphere is a result of the balance between diffusion of the electrons and the electrostatic attraction to the ions. This coupling happens over a scale of the Debye length, λ_D. On scales larger than this (typically about 10mm in the F-region), the electrons appear clustered around the ions, and so the average velocity of the cluster will be the ion velocity. The motions of these ions produce up-shifted and downshifted frequencies in the ion-acoustic waves produced from interference with the transmitted radar beam. This produces characteristic double-humped spectra. In addition, electron-acoustic waves produce weaker return signals, known as plasma lines, at much larger frequency shifts, due to their greater velocities. On smaller wavelength scales than λ_D, the electrons appear free, and the shape of the spectra is the broad single-humped shape of their thermal velocity. The Debye length is shorter at lower altitudes than the F-region, so single humped spectra are seen in these regions. The motions of the ions and the shape of the received spectra allow the calculation of parameters such as the ion and electron temperatures and velocities.

The received spectrum from the F-region typically has a double peak shape due to the Doppler shift of the ion lines. A schematic of a typical received spectrum is given in Figure 2.10, showing the ion and plasma (electron) lines. Various parameters can be obtained from various properties of the spectrum and other parameters such as the properties of the radar and the composition of the atmosphere. These are also shown in Figure 2.10. The strength of the echo is proportional to the electron concentration, N_e. The distance between the two peaks gives the Doppler broadening, and is proportional to the square root of the ratio $(T_e + T_i)/m_i$, the electron and ion temperatures and ion mass respectively. The sharpness of the peaks gives the ratio of electron to ion temperatures, T_e/T_i. In the E-region, where there is only one peak, due to the increased number of collisions between ions and neutral species, the spectral width can give the ion-neutral collisions frequency, ν_{in}. Due to the high collision rate, the temperatures are assumed to be the same, i.e. $T_e = T_i = T_n$. These ratios can be used to calculate the ion and electron temperatures individually, assuming an ion mass, which is determined by models of the composition over altitude. The ion-neutral collision frequency can then be used to estimate the neutral temperature. Other parameters such as the conductivities and electric field strength can also be calculated.

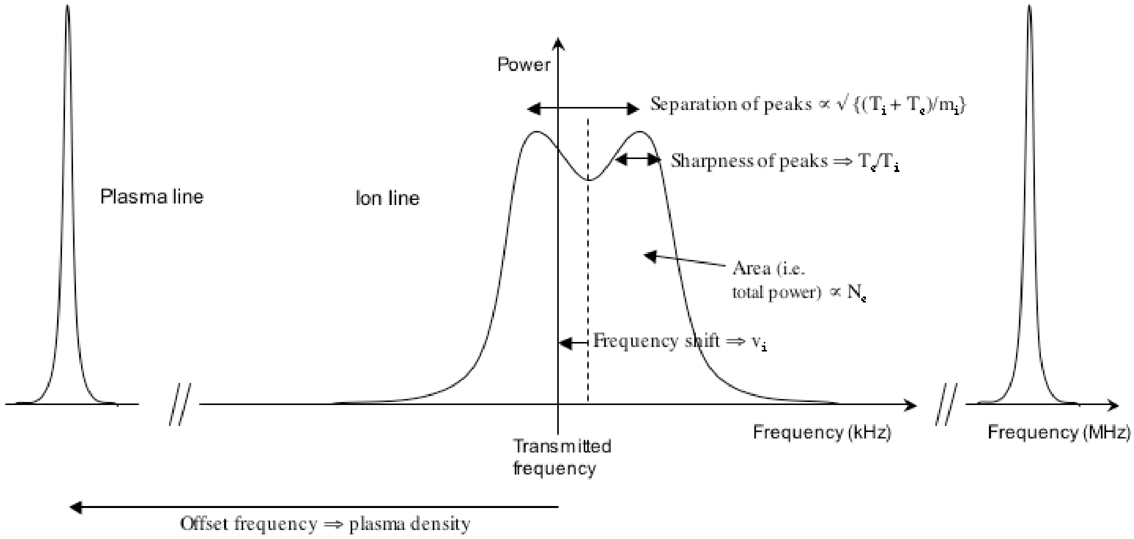

Figure 2.10 *Schematic of the spectra from a typical incoherent scattered signal from the F-region, showing the parameters obtained.*

Two transmitters are at Svalbard, at 78°09′N, 16°03′E, the EISCAT Svalbard Radars (ESR). These both transmit at 500MHz, from a fixed, field-aligned, 42m diameter parabolic dish, and from a 32m steerable dish. At Tromsø, Norway (69°35′N, 19°14′E), there is a VHF (very high frequency) transmitter at 224MHz and a UHF (ultra high frequency) transmitter at 928MHz. The VHF transmitter is made of four 30 x 40m parabolic cylinders that are steerable along the magnetic meridian. The UHF is a 32m fully steerable dish. Two additional UHF receivers are located at Sodankylä, Finland (67°22′N, 26°38′E) and Kiruna, Sweden (67°52′N, 20°26′E). Both are also fully steerable 32m dishes, which allows true tristatic measurements to be made.

2.4.2 Tomography Chain

The Sodankylä Geophysical Observatory's tomography chain measures electron densities over Scandinavia. Satellites passing over four receiver stations transmit radio signals at two frequencies, and the difference in phase between the two signals is used to calculate the total electron content (TEC). Tomographic inversion techniques are then used to create a map of the electron density along the line of the receiver chain and altitudes up to 700km.

To obtain the TECs, Russian low Earth orbit (LEO) satellites transmit coherent radio signals at 150MHz and 400MHz. The satellites orbit at 700-800km and are in a polar orbit so that they cross Scandinavia in a southwest or northeast direction. The satellites take about 20 minutes to pass over the receiver chain, and with an average of about 20 passes per day gives about one pass per hour. There are four receiver stations at the sites below:

Tromsø:	The Auroral Observatory, University of Tromsø	Norway
Kiruna:	Swedish Institute of Space Physics (IRF), Kiruna	Sweden
Luleå:	Luleå University of Technology, Luleå	Sweden
Kokkola:	Nerogas Tech Oy, Kokkola	Finland

Figure 2.11 shows typical tomography data, this is from the second night of the first tristatic campaign, the 28th February 2003, at the time of the first large peak in the FPI intensities, see section 3.2. The first plot shows the path of the satellite pass over the tomography stations. The pass is southward so it is in line with the axis of the receiver chain. It is also close to the zenith of the receivers so the data are of good quality and reliable. The northward passes tend to be at an angle from the receiving chain and so most of the pass is far from the receivers. Therefore, the results are projections onto the position of the chain, so the results in northward reconstructions are often unreliable. If the satellite is only in range of two of the receivers, the data are not analysed.

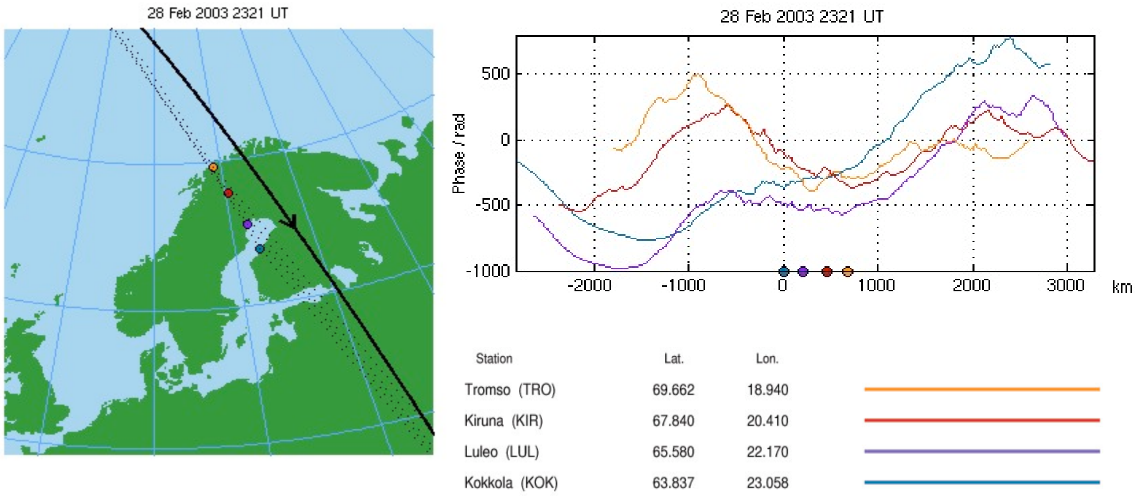

Figure 2.11 *Left: track of the satellite path (solid line), receiver stations (see key, bottom right), and minimum observing angle – 25º above the horizon (dotted lines). Right: phases of signal for the four receiver stations, and key to the stations.*

The figure on the right in Figure 2.11 above is the phase difference from the times the two signals transmitted by the satellites (150 and 400MHz) are received. The lines are for each of the receivers, with the same colour coding as the orbit pass points on the map on the left. The oscillations in the curves are indicative of gravity wave activity. These phases are used to create the reconstructions of the electron density in the plane above the receiver stations, shown in Figure 2.12 below, by means of a statistical inversion process (see Nygrén et al., 1997), using two slightly different models. The phase difference between the two signals is proportional to the total electron density along the beam between the satellite and receiver station. This phase can only be known to within a factor of 2π. This unknown multiple of 2π is treated as a random variable, for each receiver. This variable, along with the phase difference and the measurement errors are put into a matrix that is used to calculate the electron densities. However, in addition, a vertical regularisation profile is needed, which confines values in adjacent points in the grid, in the plane above the receivers. Two different profiles are used, one based on the Chapman layer profile, the other obtained from the International Reference Ionosphere (IRI) (see e.g. Bilitza et al., 1993).

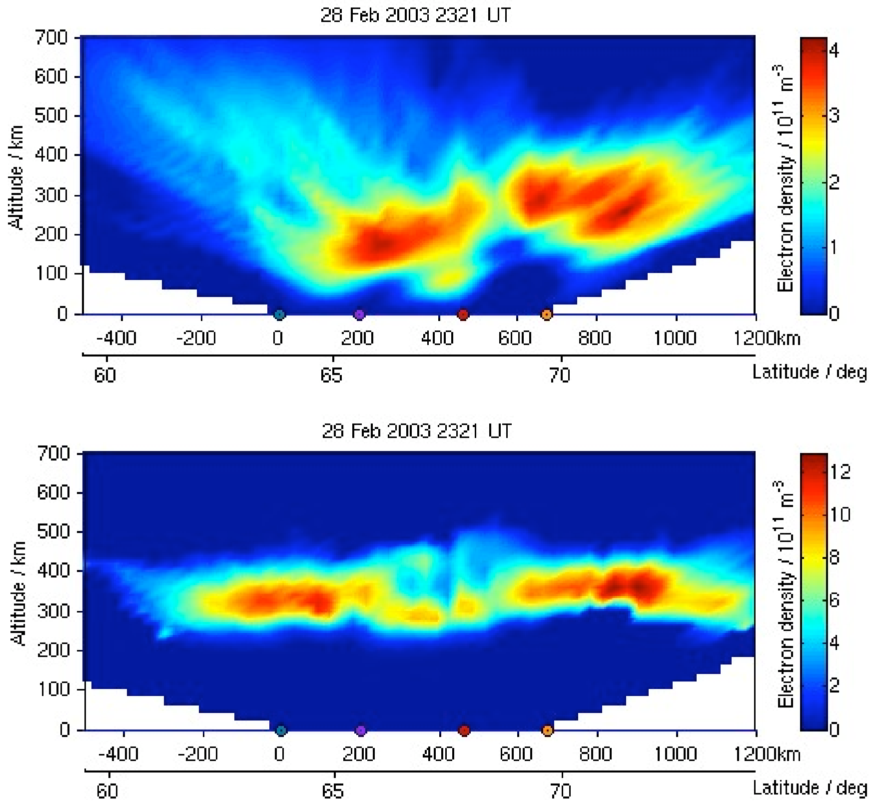

Figure 2.12 *Reconstructions for a pass on the 28th February 2003: top: Chapman reconstruction, bottom: IRI model reconstruction.*

The Chapman reconstructions have a 280km peak electron density (N_e) height and a scale height of 140km. It allows electrons to reach lower altitudes than the IRI reconstruction, as N_e is set to zero at the ground and satellite height, and so is more appropriate for active conditions when electrons are likely to penetrate further into the atmosphere. This can be seen in Figure 2.12 where the general shapes of the electron densities are similar, but the peak height is at lower altitudes in the Chapman (top) plot. The IRI reconstructions have a time dependant profile, obtained from the IRI model over the Kokkola receiver shifted down by 50km. For this model, the electron density is only forced to zero at ground level and is open at the top.

The electron densities obtained from the tomography chain, as well as those obtained from EISCAT, are used in Chapter 3. They are cross-correlated with the FPI intensities, to identify the production processes of the 6300Å emissions at F-region altitudes.

2.4.3 MIRACLE

MIRACLE – Magnetometers, Ionospheric Radars, All-sky Cameras Large Experiment – is a joint collaborative initiative run by the Finnish Meteorological Institute (FMI). It is a multi-instrument array to study mesoscale structures in auroral electrodynamics. It consists of three sets of instruments: magnetometers, all-sky cameras, and coherent scatter radars, each described in the sections below.

2.4.3.1 IMAGE

IMAGE, the International Monitor for Auroral Geomagnetic Effects (e.g. Viljanen & Häkkinen, 1997), is a chain of 29 magnetometers across Scandinavia (see Figure 2.13 further below for locations). They measure excursions from the magnetic field base line (i.e. the steady state condition) in three components. B_x is positive to the local magnetic north, B_y to the east, and B_z is positive down along the magnetic field line. As the stations cover a latitude range of 58° to 79° North, they are useful for studying the electrojets. They can be used to determine the onset time and position of a substorm, dynamics of substorms and the latitude and direction of the electrojets. Equivalent current densities, in Amps per kilometre, in the E-region can be calculated (Pulkkinen et al., 2003). Values are calculated using all the stations data, at an altitude of 100km and along a line of longitude through the centre of the magnetometer stations.

2.4.3.2 ASCs

Many of the IMAGE sites also have all sky cameras (ASCs). These are intensified video cameras with fish-eye lenses, which are pointed to the zenith so that the whole sky can be viewed. The ASCs therefore provide an overview of the conditions, such as auroral activity, light levels, and cloud cover, to provide a context for the other instruments. Some of the cameras have filters; most have a 5577Å filter to show just the green line aurora. Most stars emit at 5577Å, and cloud can be seen at this wavelength, so this is also useful for determining weather conditions, as if stars can be seen, the sky must be clear. Some sites, such as the Sodankylä ASC also have red line (6300Å) and blue (4278Å) filters and also alternate the images with white light (i.e. no filter).

Several of the sites record the data digitally. This allows keograms to be made for each night. To create these, a meridional slice is taken from the centre of an image, and these are placed together for successive images. This creates an overview of the night, with latitude plotted against time. An example is used in Figure 3.4 in section 3.2.2. There are other ASCs in the region that are not part of the MIRACLE network, but are similar in design. These include the University of Alaska's ASC in Longyearbyen, Svalbard, and APL's ASC at KEOPS, Sweden. The latter is a new colour ASC and it is adjacent to the KEOPS FPI.

2.4.3.3 STARE

The third part of MIRACLE is STARE, the Scandinavian Twin Auroral Radar Experiment – two coherent scatter radars, located in Hankasalami, Finland (62.30°N, 26.64°E) and Midtsandan, Norway (63.66°N, 10.73°E). These are described by Greenwald et al., (1978). In VHF coherent scatter radars, the radio signal is reflected from electron density fluctuations in the E-region ionosphere. Backscatter will only occur from ionospheric irregularities that are aligned perpendicular to the direction of transmission. Signal will therefore only be received for part of the time the radars are running. Incoherent scatter radars, e.g. EISCAT, do not have this restriction as they receive scatter from all electrons and ions. They are however much more expensive.

For each radar, 64 antennae transmit at 140MHz at an elevation of 6° and an azimuth coverage of 58° centred on the magnetic field line. Backscatter is produced from irregularities with a scale size of about 1 metre. 50ms pulses give a range resolution of 15 km, over a range of 495-1200km from the radar. The field of view of the Finland radar and the all-sky cameras, and the magnetometer sites, are shown in Figure 2.13.

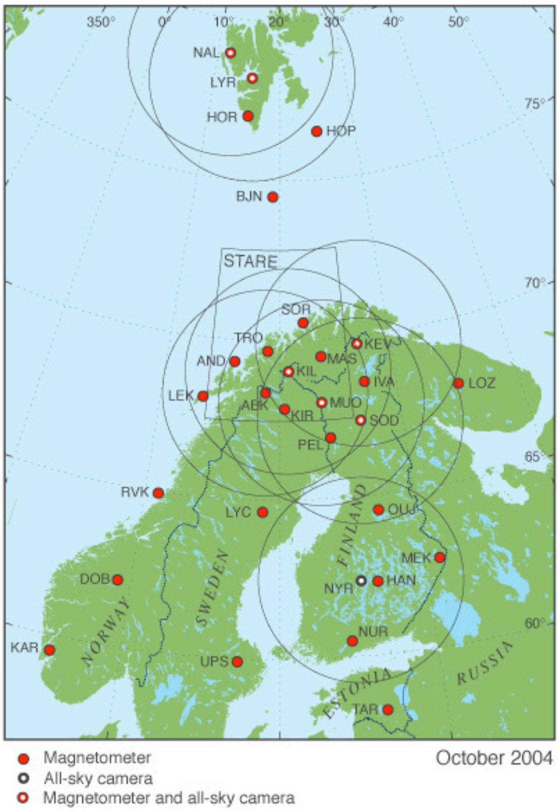

Figure 2.13 *MIRACLE instruments – magnetometer sites, all-sky camera field of views (circles) and STARE Finland radar field of view (from MIRACLE, Finnish Meteorological Institute).*

The parameters measured by the radars are the intensity of the echo (electron density), the Doppler shift (electron velocities), and the spectral width. The spectral width of the signal is dependant of the ionospheric irregularities so does not give a simple parameter. For example, large changes in the spectral width can be used to determine the location of the polar cap boundary. These are a function of range and direction with a resolution of typically 20 seconds. The two radars have an overlapping field of view so that phase differences can be obtained, which gives the electron drift velocity and so ionospheric flows. These can be used to investigate substorms and magnetospheric convection, and other ionospheric motions. STARE measures field aligned electron flows, and therefore calculates electric fields in the E-region (at altitudes of around 110km). The received echoes are co-located with electrojet currents, and Makarevitch, (2004) found that the backscatter power is a measure of electrojet intensity.

2.4.4 CUTLASS

CUTLASS, the Co-operative UK Twin Located Auroral Sounding System is part of the World-wide SuperDARN (Super Dual Auroral Radar Network) network. SuperDARN consists of fourteen radars, nine of which are in the northern hemisphere; five are in the south, at positions as seen in Figure 2.14.

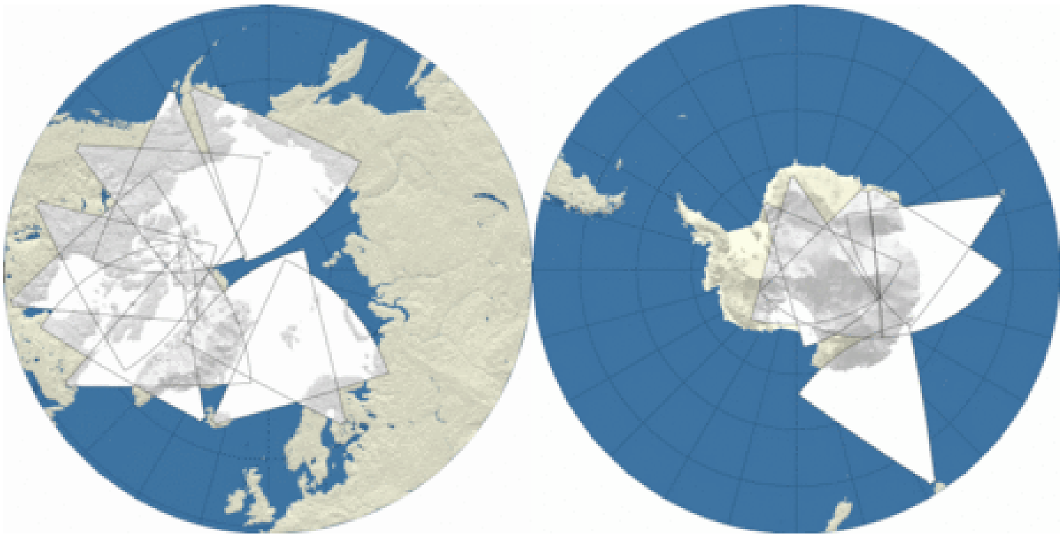

Figure 2.14 *SuperDARN radar field of views of the 9 northern (left) and 6 southern (right) radars (image is courtesy of the Johns Hopkins University Applied Physics Laboratory).*

The two CUTLASS radars, are located in Finland (Hankasalami) and Iceland (Pykkvibaer) and are run by the University of Leicester, with support from Finland and Sweden. They, like STARE, are coherent scatter radars, which means they measure field-aligned irregularities from ionosphere. They are high-frequency (HF) radars so they can detect F region as well as E region echoes. CUTLASS transmits between 8MHz and 20MHz. Sixteen transmitting antennae 3.24° apart give a field of view of 52°, centred on the magnetic meridian at 12° to the west. Pulsed transmitting at 100μs gives a range resolution of up to 15km, over a range of 180 – 3500km.

The parameters that are measured are the backscatter power, the spectral width, the line of sight plasma velocity, and the elevation angle of returned signal. The line of sight plasma velocity gives a 2D velocity vector with the two radars. The elevation angle of the signal is determined by the height of the scattering region. However, there are complications due to bending of the rays as they pass through regions of different refractive index, hence it is truly the virtual height that is measured, and models and assumptions are needed to translate this to the true height.

2.4.5 IRIS

IRIS is the Imaging Riometer for Ionospheric Studies. It is located at Kilpisjärvi, Finland (69.05° N, 20.79° E) and is run by the University of Lancaster and Sodankylä Geophysical Observatory (SGO), Finland. A riometer is a Relative Ionospheric Opacity METER and it measures the reduction in the intensity of cosmic radio noise. IRIS operates at 38.2 MHz, which is frequency that is absorbed in the D-region of the ionosphere. IRIS is an imaging riometer as it uses 64 antenna to measure the absorption, and has beam widths of 13° to 16°.

2.4.6 Satellites

There are several satellites with instruments aboard that can be useful either as they provide data at higher altitudes or as they provide in-situ measurements, rather than remote sensing. The origin of geomagnetic activity is viewed with the Solar Heliospheric Observatory (SOHO), which observes the Sun at different wavelengths and over different fields of view. This is useful both for studying solar flares and coronal mass ejections (CMEs), and as an early indicator for geomagnetic storms that may reach the Earth. SOHO and the Advanced Composition Explorer (ACE) orbit the L1 first Lagrangian point, 1/100th of the distance from the Earth to the Sun. ACE carries six high-resolution spectrometers to provide data on the solar wind before it reaches the Earth, as well as information on the inter-planetary medium and galactic cosmic rays.

In-situ measurements of the magnetosphere and the Earth's upper atmosphere are made principally by two satellites. Dynamics Explorer 2 (DE2) was a satellite in a polar orbit of altitudes from 300-1000km. It measured electric and magnetic fields, and neutral and plasma parameters with the Wind and Temperature Spectrometer (WATS), a Fabry-Perot interferometer, an ion drift meter and other instruments. The Upper Atmosphere Research Satellite (UARS) measures neutral winds with the Wind Imaging Interferometer (WINDII) and the High Resolution Doppler Imager (HRDI). UARS also measures chemical compositions, electrons, protons, Solar UV and X-rays. TIMED, the Thermosphere Ionosphere Mesosphere Energetics and Dynamics satellite, measures the MLT region from 60-180km altitude. Instruments onboard include a Doppler interferometer (TIDI), and UV and EUV imagers.

Remote sensing of the aurora from space is achieved by instruments on the Polar and IMAGE spacecraft. Their advantage over ground based instrumentation is that they view a large portion of the auroral oval at a time, and are therefore useful in viewing large scale dynamics of the aurora. The Polar spacecraft is in a highly elliptical orbit to measure the polar caps and auroral ovals. It measures, among other things, the aurora in visible (VIS instrument), ultraviolet (UVI instrument) and X-ray (PIXIE) wavelengths. The IMAGE satellite (Imager of Magnetopause-to-Aurora Global Exploration) is also in a elliptical polar orbit. It has neutral atom imagers for various energy ranges, an extreme and a far ultraviolet imagers (EUV and FUV respectively) and a plasma imager in the radio frequencies.

GOES is a geostationary weather satellite that provides visible and infrared images of the Earth's surface, though mostly over the American continent. It can be useful for providing wide scale information on cloud cover levels. The European geostationary satellite, METEOSAT, provides more global weather information, via stations including the Dundee Satellite Receiving Station (DundeeSat). This also provides high-resolution surface images in the visible, and in the infrared, so cloud levels at night time can be determined.

3 ELECTRON DENSITY CORRELATIONS WITH FPI INTENSITIES

3.1 INTRODUCTION

This chapter covers results on correlations between FPI intensities and electron density measurements from satellite tomography data and EISCAT radars. Reconstructions from the tomography chain run by Sodankylä Geophysical Observatory provide electron densities for the same region that the FPIs view. A description of SGO's tomography chain is given in 2.4.2. If dissociative recombination is the dominant mechanism at night time for production of the 6300Å emission that the FPIs measure, then the electron densities should be proportional to the FPI intensities. This can be evaluated with a cross-correlation analysis. Alternative sources of electron density data are the EISCAT radars. These data will also be correlated with FPI red line intensities.

The production methods of the 6300Å emission are given in section 1.2.4. Dissociative recombination of O_2^+ ions is the main source of F2 region 6300Å emissions. Other factors that are important at this height are processes (ii) to (iv) from section 1.2.4 – electron impact by energetic auroral electrons, excitation by electrons in the high-energy tail of the ambient thermal population and dissociation of O_2^+. The total contribution from all these will be proportional to the 6300Å intensity (I_{6300}), but (ii) to (iv) are minor contributors.

The cross correlation $\rho_{x,y}$, between two data sets *x* and *y*, is calculated using equation 30 below, which is used by Excel's 'correl' function.

$$\rho_{x,y} = \frac{1}{n}\sum_{i=1}^{n}\frac{(x_i - \mu_x)(y_i - \mu_y)}{\sigma_x \sigma_y} \qquad [30]$$

for data with mean value μ and variance σ, and where $-1 \le \rho_{x,y} \geq 1$.

The significance of the cross correlation coefficient obtained for data sets can be obtained using a non-directional hypothesis (Lowry, 1999), i.e. where no positive or negative direction, i.e. correlation or anti-correlation, are assumed beforehand. The significance of an obtained value of the correlation coefficient, *r*, is dependant on the number of degrees of freedom of the sample, i.e. the sample size, *N*. For the 5% significance (95% confidence) level, assuming a non-directional hypothesis, a correlation coefficient of 0.8 is needed for a sample size (*N*) of 5, 0.44 for *N* = 20, and levelling out to 0.35 for N > 35. Most of the data sets used in this study have a sample much greater than 35 data points.

The locations of the FPIs with respect to the EISCAT radars used in the tristatic campaigns is shown in section 3.2, including some example data from the first tristatic campaign. The correlation of the 6300Å intensities from FPI data are compared with electron densities, obtained from tomographic inversions of satellite data from the Sodankylä Geophysical Observatory Tomography Chain (section 3.3), and from EISCAT radar data in section 3.4.

3.2 FPI-EISCAT TRISTATIC CAMPAIGNS

3.2.1 INTRODUCTION

The purpose of this section is to describe the experimental setup of the FPIs used in this chapter in the correlations in sections 3.3 and 3.4 and in Chapter 4 in the gravity wave case study in section 4.4. The procedures and conditions used are described, along with some of the typical properties of the data obtained. A detailed analysis of these data is given in Aruliah et al., (2004) and Aruliah et al., (2005).

The presence of co-located Fabry-Perot Interferometers (FPIs) and EISCAT radars in three sites in northern Scandinavia has provided a unique opportunity to perform tristatic measurements of the thermosphere and ionosphere. The FPIs measure the line of sight height integrated 6300Å red emission line, and tristatic measurements will allow the location of the emitting volume to be identified. Previously, assumptions had to be made about the thermosphere such as uniform horizontal winds and zero vertical winds, but the tristatic measurements provide true vector measurements so that these assumptions will not have to be made. Independent calibration of each of the FPIs means that they are independent measurements and if they show the same values of intensity and temperature, then either they will be looking at the same volume, or the thermosphere is uniform, which is an incorrect assumption. These measurements mean that the thermospheric parameters do not have to be derived from ionospheric measurements, as is currently done from, for example, EISCAT data and the MSIS model.

This experiment aims to allow studies of why models predict much lower temperatures for the neutral gas while overestimating momentum transfer between ions and neutral particles. As well as providing true vectors at the tristatic position (point A in Figure 3.1), this setup provides a large grid of data points between which variations can be found. Both of these features can be used to find small-scale variations both at the tristatic point over time and spatially across the region covered.

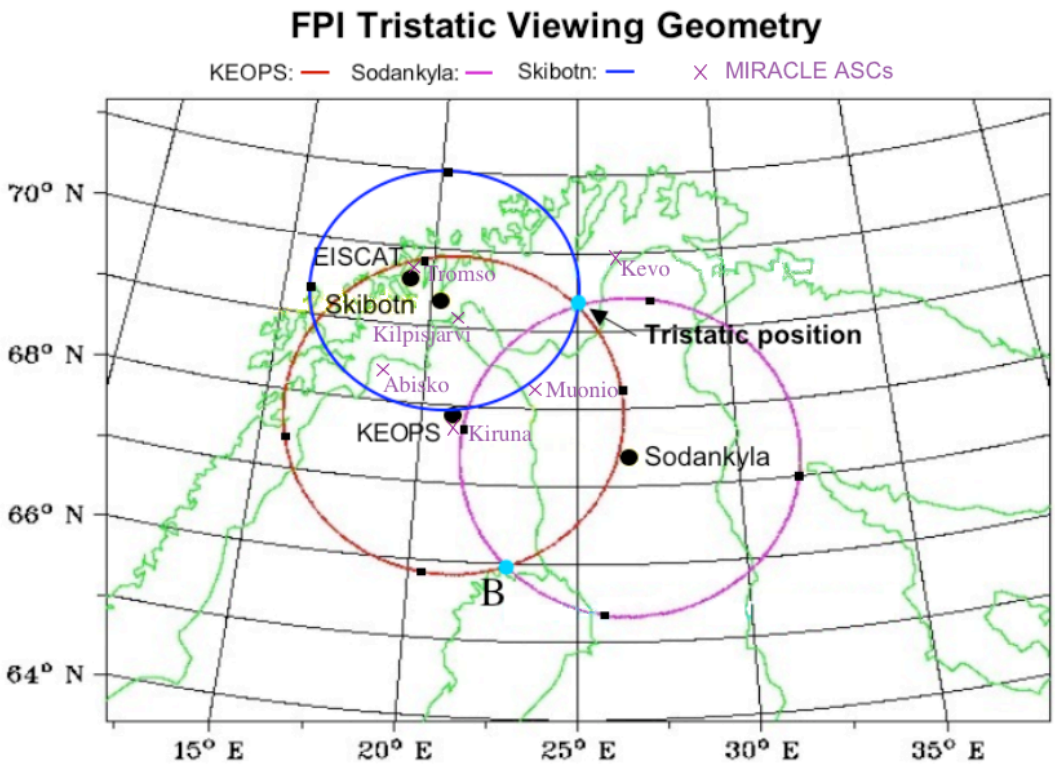

Figure 3.1 *Circles representing fields of view of the three FPIs, at Sodankylä; KEOPS, and Skibotn. Points show the measurements taken, in the six directions: north N, east E, south S, west W, and the bistatic A and B positions. The radii of the circles are due to observing the 6300Å oxygen emission line at 240km altitude. Crosses show positions of some of the MIRACLE all sky cameras.*

Figure 3.1 shows the fields of view of the three FPIs in northern Scandinavia as set up for the tristatic campaigns highlighting the separate look-directions that are viewed. It can be seen that a large grid of observations is built up from taking observations in many look directions at each of the sites. The common volume that is observed by the FPIs and the co-located EISCAT radars can also be seen in Figure 3.2 below.

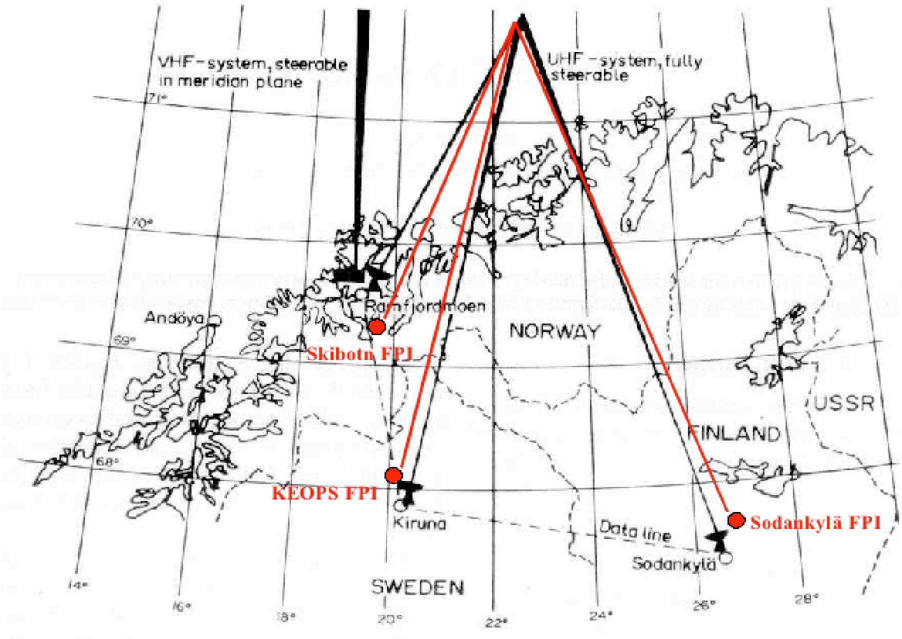

Figure 3.2 *The clear advantage of tristatic measurements is that they observe a common volume. Black lines show the line of sight of the EISCAT radars and red for the FPIs when viewing the tristatic A position.*

EISCAT radars are located near each of the FPI sites; the transmitter is near the Skibotn FPI, at Tromsø in Norway. On the nights of the 27th and 28th February 2003, the EISCAT radars were first set up to look at the same common volume that the FPIs were measuring, to provide simultaneous independent neutral and ionospheric measurements. The second campaign was on the nights of the 25th and 27th November 2003, though the 27th was overcast for the whole night. The 25th data are used for a case study of gravity waves in Chapter 4. The third tristatic campaign was on the 14th and 16th February 2004, but both these nights were cloudy. The fourth campaign was on the 8th and 9th February 2005. For this campaign, the EISCAT beam from Tromsø was scanned along the KEOPS tristatic A line of sight, so that the height of the 6300Å emission volume could be studied through correlations of FPI data with EISCAT electron densities.

3.2.2 TEMPERATURES AND INTENSITIES

To calculate vectors for the measured parameters, the three FPIs need to be looking at a common volume. The F_2 layer peak height can vary significantly through a night, but as the 6300Å line has a broad emission peak this is not usually of great concern (e.g. Solomon et al., 1988) but this can now be investigated. Confidence that the FPIs are viewing the same volume comes from plots of the intensities, in Figure 3.3.

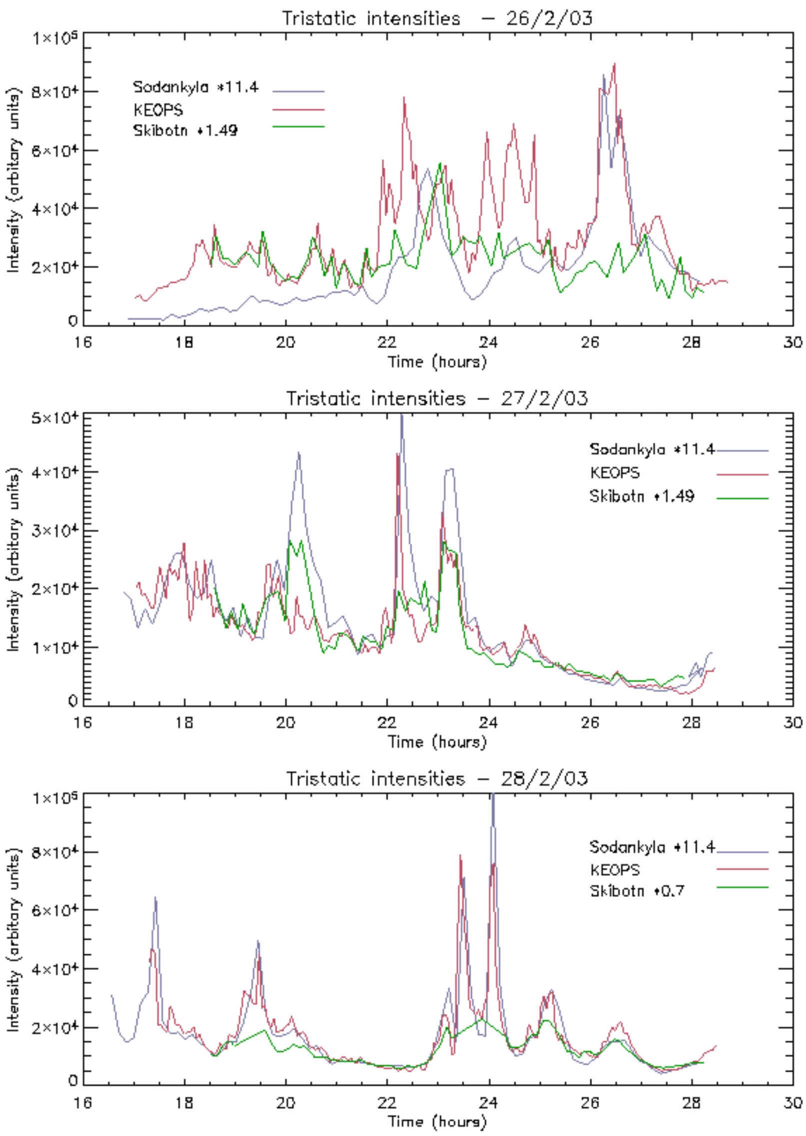

Figure 3.3 *Tristatic intensities for two tristatic campaign nights, the 27th and 28th February 2003, and the night before, the 26th, for comparison. Intensities are line of sight measurements from each of the three sites.*

The Sodankylä and Skibotn values were adjusted by a factor for these plots to get the three sites on the same scale, as the sensitivities of the detectors are not equal and intensities are not yet absolutely calibrated. It can be seen from these intensities that the 26th was cloudy, as the data from two sites do not correspond well. The 27th and 28th show much better correlation, as the nights were mostly clear, so that there is a high degree of confidence that the two FPIs are looking at the same volume. The Skibotn data (green) does not appear to match the peaks in the others as well owing to the low time resolution of this data, as despite reasonable integration times (30s), the time taken to move between look directions is several minutes due to the more complicated mechanics of the mirror mechanism. The peaks, for example at 22UT on the 27th, are from auroral arcs passing the tristatic point, as can be seen in the all sky camera keogram in Figure 3.4 below.

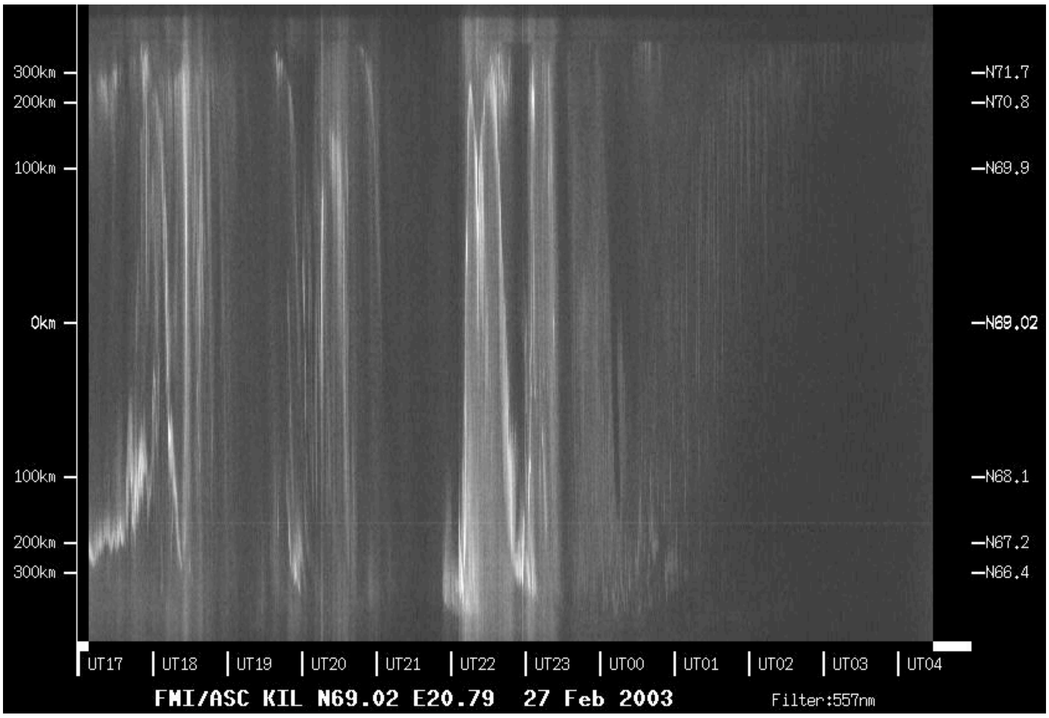

Figure 3.4 *All Sky Camera Keogram from Kilpisjärvi, which is between the tristatic A point and Skibotn, for the period of the first tristatic campaign night, the 27th February 2003. From the MIRACLE network: http://www.geo.fmi.fi/MIRACLE/ASC/*

To create the keograms, vertical slices through the centre of the ASC images throughout the night are placed side by side, which gives an overall picture of the night. The strong arcs at about 22 and 23UT on the 27th February, as well as the slightly weaker arcs from 17-19UT, can be seen in both the keogram and the intensity plots of Figure 3.3. The keogram from Sodankylä also shows this, but keograms from the other sites are not available. Individual ASC images, such as Figure 3.5 below, also show the activity, and are in addition a useful indicator of factors such as cloud cover levels and ice or snow on the dome.

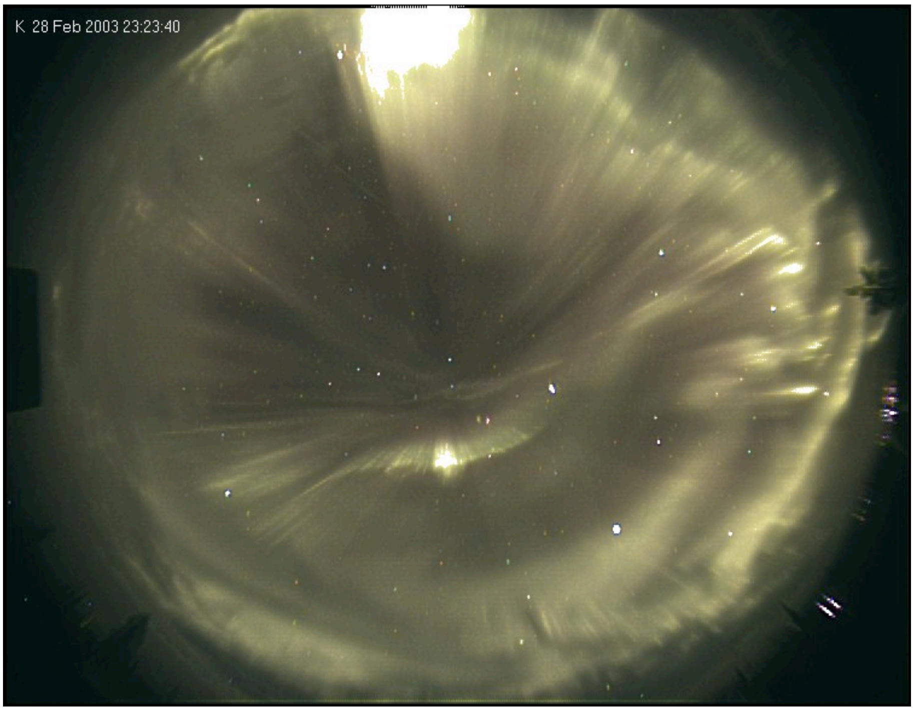

Figure 3.5 *This All Sky Camera image is from KEOPS at 2323UT on the 28th February 2003.*

This figure shows the activity at 23:30UT, the approximate time of the peak in the intensities in Figure 3.3 on the 28th (bottom plot). Visible stars confirm that the sky over Kiruna was clear.

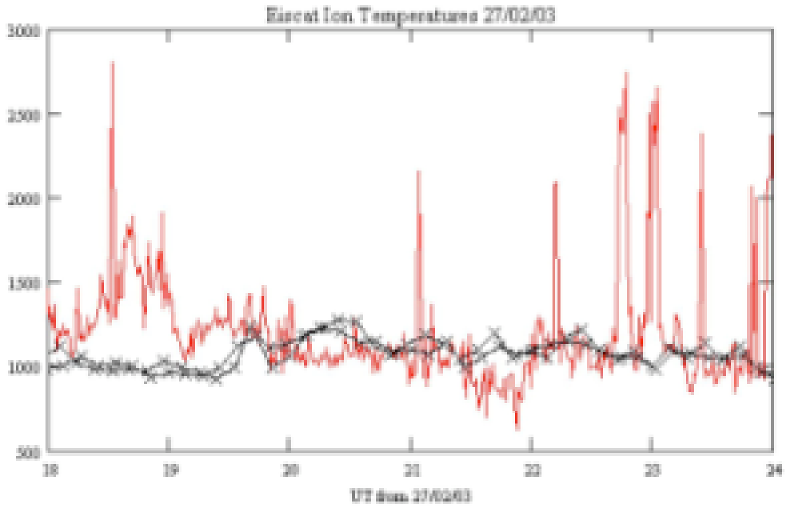

Figure 3.6 *Neutral temperatures from the Sodankylä north and tristatic A positions (two black lines with crosses) and EISCAT ion temperatures (red line) for the 27th February 2003.*

The neutral temperatures for one of the tristatic campaign nights are shown in Figure 3.6 along with the EISCAT ion temperatures (Aruliah et al., 2005). The ion temperatures vary more than T_n but the similarities can be seen between the two black FPI T_n lines that are from adjacent viewing locations. Differences between these and the other directions give us more information about the spatial structure of the atmosphere at these times. The period between 21:30UT and 22:00UT shows the ion temperature to be less that the neutral temperature, which is not realistic, as heat would be immediately transferred to the ions from the much denser thermosphere. The discrepancy during this period can be explained by the period of substorm expansion at this time, as can be seen in the IMAGE magnetometer data in Figure 3.7.

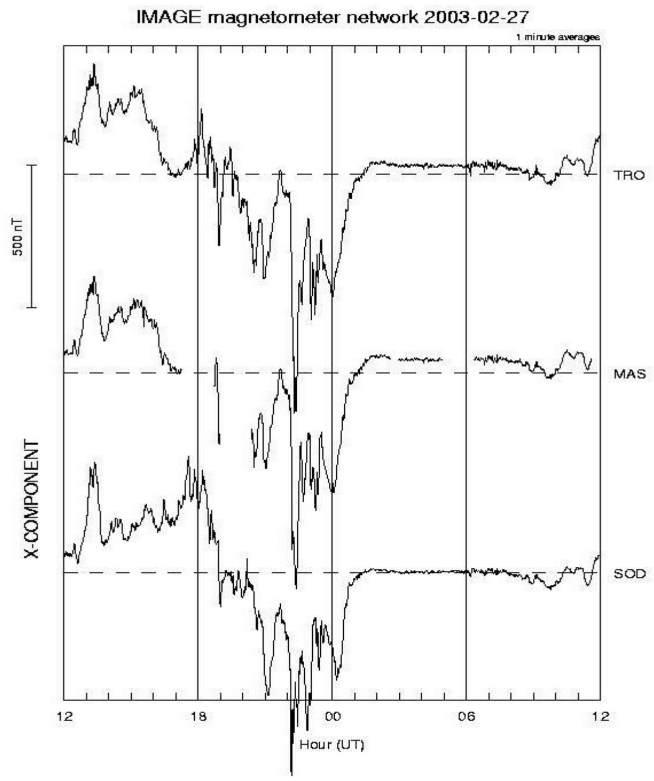

Figure 3.7 *X-component magnetic field data from the IMAGE magnetometer network, which runs from Svalbard down through Scandinavia. The east-west components at Tromsø, Masi (nearest magnetometer to the tristatic point), and Sodankylä are shown.*

The strong negative component in the magnetometers from 21:30UT to 23:00UT is due to a westward electrojet during a substorm expansion. This leads to an upwelling, which can be seen in the vertical winds, which are shown in the next section. This has the effect of increasing the molecular density at this height, and so the assumption that the ion composition is completely atomic oxygen is incorrect. The change in composition increases the recombination rate and so decreases the electron density. The failure of the assumption of the composition that is used in calculating the ion temperatures, explains the discrepancy shown in Figure 3.6. This is also true of the period between 20:00UT and 20:30UT. This is also discussed further by Aruliah et al., (2005). This, as well as the all sky camera data, is an example of how other independent instruments can provide a larger scale context from which to interpret the FPI data.

3.2.3 NEUTRAL WIND VELOCITIES

One of the other main advantages of the tristatic campaign, other than viewing a common volume, is the large grid of observation points. Each FPI makes observations at each of the cardinal positions as well as the tristatic, bistatic and zenith so that a grid covering over 600km is available. This will allow events to be tracked as they move across the grid. It also provides more points that can be compared with the tomography data, in section 3.3 below, thus creating a better statistical sample than if only one data point was available. Comparisons of the different look directions should show similarities or differences depending on the positions. Some of these can be seen in the figures of the horizontal winds below.

The meridional wind observations in Figure 3.8 are from the three FPIs but separated according to three different latitude bands. The bottom includes the bistatic B and the south positions from KEOPS and Sodankylä, the central band includes the tristatic A and north look directions from KEOPS and Sodankylä, and the top is the north Skibotn data. There are similarities in the meridional measurements within each band, but differences between the bands. As the features can be seen to be different at each of the latitudes, it shows spatial structure on the scale of ~100km, which is an order of magnitude smaller than is normally assumed for neutral winds. The width of the central latitude band, i.e. the maximum separation in latitude of the meridional wind sites, is approximately 60km, though it has a longitudinal variation of a few hundred kilometres. Gradients of winds of up to 5 ms^{-1}/km are seen between directions in this band. However, the overall wind field does not change much between the latitudinal bands, where the large-scale antisunward flow dominates. This therefore shows that within the large-scale structures, there are smaller variations in the meridional winds on scales of a few tens of kilometres, and less than 60km.

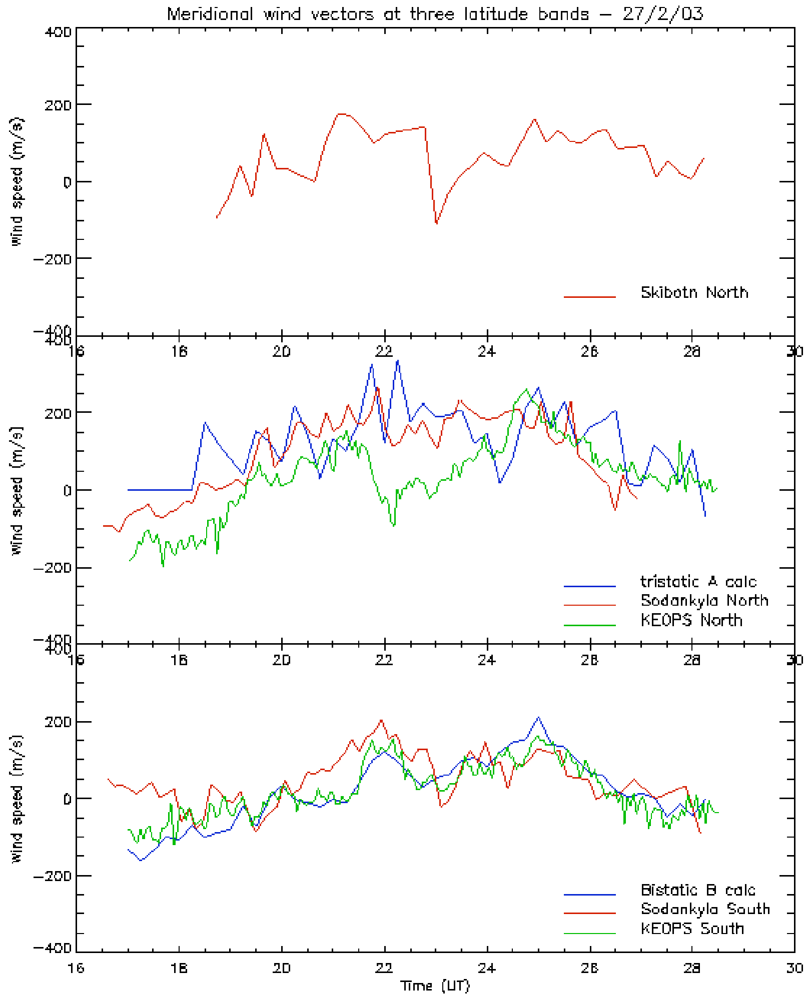

Figure 3.8 *This shows the meridional winds from three different latitude bands for the 27th February 2003 (as may be seen on Figure 3.1 if the viewing volumes are considered in horizontal bands).*

The zonal winds in Figure 3.9 show the same overall variation with time, but with a lag of the west winds of approximately one hour after the east. This lag is due to the difference in longitude of the look directions, and hence of local time. The overall convection pattern of the winds rotates around in a westerly direction. KEOPS and Sodankylä are at similar latitudes and show a similar large-scale variation with time, but Skibotn is further north and, especially in the east winds, shows a different variation over time. The Skibotn east winds have a higher time resolution than the other Skibotn look directions as this is the tristatic direction and so it has been sampled twice as often.

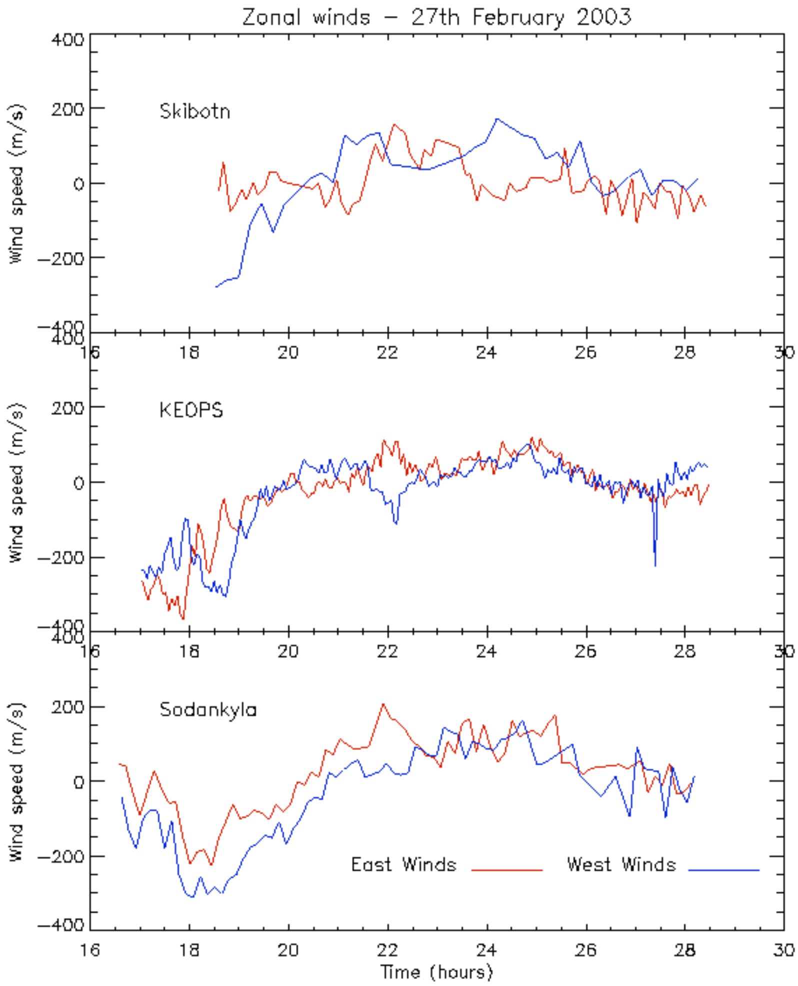

Figure 3.9 *Zonal winds for 27th February 2003 for the FPIs at Skibotn (top), Sodankylä (middle), and KEOPS (bottom).*

It can be seen in Figure 3.10 that the vertical winds are an order of magnitude smaller than the horizontal winds of Figure 3.8 and Figure 3.9, which are shown on a different scale. Although the variations in the vertical wind speeds are not as large in magnitude as the horizontal winds, they are large compared with the scale of the data, with variations of over $50ms^{-1}$, on a mean wind with a standard deviation of $\sim 20ms^{-1}$.

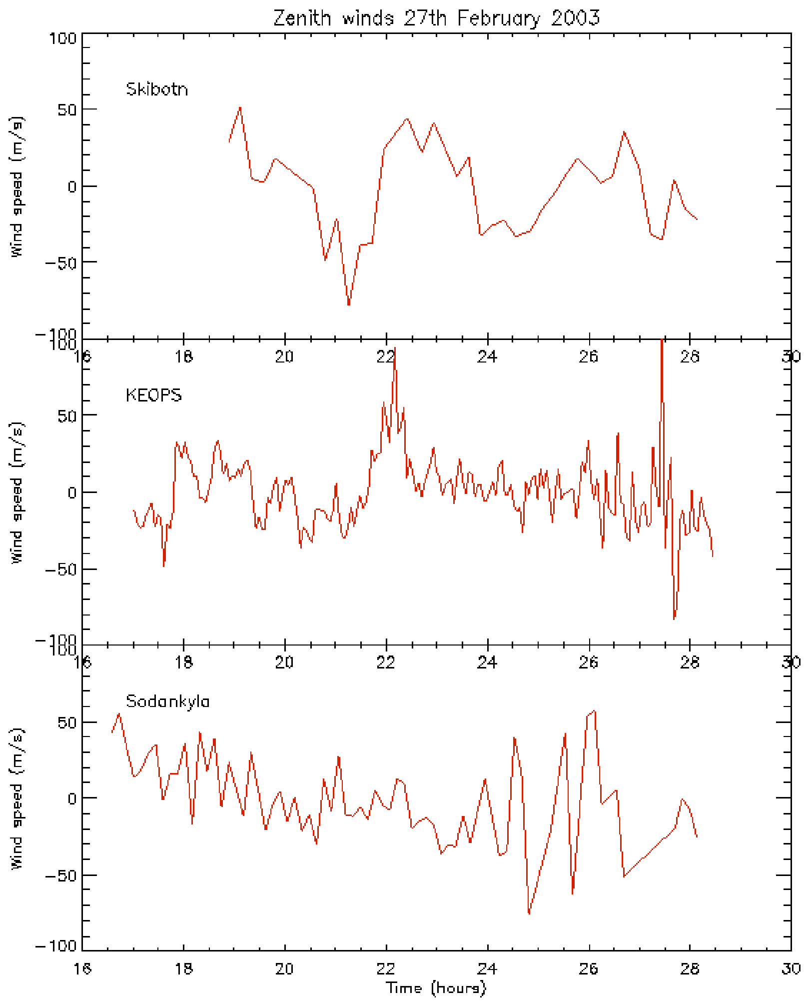

Figure 3.10 *Zenith winds measured vertically above each of the three FPIs for the 27th February 2003. The vertical winds vary with much smaller amplitude than the horizontal winds).*

KEOPS data has the highest time resolution and consequently shows more temporal structure than the other sites. High time resolution data from KEOPS are discussed further in Chapter 4. The up-welling at 22UT in the KEOPS is an example of the spatial structure between the sites, as it is seen about half an hour later at Skibotn, which is approximately 250km further north than KEOPS. These wind speeds have been used by Aruliah et al., (2005) to calculate that the Joule heating from the neutral wind dynamo contributes approximately 40% to the total Joule heating, which shows the importance of these small-scale variations.

3.3 *TOMOGRAPHY DATA CORRELATIONS*

3.3.1 DATABASE

Due to the many look directions of the FPIs producing a good grid of data, there are several points where the FPI data overlap the tomography chain as shown in Figure 3.11.

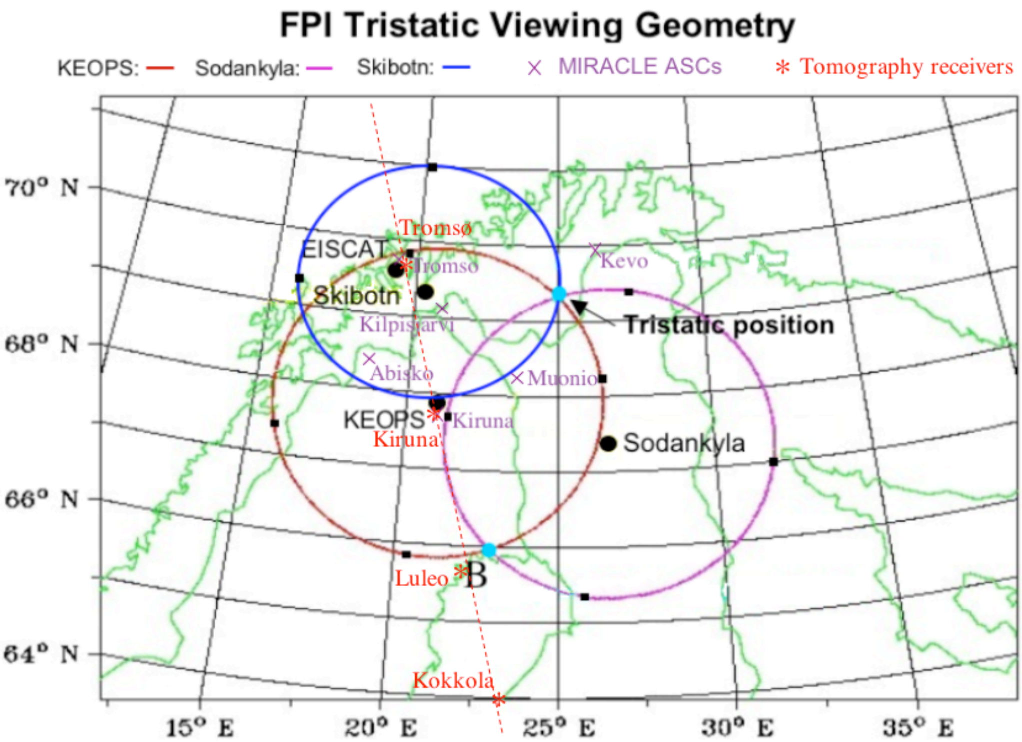

Figure 3.11 *Fields of view of the three FPIs at KEOPS, Sodankylä, and Skibotn, and SGO's Tomography receiver sites (red asterisks). The red line indicates the path of the reconstruction models. Crosses are MIRACLE ASCs.*

As the tomography chain satellite passes are slow and infrequent compared with the FPI data, many passes from different nights have been compared. SGO provided data above Tromsø and Kiruna for all the passes of sufficient quality data on clear nights where FPI data exists in 2003. These dates include five passes for the first tristatic campaign (27th – 28th February 2003), at 22:51UT, 00:37UT and 01:51UT on the night of the 27th February and at 18:57 and 23:21 (the pass in Figure 2.11 and Figure 2.12 in section 2.4.2) on the 28th February. Other clear nights with passes were on the 24th and 27th March, 21st October, 13th and 17th November 2003, and three passes for the second tristatic campaign on the 25th November 2003. This totals seventeen passes that can be used, where the nights are clear and FPI data exist, providing a reasonable statistical sample.

For each satellite pass, the electron densities are calculated at various points along the path at 20km intervals from 20km to 700km altitude. N_e values are calculated at approximately 120 points on each path. The latitude and longitudes of the look directions within the tomography chain field of view, presuming an emission height of 240km, are given in Table 3.1 below.

Site and Look direction	***Look direction latitude***	***Look direction longitude***	***Nearest tomography latitude***	***Nearest tomography longitude***	***Distance between data points***
KEOPS north	69.8ºN	20.0ºE	69.81ºN	18.80ºE	133km
KEOPS zenith	67.9ºN	21.0ºE	67.74ºN	20.53ºE	55km
KEOPS south	65.6ºN	22.0ºE	65.92ºN	21.81ºE	41km
Sodankylä west	67.9ºN	21.0ºE	67.74ºN	20.53ºE	55km
Skibotn south	67.9ºN	21.0ºE	67.74ºN	20.53ºE	55km
Skibotn zenith	69.4ºN	20.5ºE	69.29ºN	19.27ºE	137km

Table 3.1 *Positions of FPI look directions and the nearest point on the tomography chain reconstruction plane.*

The points along the satellite pass that are nearest to the FPI look direction locations are used to compare with the FPI data. These points are shown in the fourth and fifth columns of Table 3.1. Skibotn data are only available on the tristatic nights, so for the remaining passes, only the top four sites listed above are used. In addition, for a pass on the 27th March 2003 only KEOPS data are available, and for a pass on the 17th November 2003 only Sodankylä data exist. The last column gives the distance between the FPI look direction and the position of the nearest point on the reconstruction model. The field of view of the FPIs is one degree, which gives a viewing area approximately 10km wide. As the intensities determined from the FPIs are not presently calibrated, to compare the data from different sites, the data must be normalised. The data are normalised to KEOPS data by multiplying the Sodankylä data by a factor of 11.40 and the Skibotn data by a factor of 1.49.

3.3.2 Results and Discussion

The tomography data from Sodankylä Geophysical Observatory, and their method for calculating electron densities, was described in section 2.4.2. The tomography satellite passes are much longer than the FPI resolution (20 minutes for a satellite pass compared with 40 – 60 seconds for a FPI integration time). Therefore, there are various choices of which FPI data to compare to the tomography data. The FPI data point from the time of the start of the crossing can be used, or the point at the middle of the crossing time can be used. For a twenty-minute satellite path time, this would be 10 minutes after the start time. The start point of the satellite pass is where the satellite is at the edge of its range, so the mid point of the pass will be nearer to the FPI look direction. Alternatively, the average of the FPI intensities across the whole crossing time could be used.

Electron densities are determined from the satellite data by two methods: the Chapman and the IRI reconstruction models, which are described in section 2.4.2. Results for the correlations of N_e values obtained from each method, with FPI data from the pass midpoint, and with the average over the pass time, are shown in Table 3.2 below. Results are shown for the satellite passes on the first tristatic campaign (five passes on the 27th and 28th February 2003); three passes on the second tristatic campaign night (the 25th November 2003); and nine passes through the rest of 2003. The three sites are shown separately, in addition to their total (in bold).

Dates	*Site*	*Number of data points used*	*95% confidence correlation value*	*Chapman Ne c.f. pass mid-point intensities*	*Chapman Ne c.f. pass average intensities*	*IRI N_e c.f. pass mid-point intensities*	*IRI N_e c.f. pass average intensities*
1st tristatic campaign	KEOPS	15	0.51	0.45	0.52	0.90	0.90
	Sodankylä	5	0.88	0.40	0.00	0.91	-0.04
	Skibotn	8	0.71	-0.20	-0.53	0.15	-0.32
	All	**28**	0.37	0.28	0.22	**0.70**	**0.57**
2nd tristatic campaign	KEOPS	9	0.61	0.80	0.78	0.72	0.70
	Sodankylä	3	0.88	-0.64	-0.75	0.36	0.50
	Skibotn	6	0.81	0.65	0.69	0.74	0.77
	All	**18**	0.47	**0.68**	**0.66**	**0.71**	**0.63**
Non - campaign passes	KEOPS	24	0.40	0.46	0.47	0.57	0.56
	Sodankylä	8	0.71	0.56	0.60	0.79	0.83
	Skibotn	0	N/A	N/A	N/A	N/A	N/A
	All	**32**	0.35	**0.43**	**0.46**	**0.64**	**0.62**
All 2003 passes	KEOPS	48	0.35	0.59	0.62	0.79	0.78
	Sodankylä	16	0.50	0.43	0.02	0.84	0.53
	Skibotn	14	0.53	0.09	-0.26	0.38	-0.08
	All	**78**	0.35	**0.48**	**0.42**	**0.70**	**0.59**

Table 3.2 *Summary of results of correlations of Chapman and IRI reconstruction model N_e values with FPI intensities from each of the sites over various periods. Greyed out cells have correlations less that 95% confidence.*

The third column of Table 3.2 shows the number of look directions that have been used for that period. The fourth column gives the correlation value that is required for a 95% confidence (or 5% significance) level for that number of data points (see section 3.1). Cells that are greyed out have correlation coefficients less than this 95% confidence level. A number of points can be seen from this table and these are described below. Figure 3.12 is a scatter plot of FPI intensities at the times of the satellite mid-point and the IRI reconstruction electron densities. A best-fit line is also shown on the plot as a guide. The correlation coefficient for these data is 0.70.

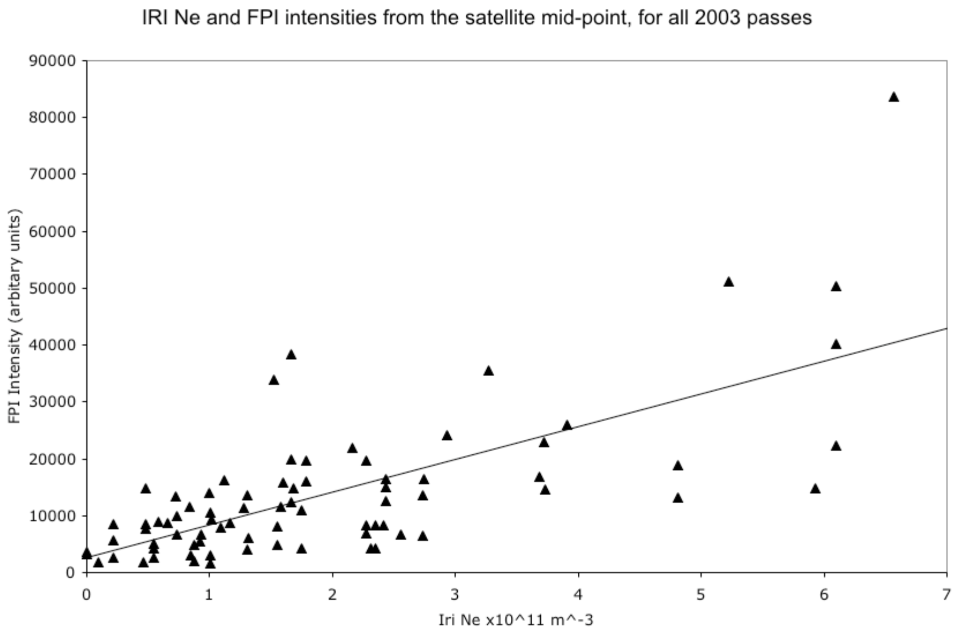

Figure 3.12 *Data from all the usable satellite passes. IRI reconstruction N_e data are compared with FPI intensities from half way through the satellite pass time.*

The Chapman and IRI reconstructions have slightly different properties, so they would therefore be expected to give different results when correlated with the FPI data. The Chapman profiles were originally thought to be the best as the scheme allows electrons to reach lower altitudes, as N_e is set to zero at the ground and satellite height, and so it is more appropriate for active conditions. However, it was found that the IRI reconstruction method of producing tomography electron densities produced higher correlation coefficients than the Chapman reconstructions. For example, the correlation value between the Chapman model N_e values and the FPI intensities for the first tristatic campaign (the 27th and 28th February 2003) was 0.28 (compared with FPI data from half way through the satellite pass), but 0.70 when correlated with the IRI results. For the three passes on the second tristatic campaign (25th November 2003) the Chapman correlation was 0.68 and the IRI gave 0.71. The only exception to this is for KEOPS data on the second tristatic campaign, where the

Chapman N_e gives a correlation of 0.80, though the IRI value is not much smaller than this, at 0.72. These are both also well above the 95% confidence level of 0.61. In addition, the Chapman model produces more coefficients less than the 95% confidence level. The IRI reconstruction model has a time dependent profile, and it has a more realistic peak electron density height, at 50km below that of the Chapman model at 280km.

Both models were correlated with the FPI data from the centre of the satellite pass and from the total period of the pass. For the Chapman reconstruction, equal numbers of coefficients from each of these, from Table 3.2, give a higher coefficient. However, many of these values are less than the 95% confidence level. The highest values of correlation coefficient from the IRI reconstruction model were obtained when the FPI data point nearest to the centre of the 20-minute pass was used, rather than when an average over the pass time was used. This is true for all but one of the periods over the 95% confidence level. The Sodankylä data from the passes not on tristatic campaign nights correlates with the pass mid-point with a coefficient of 0.79, but with a slightly higher value of 0.83 using the average FPI data over the pass. Only eight data points were available for this period, which is not a large statistical sample. The fact that the single FPI data point, rather than a 20-minute average, is better correlated with the electron densities for this set of the data, could be an indication that small-scale structure is important in the thermosphere. The satellite mid-point time was chosen as this is nearest the position of the FPI sites and so will be approximately coincident with the FPI timings. The thermosphere's behaviour is not as responsive to forcing as the ionosphere, but these results imply that further integrating the thermosphere will smooth it out excessively.

An additional variation in temporal resolution occurs with the time taken for a cycle of observations at the different sites. The Skibotn FPI mainly has less data due to the logistics of the site, as it can only be used on a campaign basis. However, even on the tristatic campaign nights, it does not correlate with the electron densities above the 95% confidence level. The average deviation from the time of the Skibotn data point to the mid point of the satellite pass for the first tristatic campaign is 10.5 minutes, compared with 1.4 minutes for KEOPS data and 3.6 minutes for Sodankylä data. This deviation is comparable with the time of the satellite pass and it can therefore be expected that it will produce similar results to comparisons with the average FPI intensities over the satellite pass time. Sodankylä and KEOPS data produce similar values implying the limit of the important variability is greater than 3.5 minutes.

The best correlation is between Sodankylä FPI data from the first tristatic campaign and IRI N_e values from the mid-point of the satellite pass. These give a correlation coefficient of 0.91. This is closely followed, at 0.90, by the IRI reconstruction electron densities both compared with the KEOPS intensities from the average of the satellite pass and its mid point, again from the first tristatic campaign. For these data sets therefore, the 6300Å intensities are well correlated with the electron density, so the source of the emission is a mechanism with a proportional amount of electrons produced per reaction. All the negative correlation coefficients (implying anti-correlation) are less than the 95% confidence level, and tend to be during periods with few data points available, where the results will not be statistically significant. The total correlation coefficients between the electron densities and the FPI intensities that are greater than the 95% confidence level, i.e. all the non-greyed out cells in Table 3.2 are shown in Table 3.3 below.

	Chapman Ne c.f. pass mid-point intensities	***Chapman Ne c.f. pass average intensities***	***IRI N_e c.f. pass mid-point intensities***	***IRI N_e c.f. pass average intensities***
Number of data points	33	48	61	56
Correlation of all data	0.48	0.42	0.70	0.59
Correlation of significant data	0.55	0.62	0.77	0.77

Table 3.3 *Correlations of Chapman and IRI reconstruction model N_e values with FPI intensities for significant results.*

The 'all data' row of Table 3.3 is the last row of Table 3.2, and the 'significant data' row is the correlation determined from the white cells for each of the sites in Table 3.2. As would be expected, this increases the correlation coefficient in all cases. The IRI reconstruction results are still significantly higher than the Chapman results. The coefficient is still however not quite unity. This would be partially due to uncertainties in the measurements. Spatial or temporal differences between the two data sets would lower the correlation coefficient. The FPI viewing volume is approximately 50km in altitude with a 10km x 10km field of view, whereas the grid spacing of the tomographic reconstruction models is 3km horizontally and vertically, resulting in higher spatial resolution that the FPIs. Changes in the FPI viewing volume would mostly be due to changes in emission height, which would result in a spatial offset with the tomography data used. The temporal variations are mostly due to the choice of FPI data used across the time of the satellite pass used in the reconstruction. After these effects have been taken in to account, a further source of the deviation from unity of the correlation coefficient would be due to another source of the 6300Å emission that does not also produce free electrons that contribute to the electron density value.

For the results shown above, the electron densities were averaged over the altitude range 240km – 300km. Table 3.4 below shows the correlations of the FPI data with IRI N_e values at 20km intervals over this range. This will indicate how important spatial variations are for the correlation.

Data period	***IRI N_e c.f. intensities from:***	***Altitude of Ne***				
		240km	***260km***	***280km***	***300km***	***Average***
2nd tristatic campaign	***Pass mid-point***	0.59	0.55	0.67	0.86	*0.71*
	Pass average	0.52	0.48	0.59	0.80	*0.63*
Non campaign nights	***Pass mid-point***	0.48	0.60	0.64	0.51	*0.64*
	Pass average	0.47	0.58	0.62	0.50	*0.62*
All passes	***Pass mid-point***	0.39	0.50	0.60	0.61	*0.70*
	Pass average	0.33	0.43	0.53	0.57	*0.59*

Table 3.4 *Correlations of FPI intensities with IRI reconstruction model N_e values at various altitudes.*

Table 3.4 shows that the correlation coefficients vary significantly with the different altitudes that the electron densities are taken. Overall, for each data set and for both choices of the FPI data with respect to the satellite pass, the coefficients are generally higher at the higher altitudes. The highest value for the second tristatic campaign night is 300km, and for the non-campaign nights is 280km. For the second tristatic campaign, the 300km value is higher than that of the average of 240-300km, implying a consistently higher emission altitude over the night. However, for the total of all the passes, the average is higher than any one altitude. This would be expected as the emitting layer, which is typically assumed to be approximately 50km wide, is broader than the 20km intervals shown, and the altitude can vary on relatively short timescales due to changes in heating and particle precipitation. Changes with altitude can be studied further by comparing the FPI data to electron densities obtained from the EISCAT radars.

3.4 RADAR DATA CORRELATIONS

A similar study with the EISCAT electron densities and 6300Å intensities was performed with the data from the first tristatic campaign, the 27th February 2003. The main advantage of using the EISCAT values of N_e is the much higher time resolution than the tomography chain, which only provides data a few times per night, at best. The EISCAT data had originally been post-integrated to one minute, but as this is much shorter than the FPI data, 180s post-integrated data are used for this analysis. For this campaign, the radars were pointed towards the tristatic A position (see Figure 3.1 in section 3.2.1). Therefore, Tromsø electron densities are compared with Skibotn East intensities; Kiruna N_e with KEOPS tristatic A intensities; and Sodankylä N_e with Sodankylä tristatic A intensities. These FPI data were shown in the middle plot of Figure 3.3. To compare the data, they are interpolated to the same times, which are chosen as that of the poorest time resolution data set, i.e. the Skibotn FPI data. The Kiruna and Sodankylä radars were pointed are at a fixed altitude, which was 240km for most of the night, but for two periods when the signal was weak, this was raised to 300km. Tromsø data are available for a range of heights along the beam from 115km to 525km.

The correlation coefficients between the FPI data and EISCAT data for the first tristatic campaign were lower than the tomography data. The highest value was for KEOPS FPI data correlated with Kiruna EISCAT data, reaching a coefficient value of 0.44. This is still statistically significant, as there are 75 data points, so a coefficient of 0.35 is the 95% confidence level. The Sodankylä data give a correlation of 0.26, and the Tromsø and Skibotn data give a value of 0.34, which are both below the 95%

confidence level. The Kiruna data are probably better correlated due to the higher time resolution of the KEOPS data, allowing more contemporaneous measurements between the two data sets.

As the three FPIs and the three EISCAT beams should all be looking at the same common volume of sky, the correlations between each of the data sets should all be the same. The correlation coefficients for all these combinations are given in Table 3.5 below.

FPI data set	***EISCAT radar***	***Correlation coefficient***
KEOPS	Kiruna	0.44
Sodankylä	Sodankylä	0.26
Skibotn	Tromsø	0.34
KEOPS	Tromsø	0.47
Sodankylä	Kiruna	0.35
Skibotn	Sodankylä	0.30
KEOPS	Sodankylä	0.38
Sodankylä	Tromsø	0.49
Skibotn	Kiruna	0.34

Table 3.5 *Correlations of FPI intensities with EISCAT radar N_e values between each combination of instruments.*

It can be seen from Table 3.5 that the correlation coefficients are not the same for all combinations of FPI and EISCAT data, as would be expected if they were all looking at a common volume. The azimuth and elevation of the radars are set presuming an emission height of the 6300Å line for the FPIs of 240km. If, however, the emission peak is not at this altitude but is lower or higher, then the volume that the FPIs do observe will undershoot or overshoot, respectively, the common volume along the line of sight beam. Small scale features such as auroral arcs will then pass the radar beams and the FPI lines of sight at different times, so the data sets will not be measuring the same feature at any given time.

As the EISCAT data have a much higher time resolution than the tomography data, they may be expected to produce higher correlation coefficients, as the timings of the two data sets can be better matched. The two highest values of correlation in Table 3.5 are from Tromsø data. The data used here were from 251km altitude. There are however many other range gates along the Tromsø beam that can be used to compare with the FPI intensities. The correlation coefficients between the intensities each of the three FPI instruments, and the electron densities from the range gates between 168km and 335km altitude, are given in Table 3.6 below. The correlations with the average of these altitudes are also shown. The altitudes shown for each range gate are the average values for all the data points throughout the night. The standard deviation of these points from the mean is 5km at the lowest range gate (118km) to 10km at 251km and up to 20km at the furthest range gate at 535km.

Tromsø Ne Altitude	***KEOPS***	***Sodankylä***	***Skibotn***
168km	0.66	0.77	0.39
184km	0.67	0.78	0.53
201km	0.73	0.82	0.57
217km	0.73	0.74	0.45
234km	0.62	0.63	0.42
251km	0.47	0.49	0.34
267km	0.42	0.46	0.33
284km	0.46	0.49	0.41
301km	0.49	0.50	0.43
318km	0.51	0.51	0.48
335km	0.51	0.49	0.48
Average 234:318km	0.55	0.57	0.45
Average 215:330km	0.68	0.72	0.52

Table 3.6 *Correlations of FPI intensities with EISCAT radar N_e values at various altitudes.*

Table 3.6 shows that the 245km range gate actually gives the some of the lowest values of correlation coefficient. 217km and 234km give the highest correlation values for KEOPS and Sodankylä FPI data. These data are plotted in Figure 3.13 below.

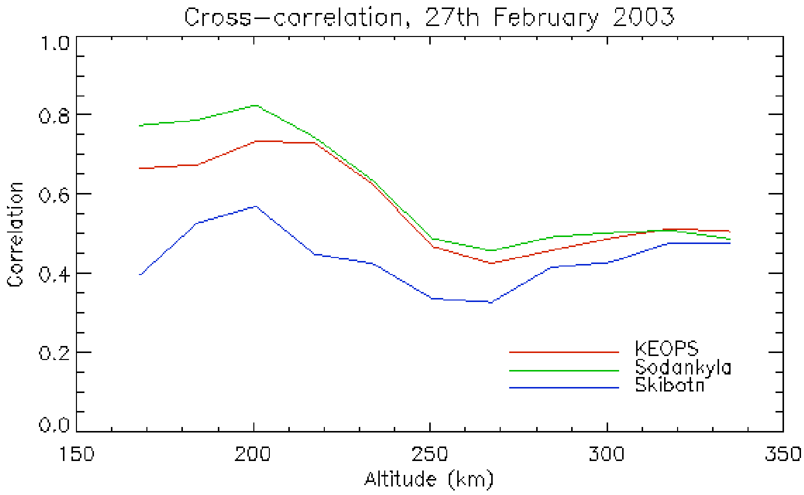

Figure 3.13 *Correlations between Tromsø electron densities at various altitudes, and FPI data from the three sites, for the first tristatic campaign night.*

There was some geomagnetic activity on this night, which would have pushed the emission height down, which would therefore result in a better match with lower altitude electron densities. However, this should only be the case for one site, as if the emission altitude is not 240km, the FPIs are viewing horizontally separated areas. Figure 3.13 shows that all three sites have peak correlations at around 200km. This would therefore imply the electron density and the 6300Å emission vary in the same way across the ~80km horizontal separation if the emission altitude is 200km rather than 240km. It should also be noted that the 6300Å emissions originate from a layer approximately 50km wide, and the height-integrated emissions along the line of sight are measured. At the top of the range, the fields of view of the FPIs would only be approximately 20 – 30km apart, with a viewing volume 10km wide. The average value of electron density over the altitude range shown also gives high values of correlation coefficient, and this is the highest value for Skibotn data. This is consistent with the activity levels and precipitation energies changing through the night, making the emission altitude change through the night, so different altitudes of electron densities correlate better with the intensities at different points in the night.

This affect cannot be seen with the tomography data, as these are only available a few times per night at best.

Correlations between the FPI intensities and the electron densities, from either tomography or radar data, will include dissociative recombination as well as other mechanisms that produce electrons. However, dissociative recombination is the dominant mechanism at the 6300Å emission altitude (see section 1.2.4). There are several reasons for the correlation coefficients not reaching unity. Observational errors will result in inaccuracy of the electron densities and the FPI intensities, changing the value of the coefficient. Errors in the time and location (e.g. altitude) of the measurements will mean the data that are compared may not be coincident. Statistical errors are also possible, particularly when the time range is limited, or sites are considered separately, resulting in small data sets.

After these errors have been considered, the most likely cause for correlation coefficients not reaching unity could be the combination of atomic nitrogen with molecular oxygen. This is shown in equation [31] below (and also in section 1.2.4).

$$N(^2D) + O_2 \rightarrow NO + O(^1D) \qquad [31]$$

This accounts for 80% of the emissions below 200km (Rees & Roble, 1986) and so may be a small contributor at the 6300Å emission layer altitude (240km). Link & Swaminathan (1992) measured this reaction in laboratory conditions and concluded it could be a source of $O(^1D)$ at thermospheric temperatures. This cannot be confirmed however without the ability to measure the nitrogen or NO levels. The cascade from higher energy atomic oxygen, $O(^1S) \rightarrow O(^1D) + h\nu$(5577Å) (equation (v) in section 1.2.4) also would not lead to electron density proportionality with the 6300Å intensity (Solomon et al., 1988), as formation of $O(^1D)$ does not also produce free electrons. This could be another source of decrease of the correlation coefficient.

4 ATMOSPHERIC GRAVITY WAVES: CASE STUDIES

4.1 INTRODUCTION

Atmospheric gravity waves are an important mechanism for the transfer of energy and momentum throughout the atmosphere. The physics of gravity waves was discussed in section 1.4. Gravity waves have mostly been studied in the stratosphere, mesosphere, and lower thermosphere regions as these have the largest powers and so have a significant role. At higher altitudes, AGWs are important as they can transport energy and momentum large distances, for example redistributing energy to equatorial latitudes. However, most high-latitude studies have been on Travelling Ionospheric Disturbances (TIDs), the ionospheric reaction to AGWs (see for example Balthazor & Moffett, 1999, MacDougall et al., 2001 and reviews by Hunsucker, 1982, Williams et al., 1993, Hocke & Schlegel, 1996). Gravity waves have been observed in the upper thermosphere over the southern polar cap, for example, by de Deuge et al., (1994) and Innis et al., (2001) with photometer observations of the 6300Å oxygen emissions and by Innis and Conde, (2002) in satellite data. Innis and Conde, (2001) observed gravity waves in vertical thermospheric winds from the Dynamics Explorer 2 (DE2) satellite.

Gravity waves at lower altitudes mostly have their origin in the troposphere. They are formed from, for example, thunderstorms or air rising over mountain ranges. The amplitudes of the waves increase with altitude as the density decreases. The gravity waves that are formed in the troposphere have mostly dissipated long before they reach F-region altitudes. Gravity waves observed here, therefore, have to be created in situ in the thermosphere. The mechanisms to create these gravity waves are auroral in origin. Large-scale gravity waves are thought to be generated in magnetic storms by one of two mechanisms. Particle precipitation in auroral regions will create localised heating that could set off waves, as could the Lorentz forces and Joule heating from electrojet currents (de Deuge et al., 1994). The theory of gravity wave behaviour and the expected properties of the waves have been reviewed by Hunsucker (1982) and Hocke & Schlegel, (1996), and are summarised in section 1.4.

This chapter shows that gravity waves have been found in the upper thermosphere in the northern auroral oval and polar cap regions, in neutral atom emission intensities as well as temperatures and winds. The array of measurements obtained will allow phases of the wavefronts to be calculated so that the source region and therefore mechanism of the waves can be identified. Several case studies are used to investigate the properties of individual waves, by using FPI data along with the other data sets available in the area, such as EISCAT, CUTLASS and STARE radars, IMAGE magnetometers and all sky cameras. The first of these (section 4.4) is the night of the 25th November 2003, where EISCAT radar data is available viewing the same volume of atmosphere as the FPIs. FPI data is normally taken in a sequence of look directions, and for individual integration periods of 30-60 seconds at each position, measurements at the same look direction are several minutes apart. This cycle time also depends on the efficiency and sensitivity of the detector and possibly the geomagnetic conditions. However, in March – April 2004, measurements were taken at KEOPS mostly in the tristatic A position, and for an integration time of 10 seconds, this provided higher resolution data, at approximately every 15 seconds (section 4.5). The differences in gravity wave activity between the polar cap, as observed by an instrument at Longyearbyen on Svalbard and the mainland sites close to the auroral oval are compared in section 4.6.

4.2 Lomb-Scargle Analysis

To identify gravity waves in data sets, a Lomb-Scargle analysis is used (Lomb, 1975 and Scargle, 1982). This technique is used because it is similar to Fourier transforms but can be used for unevenly spaced data. Its purpose in the context of the FPIs is to identify periodic oscillations in the atmospheric data. The method is a periodogram analysis (Lomb, 1975) that uses least squares fitting to sine waves, and it reduces to a Fourier transform in the limit of equally spaced data. Scargle (1982) normalised this periodogram using the variance of the original data, and Horne and Baliunas, (1986) made an estimate of the significance of the height of peaks in the power spectrum. Press and Rybicki, (1989) provide a fast algorithm for calculating the periodogram, which Wilms, (2000) has used to create an IDL code to produce the periodograms.

The time resolution of the FPIs means that generally periods of less than about 20 minutes cannot be identified, as at least two data points are needed per cycle, and measurements in the same look direction are around 8 – 10 minutes apart at best (see section 2.2.4). The minimum period is physically limited by the Brunt-Väisälä period (section 1.4.2), taken to be 12 minutes at F2 region altitudes (de Deuge et al., 1994). The maximum period is determined by the detection of atmospheric tides. The semi-diurnal tide of 6 hours may be detectable for the longest data sets in the mid-winter, but as at least twice this period is needed for its detection, and this is longer than many of the nights can observe, especially at the beginning and ends of the season. Several consecutive nights of data would be needed to detect the 12-hour tide. Due to the difficulties in discerning between tides and gravity waves, the main tides at 12, 8 and 6 hours are excluded from the analysis by only considering periods less than 6 hours. Although tides with periods of 3 or 4 hours are included in this, they often have very small amplitudes. Any waves found with either of these periods will have to be considered carefully. The main method of distinguishing gravity waves from tides is by considering that the gravity waves are localised in nature, whereas tides are global phenomena. A gravity wave would be expected to dissipate, at least partially, over the distances between look directions or instruments (several hundred kilometres), so its amplitude would be different at the different look directions. A tide would have roughly constant amplitude over the field of view of the instruments.

The time series analysis performed on these data was a Lomb-Scargle least squares frequency analysis of unevenly sampled data, first formulated by Lomb, (1976) and further developed by Scargle, (1982). This method was used due to the uneven sampling of the data, which is due to the many look directions observed in a cycle, the absence of data points due to either cloud cover or the non-fitting of spectra due to low intensities. An IDL routine (Wilms, 2000) has been used to calculate the Lomb – Scargle Periodogram (LSP) for this work. The main disadvantage to the Lomb-Scargle periodogram over Fourier transforms is the calculation time, as this is dependant on the product of the number of frequencies sampled and the number of data points, and also has many trigonometric operations. The fast algorithm calculated by Press & Rybicki (1989) is included in the analysis program, however the data sets used here are not sufficiently large for this to become necessary, and frequencies down to the Nyquist frequency are not often needed as this is above the Brunt-Väisälä frequency of the atmosphere.

The spectral power ($P_N(\omega)$) of the normalised periodogram as a function of angular frequency, ω, is given by (Press & Rybicki, 1989).

$$P_N(\omega) \equiv \frac{1}{2\sigma^2}\left\{\frac{\left[\sum_i (h_i - \langle h\rangle)\cos\omega(t_i - \tau)\right]^2}{\sum_i \cos^2\omega(t_i - \tau)} + \frac{\left[\sum_i (h_i - \langle h\rangle)\sin\omega(t_i - \tau)\right]^2}{\sum_i \sin^2\omega(t_i - \tau)}\right\} \quad [32]$$

For amplitudes h_i, $i = 1, \ldots,$ N data values at times t_i, with a mean of $\langle h\rangle$, variance σ^2 and a time offset, τ, defined by:

$$\tan(2\omega\tau) \equiv \frac{\sum_j \sin 2\omega t_j}{\sum_j \cos 2\omega t_j} \qquad [33]$$

This has two important effects, it makes the power spectra time shift invariant, so the time series can be shifted by any constant without affecting the results, so making calculations simpler. It also makes $P_N(\omega)$ (equation [32]) identical to that obtained if a linear least squares fitting was calculated, $h(t) = A \cos \omega t + B \sin \omega t$. This weighs the data on each point rather than each time interval, which is why this method is more accurate than a simple Fourier transform, for unevenly sampled data.

The spectral power is also calculated with the normalisation using the variance (σ^2) of the data. This is the total variance of the data, and not either the variance after an offset has been removed, or from the uncertainty of the measurements (Horne & Baliunas, 1986.) The variance is given by equation [34].

$$\sigma^2 = \frac{1}{N-1}\sum_{1}^{N}(h_i - \bar{h})^2 \qquad [34]$$

The amplitudes of the waves, as a function of frequency, can be calculated with the following equation (Hocke, 1998), from the spectral power and the standard deviation of the data:

$$A(\omega) = \sqrt{\frac{2}{n} 2\sigma^2 P_N(\omega)} \qquad [35]$$

The program used is both time and intensity shift invariant as the initial time value and the mean of the intensity are removed from all data points. The maximum and minimum periods or frequencies that are sampled can either be user defined, or they can be calculated from the data. The minimum frequency sampled is simply the inverse of the length of the data set being analysed. The maximum frequency sampled is approximately the Nyquist frequency given by equation [36], for time resolution of the data, Δt, and is determined in the program from the number of the data points (N), and the total length of the data set, T.

$$\nu_{Ny} = \frac{1}{2\Delta t} = \frac{N}{2T} \qquad [36]$$

The number of independent frequencies that are sampled is also based on the time resolution of the data, and the total length of the data set. The formula used was calculated by Horne & Baliunas, (1986), who simulated data sets with various spacings of data, and various numbers of data points. They performed periodogram analyses on the data sets, up to the Nyquist frequency, and using the highest peaks in each periodogram, for many periodograms, calculated an empirical formula for the number of independent frequencies to sample, as a function of the number of data points in a series.

Error analysis of the spectral power is achieved by the false alarm probability. The distribution function has been shown (Horne & Baliunas, 1986) to be proportional to e^{-z}, where z is the height of a peak in the periodogram. The probability that a peak is smaller than z is $1 - e^{-z}$ so for all frequencies, N_i, the false alarm probability is:

$$F = 1 - \left(1 - e^{-z}\right)^{N_i} \qquad [37]$$

Rather than assuming that there is a signal in the data, the false alarm probability is the probability of a signal occurring assuming the data is pure noise. A probability of 1 is therefore noise, and a calculation of significance will be 100% for noise. Confidence levels are used, which is (100 – significance), and is an indication of the confidence that a signal detected is real. The value of spectral power that a particular confidence level will have will be dependant on N_i, from equation [37]. This is the number of data points in the time series, which is dependant on both the time resolution of the data, and the length of the data set, i.e. the length of the night.

A cross correlation analysis of the data can be used to obtain information on the wave speed, direction, and wavelength. A cross correlation between the north and south directions and the east and west directions will give, respectively, the meridional and zonal time lags τ_y and τ_x across the field of view of the FPI (e.g. Oliver et al., 1995). These lags will be dependant on the distance between the look directions (d), the phase speed of the wave (v) and the phase angle (θ, measured anticlockwise from the east). These parameters can therefore be calculated with the following equations (Oliver et al., 1995):

$$v = \sqrt{\frac{d^2}{\tau_x^2 + \tau_y^2}} \qquad \theta = \arctan\frac{\tau_x}{\tau_y} \qquad [38]$$

The distance, d, between the FPI look directions, presuming a fixed emission altitude of 240km over the viewing area and with an elevation angle of 45°, is approximately 480km. The horizontal speed of the wave can then be used, along with the frequency information from the Lomb-Scargle periodogram, to calculate the horizontal wavelength, λ_h, using $v = \omega / k$.

4.3 SYNTHETIC DATA TEST

To have confidence in the results obtained from a Lomb-Scargle analysis, it is important to test the program with synthetic test data. These are simulated time series of a constant background level, with noise added, at various amounts, and waves superimposed through the series. Therefore, with a known input frequency and amplitude, the accuracy of the analysis can be determined by the frequencies and amplitudes output. Files of synthetic waves have been created by Morgan, (2003), with waves inserted with frequencies of 30 minutes, 3 hours and 6 hours, of varying amplitudes or background noise levels. Morgan, (2003) gives a thorough investigation of the analysis of these data with a Lomb-Scargle analysis program written in Fortran. The periods and amplitudes of the output waves with this test data from the Lomb-Scargle analysis used in this thesis are presented in this section.

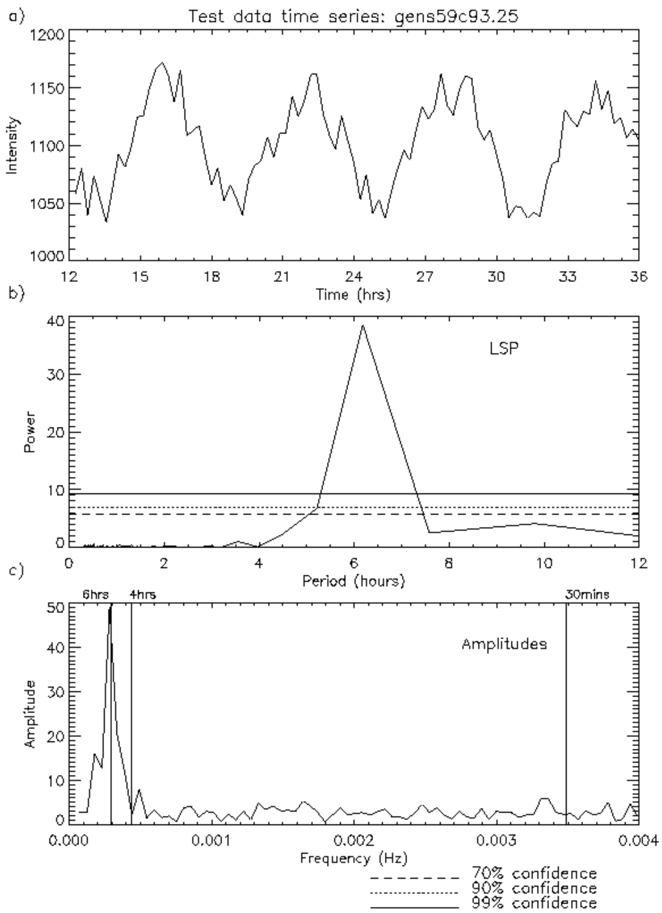

Figure 4.1 *a) Time series of test data of a period of 6 hours and noise levels 50% of wave amplitude, with its LSP (b) and amplitudes (c).*

Only waves that have large amplitudes in relation to the background noise variations will be detected. Using synthetic waves over known and various random amplitudes is a method of determining what sensitivity is achievable with the analysis. Figure 4.1a shows a wave of period 6 hours with a noise level half that of the wave amplitude. The 6-hour wave is clearly observed, in Figure 4.1b, well above the 99%

confidence level. The amplitude of the wave plotted against frequency is shown in Figure 4.1c. It can be seen that the spectral power, equation [32], depends on the sampling frequency of the data. Empirically, by making comparisons of high-resolution data with data integrated over a longer period, it can be shown that the spectral power obtained is proportional to the time resolution of the data. For example, 10-second magnetometer data were compared with 1-minute averaged data, and the powers of the peaks were six times smaller than the 10-second data.

	Input				**Output**			
Filename	Base	Random amp.	Period (minutes)	Signal amp.	Period (hours)	Power	Amplitude	Error in time (%)
s59c98.25	1100	0	360	100	6.18	43.5	48.9	2.9
s59c97.25	1100	5	360	100	6.18	43.5	49.3	2.9
s59c96.25	1100	10	360	100	6.18	43.3	49.3	2.9
s59c95.25	1100	20	360	100	6.18	42.2	48.2	2.9
s59c94.25	1100	30	360	100	6.18	41.1	49.2	2.9
s59c93.25	1100	50	360	100	6.18	38.5	49.2	2.9
s59c92.25	1100	100	360	100	6.18	28.1	52.7	2.9
s59c91.25	1100	0	240	100	3.98	45.3	49.9	-0.6
s59c90.25	1100	5	240	100	3.98	45.2	49.9	-0.6
s59c89.25	1100	10	240	100	3.98	45.0	49.0	-0.6
s59c88.25	1100	20	240	100	3.98	44.2	51.3	-0.6
s59c87.25	1100	30	240	100	3.98	43.0	50.6	-0.6
s59c86.25	1100	50	240	100	3.98	39.0	45.6	-0.6
s59c85.25	1100	100	240	100	3.98	31.1	51.7	-0.6
s59c84.25	1100	0	30	100	0.54	43.2	48.6	8.1
s59c83.25	1100	5	30	100	0.54	43.1	49.0	8.1
s59c82.25	1100	10	30	100	0.54	42.7	48.3	8.1
s59c81.25	1100	20	30	100	0.54	42.1	49.5	8.1
s59c80.25	1100	30	30	100	0.54	41.7	49.6	8.1
s59c79.25	1100	50	30	100	0.54	36.0	47.8	8.1
s59c78.25	1100	100	30	100	0.54	23.8	42.6	8.1
s59c77.25	1100	200	360	100	6.18	17.6	60.3	2.9
s59c76.25	1100	400	360	100	0.41	4.3	52.7	-1361.4
s59c75.25	1100	200	240	100	3.98	10.9	44.2	-0.6
s59c74.25	1100	400	240	100	6.18	3.9	51.7	35.3
s59c73.25	1100	200	30	100	0.54	17.9	62.5	8.1
s59c72.25	1100	400	30	100	0.54	6.6	62.1	8.1

Table 4.1 *Waves obtained from a Lomb-Scargle analysis of synthetic waves of 0.5, 4 and 6 hours, with varying noise levels.*

Table 4.1 shows the results of the Lomb – Scargle analysis described above on waves with an amplitude of 100 units, with periods of 30 minutes, 4 and 6 hours. The base level is 1100 units, with noise levels ranging from 0 to 400 units. The sixth, seventh and eight columns show the period, spectral power, and amplitudes of the waves detected for each wave. The percentage difference in the period of the wave inserted into the synthetic data set and the period detected is shown in the last column. For the files ‘s59c98.25’ to ‘s59c78.25’, which have noise levels up to the size of the wave (i.e. 100 units), the error in the frequency detection is always less than 10%. In the 30-minute period wave data, the strongest period detected is at 0.54 hours (32.6 minutes). On inspection of the whole periodogram (see Figure 4.2), there is also a wave of 0.50 hours. Detection of these two waves around this period is due to aliasing, as the 0.54-hour period is the sampling frequency of the data set.

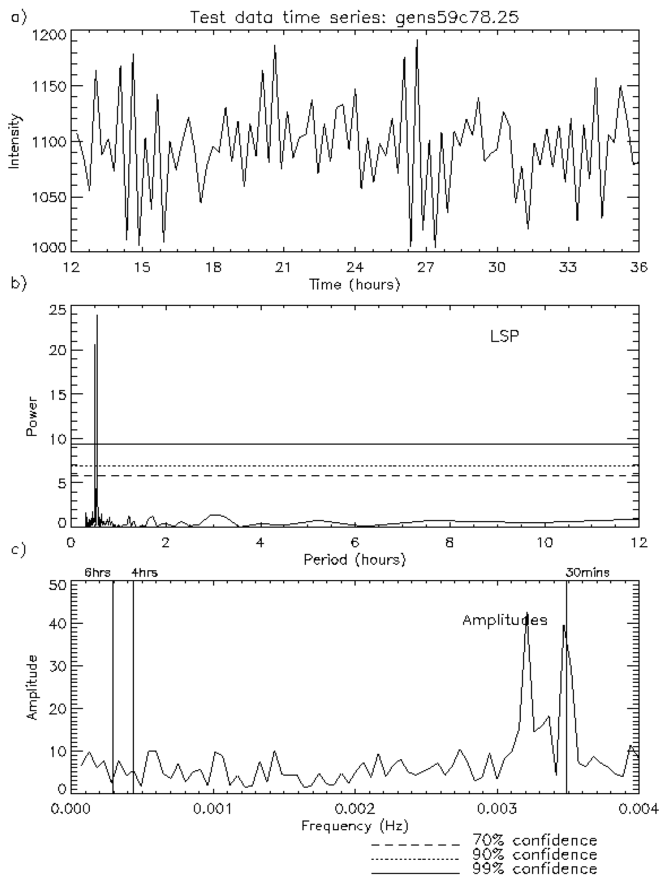

Figure 4.2 *a) Time series of test data of a period of 30 minutes and equal noise levels to wave amplitude, with its LSP (b) and amplitudes (c).*

The file labelled 's59c76.25', shown in Figure 4.3, contains a 6-hour wave and noise levels 4 times the amplitude of the wave. The error is extremely large, which shows that aliasing has occurred in this data set, and power from the correct wave has leaked down to the higher frequencies. The spectral power at this period is very small (4.3), and less than the 70% confidence level at 5.7. In Figure 4.1a, the wave pattern can be clearly seen by eye, however in Figure 4.3a, the signal is hidden by the greater noise levels.

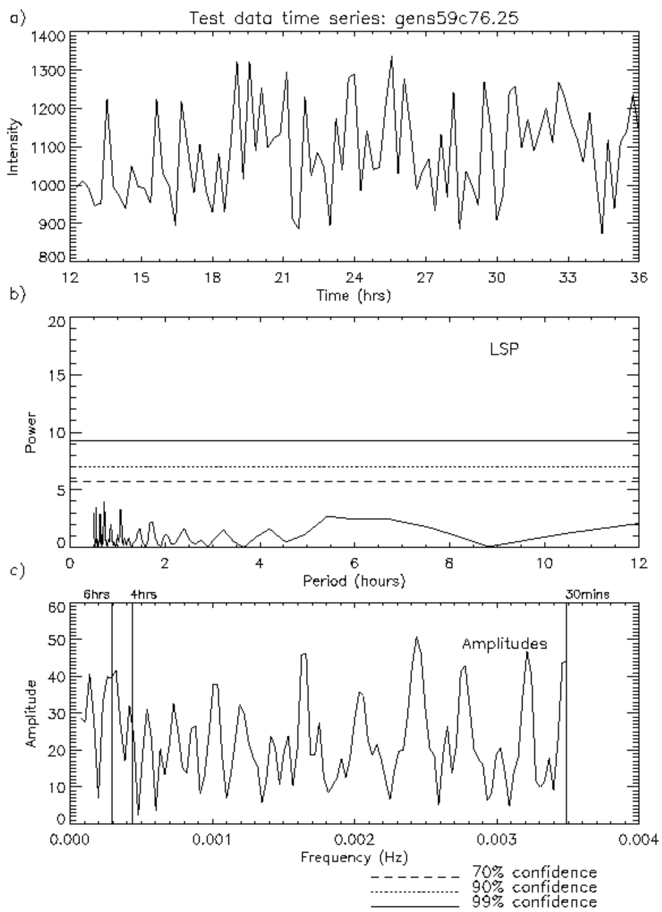

Figure 4.3 *a) Time series of test data of a period of 6 hours, with its LSP (b) and amplitudes (c), with large noise levels (4 times wave amplitude).*

The error in the frequency found in the periodogram (the last column of Table 4.1) is dependant on how far the maximum frequency sampled is from a harmonic of the frequency of the wave. Smaller values of this maximum frequency chosen will increase the likelihood of it being near to a harmonic, and so increasing the error. However, it will mean the higher frequencies will not be tested, leading to shorter period waves not being detected. The smaller of the Brunt-Väisälä and the Nyquist frequencies are used. The latter is the default value, unless it is seen that this is not effective in observing the waves in sufficient detail for the time resolution and length of the data set.

In summary therefore, waves can be detected in data with up to approximately twice the amplitude of the noise to that of the signal, i.e. a signal to noise ratio (SNR) of 0.5. A set of synthetic data with 80-minute waves of varying amplitudes shows that a wave can be detected to the 99% confidence level with a SNR of 0.43, and to the 70% level with a SNR of 0.32. However, the amplitude of the wave detected is never as large as the input amplitude. It is generally around 50% of the input, and at most 60%. This is thought to be due to the width of the peaks, as the input amplitude is spread over the periods across the width of the peak.

Knowing the dependence of the analysis on the background level is important in the statistical analysis (Chapter 5) as data from different seasons are compared. The wind data are not affected by variations in background as they are calibrated with the zero Doppler shift position of the calibration lamp. For more recent temperature data, laser calibrations have been used to determine the instrument function of the FPI, so absolute temperatures have been determined. However, because the background levels will vary with a number of factors, older temperature data and all the intensity data are not calibrated and are simply relative values. The greatest of these variations is time of day, as obviously the background airglow is much greater when the Sun is nearer the horizon near dusk and dawn. The Moon also reflects sunlight, increasing airglow, so position in the lunar cycle is important. The Sun reaches further below the horizon at the solstice than the equinoxes, so the mid-winter period will have a lower background. Cloud cover will scatter any light increasing backgrounds, but these nights are already excluded from analysis. Detectors have been replaced and upgraded when they fail or improvements in technology have allowed, and each detector has had a different sensitivity. For some periods however, the same detector has remained in place for many seasons, but the sensitivity still cannot be assumed constant, as the detector will degrade over time. However, in the same way that the program is time – shift invariant, the intensities are also shifted. The mean value of the data set is removed from the data points before the periodogram is calculated. This has three advantages for the data used here. Relative, as well as absolute, temperatures can be used, so the analysis is not restricted to the latest, calibrated data. The wind and temperature data can be smoothed to remove large-scale structures from the data (see below). The whole database can be analysed and different seasons, detectors and sites can be compared i.e. a statistical survey of the database can be performed (see the next chapter).

To see how multiple waves in a data set affect the analysis, synthetic data sets were created of two waves of either 6, 4 or 0.5 hours. Table 4.2 shows the input waves and periods detected for data sets containing two waves of different periods. The shorter period waves have amplitudes of 100 units, and the longer period wave has half this amplitude. Files have different levels of background noise for each combination of wave.

Filename	Base	Random amp	wave1 period	wave2 period	wave1 detected period	power	amp.	wave 2 detected period	power	amp.	wave 3 detected period	power	amp.
s60c01.25	1100	0	6	4	6.18	8.63	24.4	3.975	36.2	49.9			
s60c02.25	1100	5	6	4	6.18	8.15	23.9	3.975	36.3	50.3			
s60c03.25	1100	10	6	4	6.18	11.5	28.3	3.975	36.2	50.2			
s60c04.25	1100	20	6	4	6.18	11.3	28.9	3.975	35.7	51.4			
s60c05.25	1100	30	6	4				3.975	34.7	50.5			
s60c06.25	1100	50	6	4	6.18	8.97	25.5	3.975	33.5	49.3			
s60c07.25	1100	100	6	4				3.975	22.4	47.6			
s60c08.25	1100	200	6	4				3.975	18.4	62.8	2.493	6.793	38.2
s60c09.25	1100	400	6	4									
s60c10.25	1100	0	6	0.5	6.18	8.95	24.7	0.544	34.0	48.2	0.503	23.87	40.4
s60c11.25	1100	5	6	0.5	6.18	8.71	24.5	0.544	33.8	48.3	0.503	26.29	42.6
s60c12.25	1100	10	6	0.5	6.18	8.65	24.1	0.544	34.6	48.1	0.503	25.31	41.1
s60c13.25	1100	20	6	0.5	6.18	9.01	25.5	0.544	32.7	48.6	0.503	27.29	44.4
s60c14.25	1100	30	6	0.5	6.18	8.48	24.2	0.544	32.8	47.6	0.503	24.07	40.7
s60c15.25	1100	50	6	0.5	6.18	8.48	25.7	0.544	30.3	48.6	0.519	6.276	22.1
s60c16.25	1100	100	6	0.5	6.18	7.06	29.0	0.544	22.7	52.0	0.503	16.82	44.8
s60c17.25	1100	200	6	0.5				0.544	10.9	47.3	0.503	8.452	41.6
s60c18.25	1100	400	6	0.5									
s60c19.25	1100	0	4	0.5	3.975	9.61	25.6	0.544	33.6	48.0	0.503	28.11	43.9
s60c20.25	1100	5	4	0.5	3.975	8.81	24.6	0.544	35.0	49.1	0.503	25.6	42.0
s60c21.25	1100	10	4	0.5	3.975	9.38	25.8	0.544	33.5	48.7	0.503	25.15	42.2
s60c22.25	1100	20	4	0.5	3.975	9.44	25.9	0.544	33.3	48.6	0.503	27.25	43.9
s60c23.25	1100	30	4	0.5	3.975	8.73	24.8	0.544	32.3	47.7	0.496	23.62	40.7
s60c24.25	1100	50	4	0.5	3.975	7.48	24.8	0.544	30.0	49.6	0.496	22.15	42.6
s60c25.25	1100	100	4	0.5	3.975	5.88	25.6	0.544	21.1	48.4	0.496	17.05	43.5
s60c26.25	1100	200	4	0.5				0.544	13.0	53.3	0.496	10.58	48.2
s60c27.25	1100	400	4	0.5									

Table 4.2 *Waves obtained from a Lomb-Scargle analysis of synthetic data with multiple waves of periods of 0.5, 4 and 6 hours, with varying noise levels.*

An example of one of these synthetic data sets, labelled ‘s60c24.25’ in Table 4.2, can be seen in Figure 4.4 below. This shows a 30-minute wave with the addition of a 4-hour wave, and a random noise level of 100 units, which is equal to the amplitude of the 4-hour wave, but half the amplitude of the 30-minute wave.

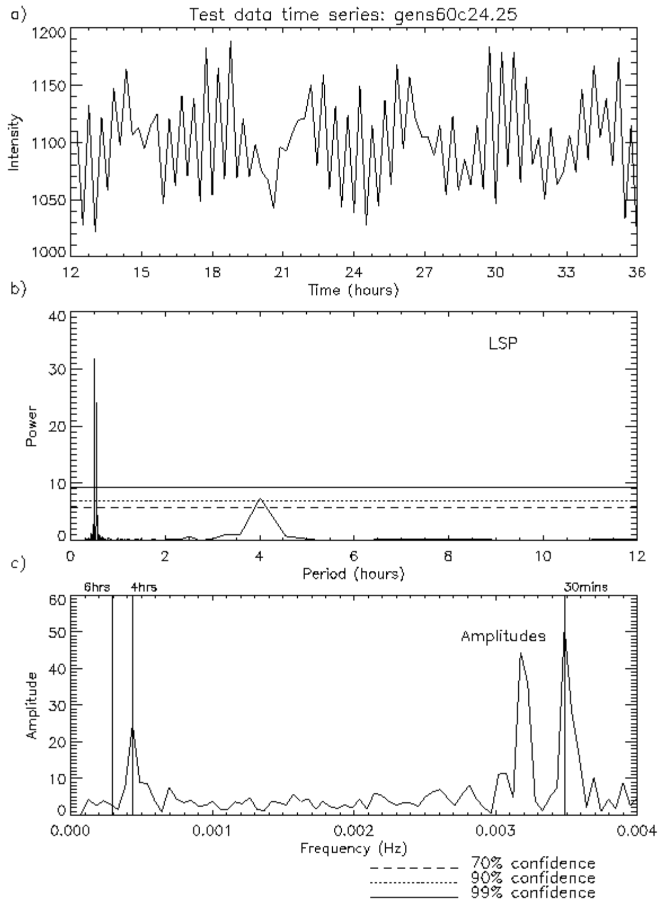

Figure 4.4 *Test data with waves of periods of 4 hours and of 30 minutes, with noise levels equal to that of the 4 hour wave amplitude: a) time series, (b) its LSP and (c) wave amplitudes.*

By comparing Figure 4.2 and Figure 4.4, it can be seen in both cases that the 30-minute waves also have a peak at 32 minutes. Table 4.2 shows that for the data sets that have the largest noise levels, of 400 units, the waves are not detected above the 70% confidence level. Only the half hour waves are detected with noise levels of 200units. The amplitudes again are approximately half that of the expected (i.e. input) wave. This is the case for both of the two waves: the longer period wave input at 50 units ranges from 24 to 29, and the shorter period input at 100 units ranges from 40 to 62. Therefore, these results are consistent with the single waves, so additional waves in the data set do not affect the results of the analysis.

4.4 SECOND TRISTATIC CAMPAIGN

The geometry of the tristatic setup was described in section 3.2 and can be seen in Figure 3.1. On the night of the 25h November 2003, a second tristatic campaign was performed, with the FPIs in KEOPS, Sodankylä, and Skibotn. All the FPIs are co-located with EISCAT radars (e.g. Rishbeth & Williams, (1985) and section 2.4.1), and for the night of the data presented here, the EISCAT tristatic point was at the KEOPS zenith position. These data are also presented in Ford et al., (2006).

A range of other instrumentation in the region supplements the FPI data. The fields of view of the CUTLASS and STARE coherent scatter radars also cover this region (see sections 2.4.4 and 2.4.3 respectively). CUTLASS measures plasma velocities using HF radar. It receives echoes from E and F region ionospheric irregularities, perpendicular to the direction of transmission, while STARE using VHF measures the E region to calculate electric fields. The IMAGE magnetometer network consists of 29 magnetometers across Scandinavia, covering the same region as the FPI look directions (e.g. Viljanen & Häkkinen, (1997)) and there are all sky cameras (ASCs) at several of these locations, as part of the MIRACLE network (see section 2.4.3).

The KEOPS FPI detector had the most sensitive detector at the time of this campaign, allowing integration times of 20 seconds, while Sodankylä had 40 seconds and Skibotn has 60 seconds. Due to the cycles that the data are taken in, this leads to cycle times of 3.5 minutes for KEOPS, 8.5 minutes for Sodankylä, and 13.9 minutes for Skibotn (Skibotn takes longer due to a more complicated and slower mirror control system). These cycle times in principle allow waves to be detected in the data of as short as 7-minute periods for KEOPS or 17 and 28 minutes for Sodankylä and Skibotn respectively, i.e. twice the period of observations. Waves should therefore be detectable down to the Brunt-Väisälä period, which at this altitude (240km) is around 12 minutes (e.g. Hargreaves, 1979; Innis et al., 2001). This gives a maximum detectable frequency (the Nyquist frequency) of 8.6 cycles per hour, for KEOPS data. The longest periods detectable are up to half the period of darkness, which in the polar winter night extends up to 24 hours. In practice though, only periods less than 6 hours are included, as periods longer than this are hard to distinguish from atmospheric tides. The time series analysis performed on these data was a Lomb-Scargle least squares frequency analysis of unevenly sampled data as first formulated by Lomb (1976) and further developed by Scargle (1982), described in section 4.2.

4.4.1 THERMOSPHERIC DATA

The second APL tristatic campaign, where the UHF EISCAT radars viewed the same common volume as the three FPIs, was on the night of the 25th November 2003. This experimental setup was described in section 3.2. The EISCAT data from this night are shown in Figure 4.5. It is a potentially good night to find gravity waves because TIDs appear to be present as clear wavefronts in the electron density (Figure 4.5a). An example of a wavefront, between 21:50UT and 22:30UT, is highlighted with a dotted line. The wavefronts can also be seen, though not as clearly, in the electron temperature data in Figure 4.5b. The effect is even weaker in the ion temperatures of Figure 4.5c. Figure 4.5d shows the line of sight velocities along the Tromsø beam (which at this time was towards the tristatic point over the KEOPS site). These do not show the wave structures due to the overwhelming effect of the diurnal change in direction due to the two-cell convection pattern of the polar cap region. These data are discussed further in section 4.4.3. Previous work on observations of TIDs in EISCAT data are discussed in e.g. Williams et al., (1993), Lanchester et al., (1993), Shibata, T. & Schlegel, (1993), and Mitchell & Howells, (1998).

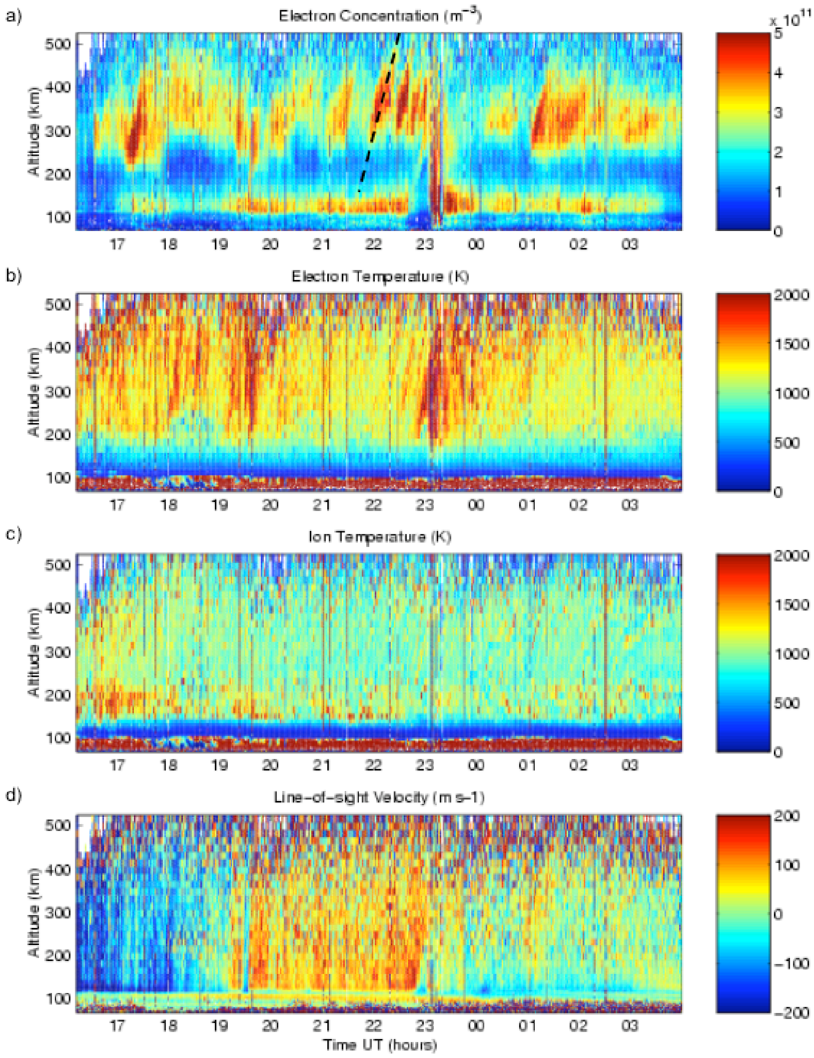

Figure 4.5 *EISCAT data from 25th November 2003, showing a) electron density, b) electron temperature, c) ion temperature, and d) ion velocity. A wavefront of a gravity wave in electron density is shown with a dashed line.*

Figure 4.6 shows the 6300Å intensities from a tristatic campaign on the 25th November 2003. The look directions shown are towards the tristatic point above KEOPS, which is to the south of Skibotn and west of Sodankylä. A good correlation is seen between the sites, which indicates that they are looking at a common volume of atmosphere. The slight discrepancies at 15-16UT and 17-19UT are most likely caused by the different instruments no longer looking at a common volume. This is often due to cloud cover which results in a loss of directional information due to scattering, and a decrease in signal to noise ratio. However, no clouds were observed during these periods, and all sky cameras and satellite images show clear skies. The increase in Sodankylä data at 15-16UT is likely to be due to viewing the sunset in the west. The other reason for the intensities not matching is due to the height of the atomic oxygen layer changing. This will result in the viewing volumes over- or under-shooting the tristatic volume so that they are no longer all viewing a common volume. This is covered thoroughly by Aruliah et al., (2004).

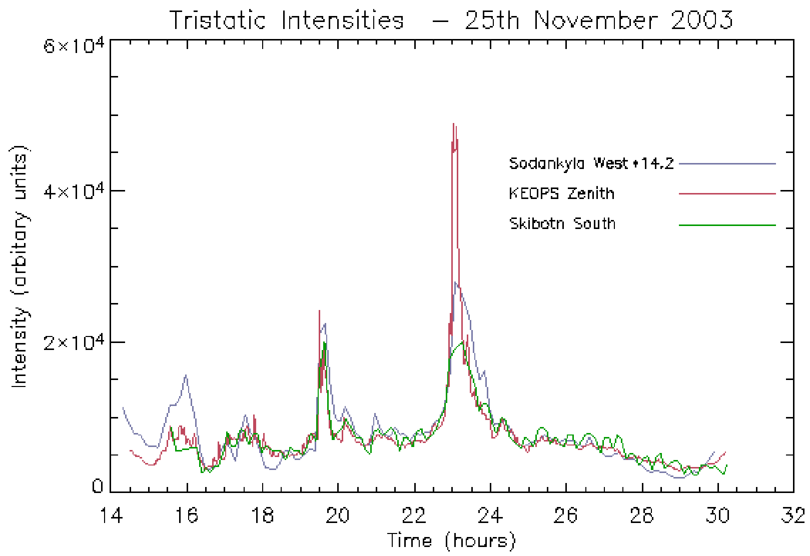

Figure 4.6 *6300Å intensities from KEOPS, Sodankylä, and Skibotn on the 25th November 2003 towards the tristatic point above KEOPS. As the intensities are not calibrated, the Sodankylä data are scaled to match the other sites by a factor of the ratios of the median value of Skibotn to Sodankylä data for this night.*

Figure 4.7 shows all the data from KEOPS on this night. Figure 4.7a shows the 6300Å intensities, Figure 4.7b shows the neutral winds, and the temperatures are in Figure 4.7c. The coloured lines are the different look directions that are viewed during a cycle of observations, and are the same for all such plots in the following sections. The same periods of precipitation that were seen in Figure 4.6 can also be seen in Figure 4.7a. The strength of the emission is largest in the more northern look directions, and smaller in the south and bistatic B positions, as would be expected from the usual location of the auroral oval at higher latitudes.

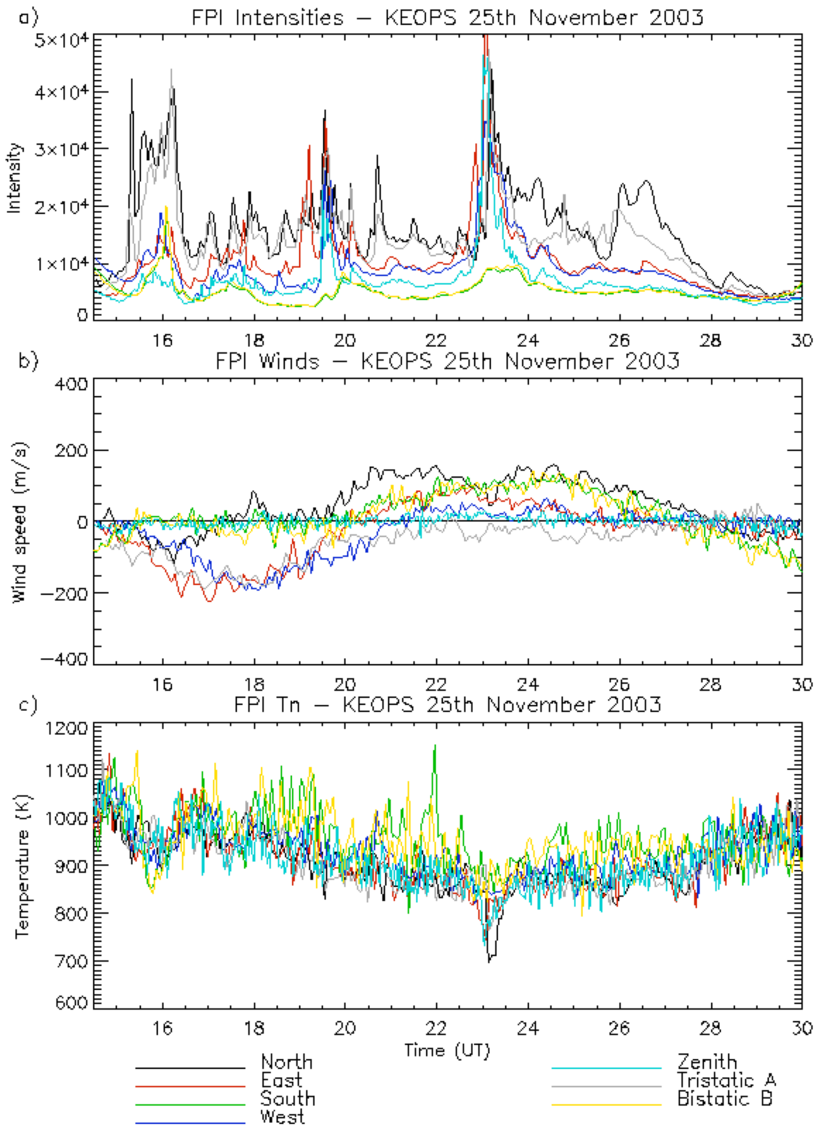

Figure 4.7 *6300Å intensities (a), neutral winds (b), and temperatures (c) for all the look directions from KEOPS on the 25th November 2003.*

The winds show the influence of the two-cell convection pattern via ion drag, with the east (red line) and west (blue line) winds being negative, and so westward, in the evening sector, and eastwards in the second half of the night. The zenith winds (turquoise line) are an order of magnitude smaller than the horizontal winds, which is usual, due to the large amount of potential energy that would be needed to create large vertical motions of the atmosphere. The overall structure of the temperatures in Figure 4.7c show a minimum around midnight, with larger temperatures at the start and end of the night, due to solar heating on the dayside transferring energy to the dawn and dusk sectors through the inertia of the thermosphere.

The periodograms obtained from the Lomb-Scargle analysis on the 6300Å line intensities from the 25th November 2003 are shown in Figure 4.8. The coloured lines are the spectral power as a function of period for each of the look directions, with the same colour coding used for all plots. The horizontal lines show the 70%, 90% and 99% confidence levels. Two clear peaks with greater than 99% confidence can be seen at 1.8 and 3.7 hours in all look directions in KEOPS data, most of Sodankylä and some of Skibotn. Peaks are also seen at 1.2, 1.4 and 2.5 hours. Periods greater than about 7 hours are not real, as this is greater than half the total length of the data set and so are not sufficiently sampled. Periods less than about 20 minutes are also not real, due to aliasing. As the same periods are seen in more than one site, this gives confidence that the periods are atmospheric effects and not instrumental. The periods are not seen so well in Skibotn data due to the low time resolution and the poor signal to noise ratio of the data. The power of the KEOPS zenith data periodograms is generally higher than the other directions as zenith measurements were taken twice in each cycle, so the time resolution was twice as high.

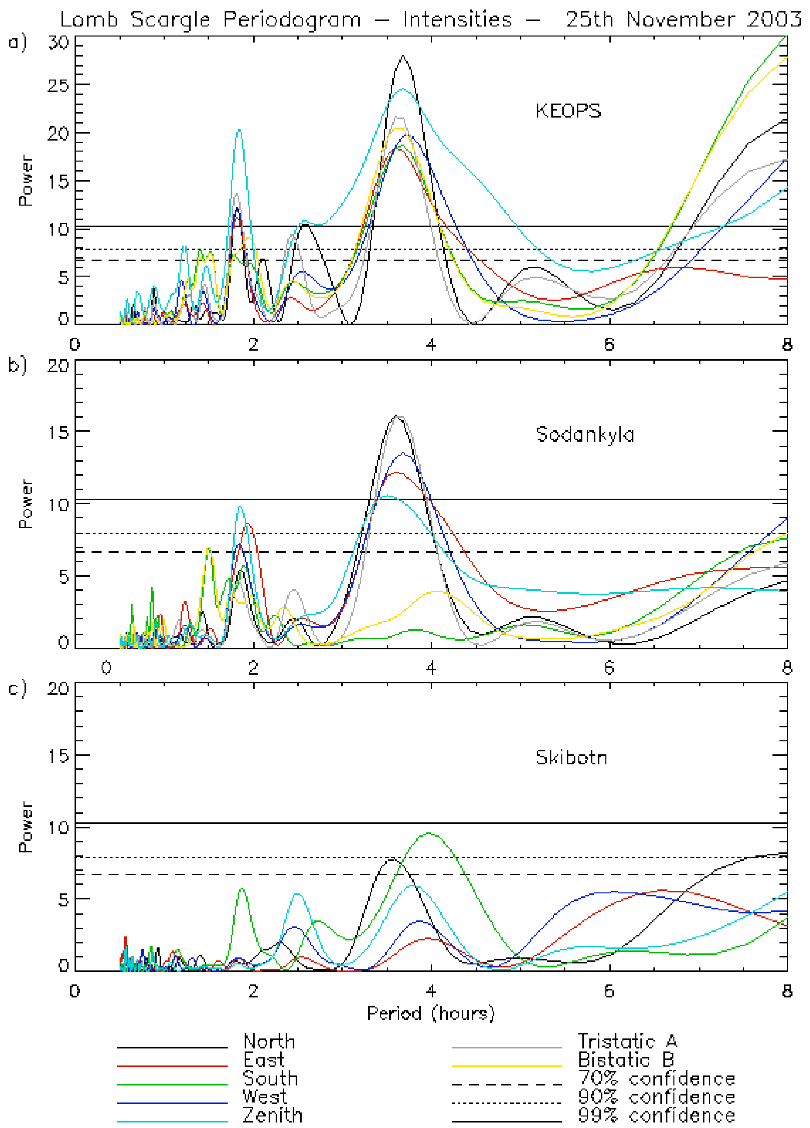

Figure 4.8 *Lomb-Scargle periodograms for a) KEOPS b) Sodankylä and c) Skibotn FPI intensities on the 25th November 2003. The colours represent each of the different look directions observed, and the horizontal lines are 70%, 90% and 99% confidence levels.*

The 3.7-hour wave is present in all look directions in KEOPS but in Sodankylä appears in all but the south and B positions. The intensities were lower at the south and bistatic positions as these are further from the auroral oval. The detector on the KEOPS FPI is more sensitive than the Sodankylä detector which means it is able to measure lower intensities, which have a lower signal to noise ratio, more accurately. This explains the wave not being observed in the Sodankylä south and bistatic positions while it is in KEOPS data. Skibotn sees the wave mainly in the north and south. This is consistent with the source being the auroral oval which lies across Skibotn and the north part of KEOPS and Sodankylä, as can be seen in Figure 4.9 which shows the All Sky Camera (ASC) keogram for Kevo, which is near to the tristatic A point, as shown in Figure 3.1.

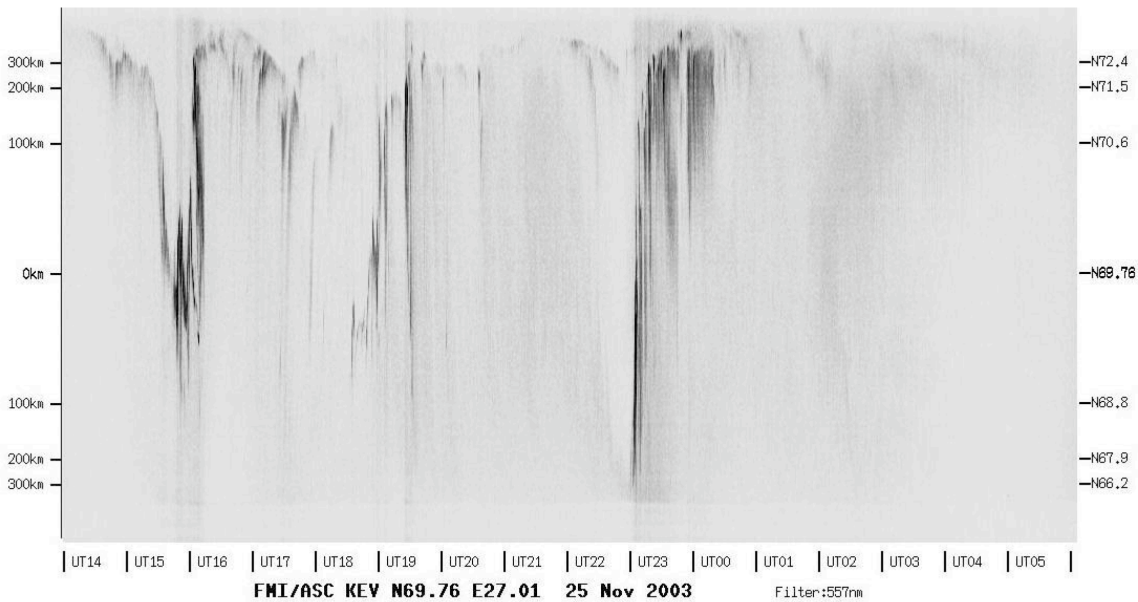

Figure 4.9 *All Sky Camera Keogram from Kevo for the 25th November 2003. The image shows 5577Å aurora, and has been made negative for clarity.*

Keograms take a north-south slice from the centre of each ASC image through the night and hence provide an overview of auroral activity (through a green line filter). The auroral arcs at 16UT, 19UT, and 23UT move south, are overhead at Kevo, and so are over the KEOPS and Sodankylä north latitude. For the rest of the night the oval is to the north of Kevo. Sodankylä is at south of Kevo, at about the latitude of the lower 300km mark on this plot. As there is little activity at the bottom of the plot, overhead of Sodankylä, this means that the south and B positions are far from the auroral oval. An auroral source of the waves that have dissipated by the time they reach the south and B positions could explain why the 3.7-hour wave is not detected in these data. The ASC keogram shows the oval to be on average 200-300km north of Kevo, which would be 500-600km north of the bistatic B and KEOPS South positions. Although gravity waves can have wavelengths of over a 1000km (e.g. Fagundes et al., (1995), Innis et al., (2001)), the source on the 25th November 2003 could be at a more distant longitude. In addition, Innis & Conde (2001) show that waves propagating poleward penetrate further than those propagating equatorward, where there are greater chances of dissipation due to background winds.

The 1.8-hour wave is detected in Sodankylä south and B, but is only seen in the south in Skibotn. This implies that it either originated further south and dissipated before it could reach the more northern look directions of Skibotn; or it was propagating along the oval, in a eastward or westward direction. The former is not likely as the auroral oval was to the north, implying the wave had propagated along the auroral oval.

Figure 4.10 shows the LSP for the 6300Å intensities for the same data of the 25th November 2003, but showing the amplitudes of the waves rather than the spectral power. This shows the relative strengths of the waves in relation to the background intensities, so waves between different look direction and instruments, with different time resolutions, can be directly compared. These are plotted against frequency in Hertz, rather than period as the wave periods of interest are more evenly spaced in frequency than period. For both the 1.8-hour (9.7×10^{-4} Hz and 3.7-hour (4.6×10^{-4} Hz) waves, KEOPS shows larger amplitudes in Figure 4.10 to the north, which decrease towards the south, through A, east and west, zenith, and the smallest amplitudes are seen at the south and B positions. This again is consistent with an auroral oval source region, which is in the north throughout the night. The relative sizes of the amplitudes between each of the look directions at Sodankylä are similar to those at KEOPS, which is at a similar latitude, and Sodankylä also has greater amplitudes to the north and A positions than the other look directions. The FPI intensities are not calibrated at present, so the intensity is dependant on detector sensitivity, so the intensities, and hence the amplitudes, are an order of magnitude smaller at Sodankylä than KEOPS.

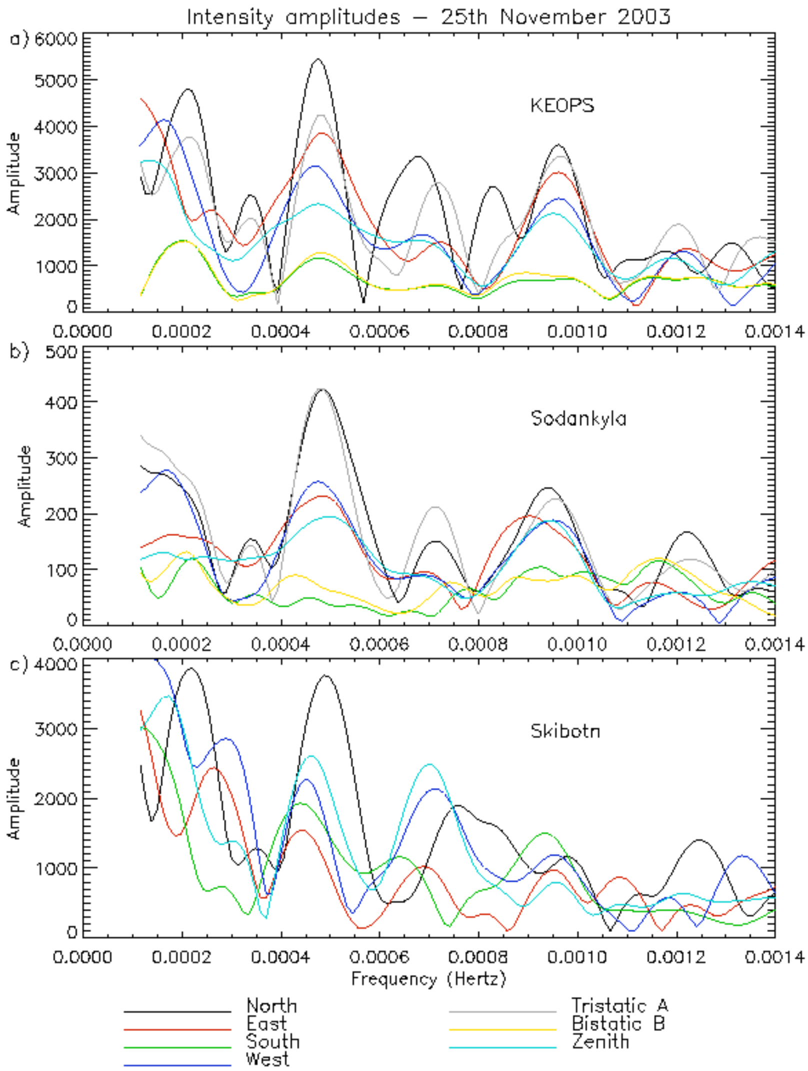

Figure 4.10 *Amplitudes of Lomb-Scargle periodograms for FPI intensities for a) KEOPS, b) Sodankylä, and c) Skibotn, on the 25th November 2003.*

The amplitude of the 3.7-hour wave is 12% of the maximum intensity at KEOPS and 11% for Sodankylä, so the wave amplitudes are comparable at both locations, as would be expected for sites at similar latitudes. The amplitudes in Figure 4.10 of the 3.7-hour wave in the south and B positions are very small, in agreement with the low power in the periodogram. Skibotn sees larger amplitudes in all look directions for the 3.7-hour wave. The north direction again has the largest amplitude, but this is consistent with the keogram from Ny Ålesund on Svalbard in the polar cap (not shown), which shows activity, confirming that the oval spreads across this whole region observed. The auroral oval covers all of the look directions for Skibotn. The 2.5-hour wave is more prominent than the 1.8-hour wave.

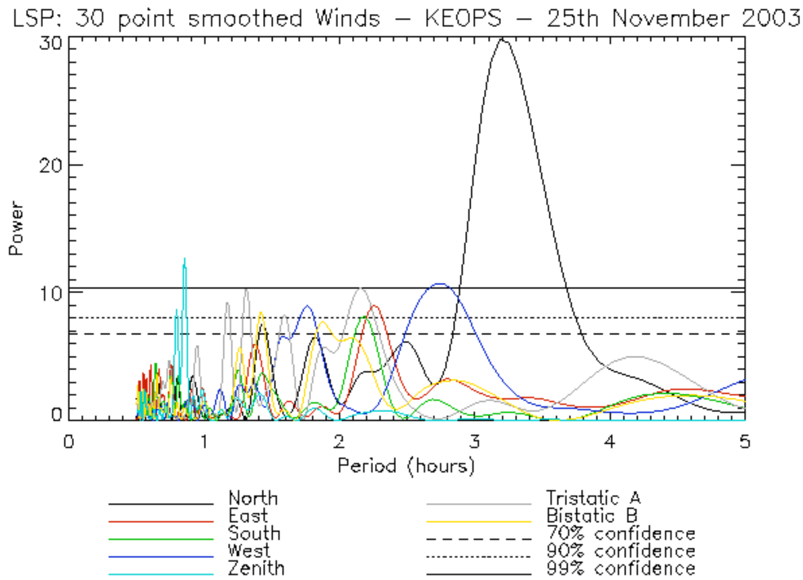

Figure 4.11 *Periodogram from KEOPS wind data with a 30-point (~120 minute) smoothed value removed, for the 25th November 2003.*

To view the periodicities in the wind data from the FPIs, the data first need to be smoothed. This is because the two-cell pattern of thermospheric winds due to magnetospheric convection across the polar cap, when there is some geomagnetic activity, creates a predominantly westward wind in the evening sector and eastward in the morning hours. This means that the majority of the power of the periodogram goes into this 24-hour period wave-like pattern. To compensate for this, a 30-point running smoothing of the data is subtracted from the data. This corresponds to an approximately 120 minute smoothing. This value is used as it removes the larger trends, and so removes power from the longest periods, those not associated with gravity waves, but is not so small as to remove power from the periods of interest. These data, normalised for diurnal variations, for KEOPS on the 25th November 2003 are shown in Figure 4.11. This shows a strong wave with a period of 3.1 hours in the north winds, but does not show the 1.8 and 3.7-hour waves seen in the intensities. Many other periods above the 70% confidence level are seen though, for instance around 1.5 and 2.3 hours. Gravity waves in the winds are complicated by the fact that the wave equations are dependant on the background winds (see section 1.4.2, Nappo (2002), or Lindzen (1990)). The frequencies of waves that are present in the intensities are therefore not necessarily expected to be present in the wind speeds. Innis and Conde, (2001) observed gravity waves in vertical winds from the DE2 satellite. However, the FPI data presented here is believed to be the first observation of gravity waves detected in the horizontal upper thermospheric winds.

Smoothing the data is also useful for identifying waves in the thermospheric temperatures. Gradients are often seen over a night, as the temperatures are relatively slow to respond to changes in geomagnetic activity levels. Figure 4.12 shows the results for the temperatures from the same night of the tristatic campaign, from KEOPS. The top plot is the data smoothed with a 30-point running average, subtracted from the absolute temperatures. A wave structure can be clearly seen in the temperatures, and in all the look directions. The waves end at 04UT, which is just due to the sky becoming overcast at this time. The periodogram is shown in the lower plot. The periodicities in the different look directions are not quite as well defined as for the intensities and winds, but are within the minimum error, of 4.5 minutes between data points. The most predominant periods are at 1.4, 1.8 and 2.3 hours, with other periods above the 70% confidence level at 1.0, 1.1 and 1.3 hours. The response time of the three parameter are different, but it is not clear whether the same periodicities would be expected in the winds as intensities, and this will be investigated further by a statistical survey of the FPI data sets (Chapter 5).

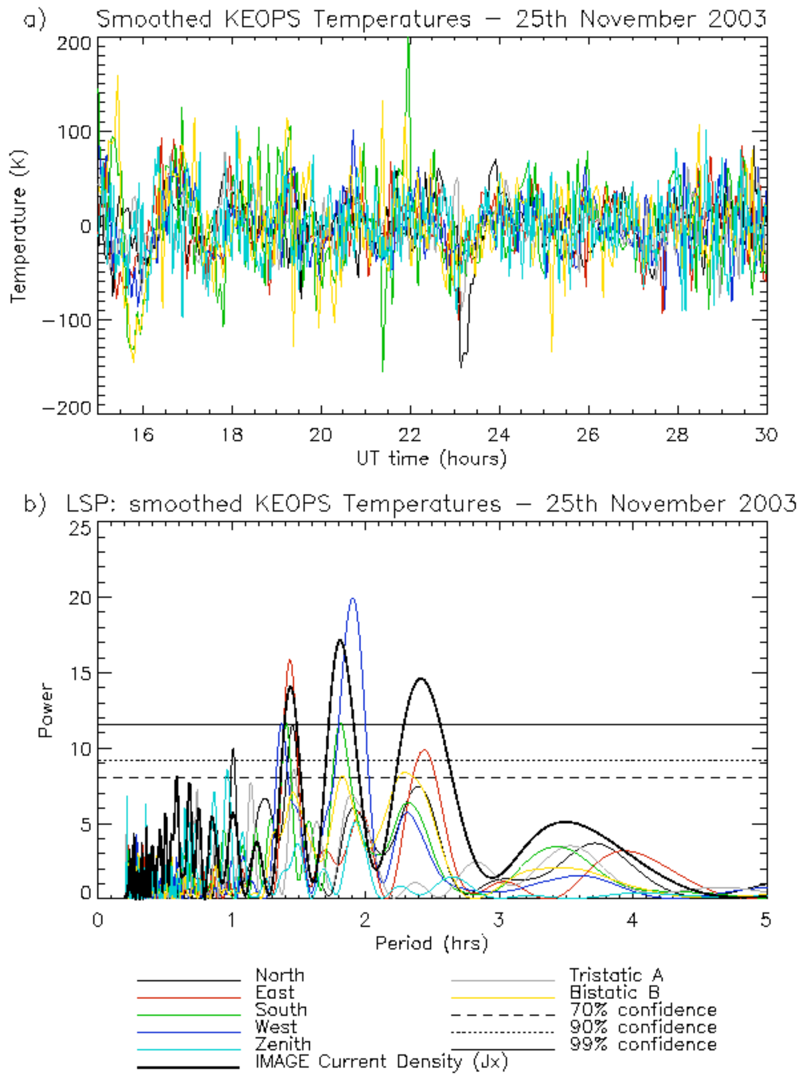

Figure 4.12 *Neutral temperatures with a 2-hour smoothed value removed (a) and their periodogram (b) from KEOPS from the 25th November 2003. The bottom plot also shows the periodogram for the equivalent current densities (thick black line) from IMAGE data (smoothed and reduced by a factor of 2 for comparisons).*

Williams et al., (1993) studied data taken as part of the World-Wide Atmospheric Gravity Wave Study (WAGS) campaign (section 1.4.4). They show that two bursts of particle precipitation could cause gravity waves with the same period as the time between the precipitation periods. The two times of strongest particle precipitation on the night of the 25th November 2003 in Figure 4.6 were at 19:27UT and 23:15UT. These times are 3.7 hours apart, and correspond to one of the strongest periods seen the periodogram for that data. There is a problem though with this theory, which can be seen from the smoothed temperatures in Figure 4.12a, which shows that the waves start before the start of the data set, i.e. before the first period of precipitation seen in the intensities at 19:27UT.

However, APL has a colour All-Sky Camera (ASC) at the site of the KEOPS FPI. Data is available for this night from as soon as the Sun sets. The FPIs require lower background light levels and so data is only collected from an hour after sunset. An image from the ASC 3.7 hours before the first period of precipitation seen in the FPI intensities is shown in Figure 4.13, at 15:49UT on the 25th November 2003. This figure also shows the positions of the FPI viewing volumes. Green lines show the look directions (north (N), east (E), south (S), west (W), zenith (Z), and tristatic (A) and bistatic (B) positions with Sodankylä. The red dots show the 6300Å emitting volume observed presuming a 240km emission height and circles show the distance of this region assuming different emission altitudes. This image shows an auroral arc to the north of KEOPS, which again agrees with the amplitudes of the waves being larger to the north (Figure 4.10). The image showing auroral precipitation in the ASC (Figure 4.13) is 3.68 hours before the first period of precipitation seen in the FPI data (at 19:27UT). The ASC images also show a third period of precipitation this night, at 23:09UT – 3.70 hours after the second. The ASC data images are only taken every 5 minutes, so these periods are the same as the FPI periodograms within the uncertainty of the timing. This therefore, for the 3.7-hour wave, corroborates the theory of Williams et al., (1993) that the wave is created from forcing from evenly spaced bursts of heating due to particle precipitation.

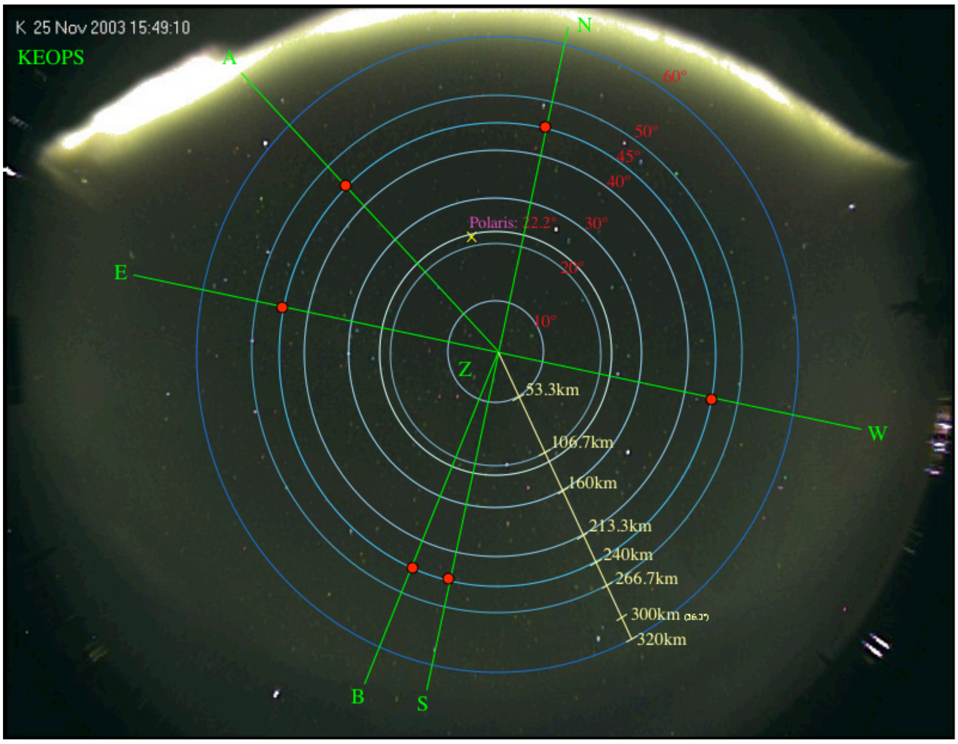

Figure 4.13 *All-Sky Camera image from KEOPS on the 25th November 2003, showing an auroral arc to the north of KEOPS. Green lines show the look directions; red dots show the 6300Å emitting volume presuming a 240km height; circles show other height distances.*

There are no periods of enhanced precipitation in the intensities (Figure 4.6), for example, 1.2, 1.8 or 2.5 hours apart, to produce these other peaks in the periodograms. Therefore, heating from precipitation does not seem to be a likely cause of the gravity waves on this night, other than for the 3.7-hour period wave.

A cross correlation was performed on the smoothed neutral temperatures to obtain the phase speed and direction and horizontal wavelength (see equation [38]). The peaks of the curves obtained give the zonal and meridional lags, τ_x and τ_y respectively (Oliver et al., 1994). To perform the cross correlation, the data from the two look directions need to be at the same times, therefore the data were interpolated, to 0.1-hour intervals. This period was chosen as it is just longer than the time resolution of the data. The first two hours of observations of the temperatures, when smoothed

(Figure 4.12a), show just the 1.8-hour period wave, so this period (14:24UT – 17:30UT) is used to determine the wave parameters. The results of this are shown in Figure 4.14. This shows a lag for the east – west correlation (dotted line) of approximately 0.2 hours, but the north – south correlation (solid line) shows a dip in the peak at 0.3 hours, so that the lag cannot be determined without further investigation, as follows.

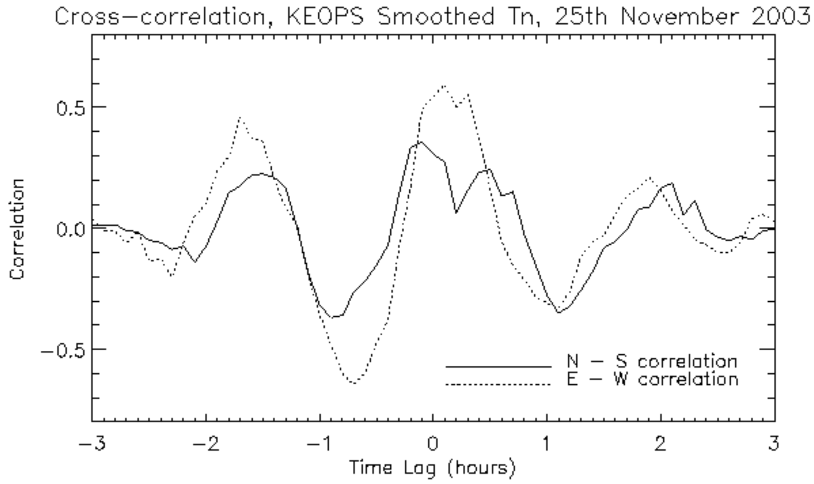

Figure 4.14 *Cross correlations on smoothed temperatures from KEOPS on the 25th November 2003: north and south (solid line), east and west (dotted line).*

Figure 4.15 shows the temperatures after they have been smoothed again with a 40-minute running smoothing, to decrease the noise levels further so that the peaks can be better determined. By examination of these smoothed temperatures, it can be seen that the north direction shows a skew between 15UT and 16UT, away from a sinusoidal curve. This skew is likely to be due to the superposition of another wave on top of the 1.8-hour wave that can be clearly seen. Subtraction of a sinusoidal wave with a period of 3.7 hours and amplitude of the standard deviation of the data removes this skew for the first oscillation though distorts the second and third oscillations. Removal of this skew decreases the size of the dip in Figure 4.14, though as a simple removal of a wave changes the successive oscillations, this does not remove it completely. This will also be the cause of the shift of the peak at 16:30UT as the beginning part of the peak is removed as it is decreased by the secondary wave. This will also account for the smaller amplitude of this wave at this time.

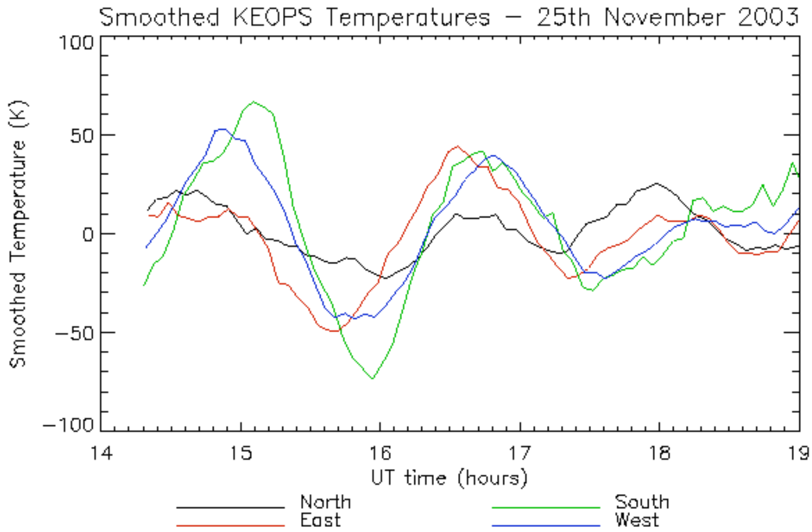

Figure 4.15 *Neutral temperatures with a 2-hour smoothed value removed, then with a secondary 40 minute smoothing, from KEOPS on the 25th November 2003.*

Therefore, to determine the lags, the peaks in Figure 4.15 between 14UT and 15UT are used. This gives values of 0.5 hours for τ_x and 0.2 for τ_y. This gives a difference in the lags of 0.3 hours, which is approximately the lag seen on the secondary peaks in Figure 4.14. Putting these lags into equation [38], and using a distance between the look directions of 480km, gives a phase speed of 250±50ms^{-1} and a phase angle ϕ of 302°±15°, measured as the direction the wave is travelling, at an angle anticlockwise from the east. It is therefore propagating in a south-south-eastward direction, consistent with an auroral oval source as this is to the north at this time. For the wave with period 1.8 hours, this gives a horizontal wavelength, λ_h of 1600km. This is consistent with previous observations and theory of TIDs, discussed in section 1.4.4, or for example Hocke & Schlegel, (1996).

Lilensten & Amblard (2001) suggest that eddies could be the cause of gravity waves in neutral winds (see section 1.4.4). If this is the case, they expect the oscillations have a $\pi/2$ phase difference between the meridional and zonal components of the neutral wind. This is not seen in these FPI data on this night, though the effect could be hidden due to larger effects of the background winds. Lilensten & Amblard (2001) say that the eddies occur at practically any time and this is a normal feature of the wind propagation. Either this is not the case and the eddies just were not present on this occasion, or they are too small scale a feature to be observed with these data. The second of these possibilities can be determined by looking at further data sets.

The fact that there is a lag between the different look directions is also evidence that the oscillations are due to gravity waves, rather than being, for instance, an atmospheric tide that has been modulated by a planetary wave. Tides and planetary waves are large-scale phenomena with very large horizontal wavelengths of several thousand kilometres. These would therefore not vary over the scale of the distance between FPI look directions, which for one site is around 480km.

4.4.2 ELECTROJETS

To investigate whether Joule heating from the electrojets could be the generation mechanism of the other period gravity waves, IMAGE magnetometer data were analysed. Electrojet activity does not necessarily coincide with the location of particle precipitation, so this would give a different source and provide different periodicities to a precipitation source. There are many magnetometer stations across Scandinavia, including at the sites of the FPIs and EISCAT radars (see Figure 2.13 in section 2.4.3). Due to the large array of magnetometer stations, equivalent current densities can be calculated (Pulkkinen et al., 2003). This gives values, for J_x (north component) and J_y (east component), in A/km. Values are calculated along a line of longitude through the centre of the magnetometer stations, and at an altitude of 100km. The current densities will be high along the electrojets, so are a good indication of their location. LSPs are therefore performed on data near the position of Kiruna (at 68.62°N, 22.06°E).

The thick black line in Figure 4.12b is the periodogram for the J_x component of the equivalent current densities. Data from 14-24UT on the 25th November 2003 were used. To again increase powers in the shorter periods, 80 minute smoothed data was subtracted from the data before the periodogram was run. The LSP for the equivalent current densities, divided by a factor of 2, is plotted over the periodograms for the normalised temperatures, for comparisons. A very good match can be clearly seen with many of the periodicities detected. The strong periods at 1.4, 1.8, and 2.4 hours are seen in the LSPs of the equivalent current densities as well as the temperatures. Even some of the shorter periods are seen, such as those at 0.8, 1.0 and 3.5 hours. The 3.5-hour peak has a lower power for the temperatures than the 6300Å intensities due to the smoothing, which takes power out of the longer periods but allows the shorter periods to be seen more clearly.

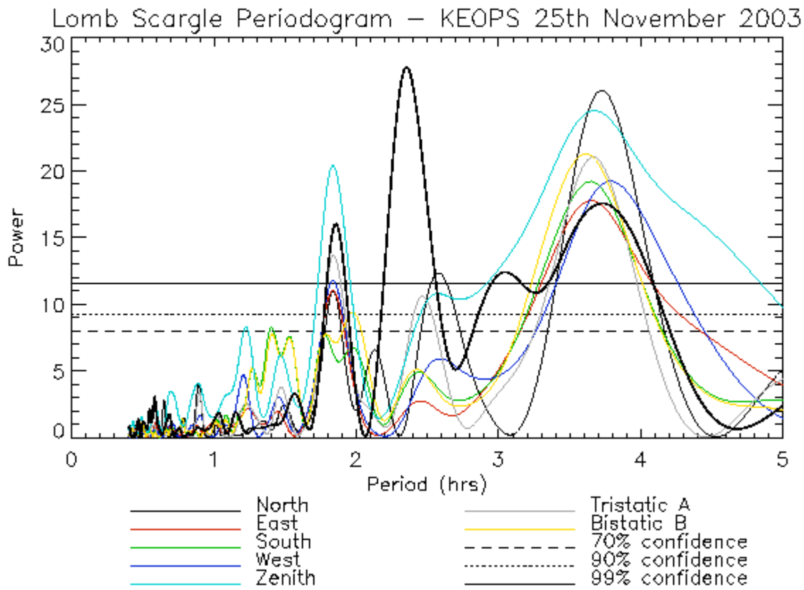

Figure 4.16 *The LSP for FPI intensities from KEOPS from the 25th November 2003. The periodogram for the equivalent current densities from IMAGE data (thick black line) has a 200-minute smoothing removed, and the periodogram was divided by 4 for comparisons.*

Figure 4.16 shows the periodogram for the KEOPS FPI intensities, with the smoothed current density overlaid with a thick line as for the temperatures. The J_x component from 12UT – 06UT is used, as it is the north/south component and will show the locations of the electrojets, with a 200-minute smoothed value removed, and scaled down by a factor of four, for comparisons. The LSP for the equivalent current densities has a larger power because it has approximately four times higher time resolution, with data points every minute. The 1.8 and 3.7 hour peaks match very well, but there is also a large peak at 2.4 hours that is not seen in the intensities. The temperature data would be expected to match well with the current densities (Figure 4.12) as currents cause heating. However, the intensities may not correlate so well with the current densities (Figure 4.16), as precipitation causes heating, but not vice-versa, and so the currents would only be proportional to the electron densities and hence 6300Å intensities if the currents were formed through precipitation, rather than for instance changes in conductivities. The 3.7 hour wave has a higher power here than in Figure 4.12b, as the intensities are not smoothed, since there are no long term trends in the intensities, as are seen in the winds and temperature data over a night.

By comparing the vertical perturbations (B_z) from magnetometers at different latitudes, the position of the centre of the electrojet can be determined as a function of local time, i.e. its variation in latitude. At the location of the westward electrojet, the B_z component to the north will be decreased, and increased to the south. Conversely, for the eastward electrojet, B_z will increase to the north and decrease to the south. Figure 4.17a shows the latitudinal location of the centre of the electrojet through the night, estimated at 10-minute intervals.

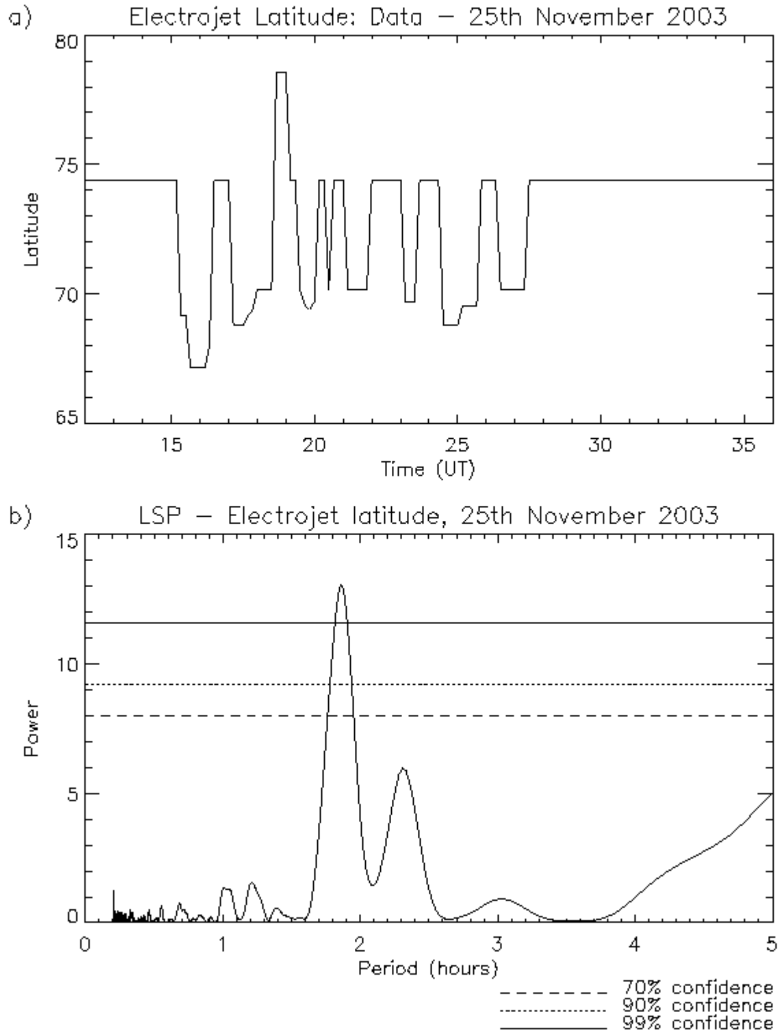

Figure 4.17 *Positions of the electrojets (a) and periodogram (b) from IMAGE B_z data, from the 25^{th} November 2003.*

A Lomb-Scargle analysis of the electrojet position gives the periodogram in Figure 4.17b. This shows periods of 1.8 and 2.3 hours, which correspond to the smoothed temperature periods from KEOPS. Therefore, the location of the electrojet correlates with the FPI data as well as the intensity of the current density. At any one location, the current density will vary as the electrojet moves northward and southward over the location, so these properties are connected.

4.4.3 ION-NEUTRAL COUPLING

Due to the relative difficulty in obtaining thermospheric data with respect to ionospheric data, thermospheric gravity waves have often been studied using neutral data derived from ionospheric measurements, such as radar data. As these measurements are not independent, the coupling between the ions and neutrals cannot properly be studied. However, for the night of the 25th November 2003 the EISCAT UHF radars were in tristatic mode, viewing the same volume of sky as the FPIs at the tristatic point above KEOPS. This provides a good opportunity to study gravity waves and their ionospheric counterpart, TIDs, at the same time. CUTLASS and STARE coherent scatter radar data are also available for this period.

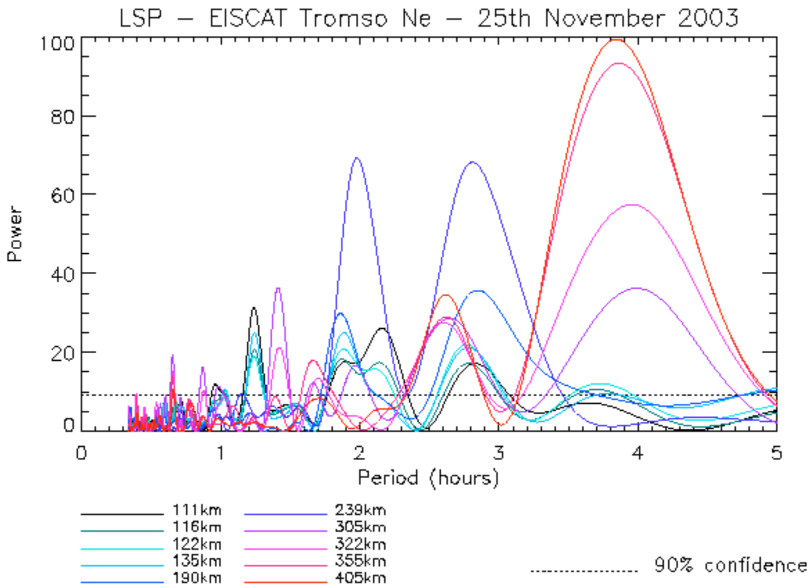

Figure 4.18 *Lomb-Scargle Periodogram from EISCAT electron densities from the 25th November 2003. Colours are for different heights along the Tromsø beam, blues are low altitudes, red high altitudes.*

The EISCAT UHF data from the 25th November 2003 is shown in Figure 4.5. Several wavefronts can be seen, most clearly in the electron densities (top plot), where one wavefront is highlighted with a dashed line (between 21:40UT and 22:20UT). The other parameters also show waves to some extent. Lomb-Scargle analyses of the electron densities (N_e) for different heights along the Tromsø beam are shown in Figure 4.18. The altitudes shown range from the E-region, in greens and blues up to high F-region altitudes, in reds. 111km (dark green) is the nearest range gate to the equivalent current densities of the magnetometer data. The highest altitude shown is at 405km in red.

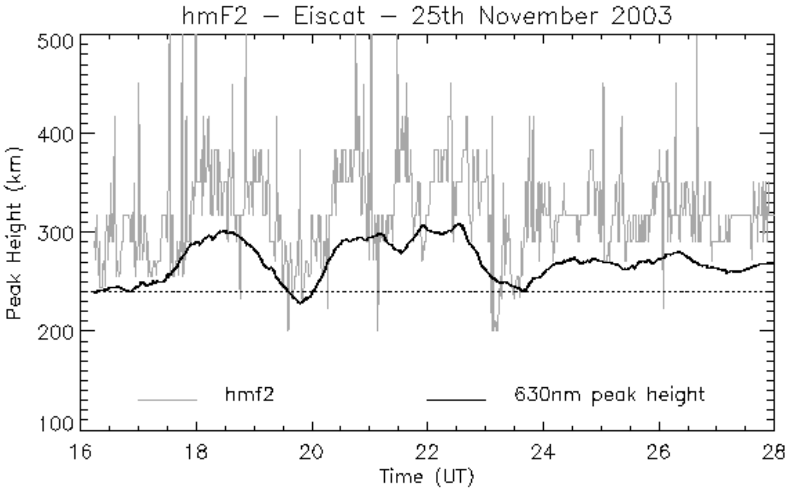

Figure 4.19 *EISCAT hmF2 values for the 25th November 2003 (grey) and estimated OI emission altitude (black). The presumed altitude of 240km is shown with a horizontal dotted line.*

The height of the maximum electron density in the F-region (hmF2) is shown in Figure 4.19 in grey. The 6300Å emission is estimated as being approximately one scale height (50km) below hmF2 (an hourly smoothed value is plotted). The emission altitude is approximately one scale height thick, and the peak can be seen to vary between 230km – 300km over this night. The LSPs corresponding to these altitudes are shown in the EISCAT analysis (Figure 4.18) in blue and purple (239km and 305km).

The periods seen in the EISCAT data (Figure 4.18) match several of those seen in the FPI data. The first periodicity seen in N_e data above the 90% confidence level, which is at 0.6 hours, however is not present in FPI data. The period with a strong power at 1.2 hours is only seen in the lower altitudes, from 111km – 135km. This period is weakly seen in the current densities, in Figure 4.12, but not strongly in the FPI data. This could imply that it has dissipated by the time it reaches the 6300Å emission altitude.

The 1.5-hour wave that is seen in the FPI temperatures and the electrojet, from the equivalent current densities, is also seen in the upper altitudes of EISCAT N_e data. This may possibly indicate that there is a vertical component in the direction of propagation of this wave. The STARE coherent scatter radar (Greenwald et al., 1978) provides electric field values, which are also at the electrojet altitude. The E_x values (geomagnetic north component of the electric field, in mV/m) are shown in Figure 4.20a. The horizontal lines mark the latitudes of KEOPS (full line) and the tristatic A point and the Skibotn north location, which is the northernmost look direction of the three FPIs. The LSP for the E_x values is shown in Figure 4.20b. This shows the periodogram powers in colour, for all the latitudes in the STARE radars field of view, at a longitude of 20.4°E, which is the location of the KEOPS FPI. Vertical yellow lines mark waves with periods of 1.5, 1.8, 2.3 and 3.3 hours.

This therefore shows that the 1.5 hour wave is likely to be of electrojet origin as it is seen in the electrojet equivalent current densities and the electric fields, both at E-region heights. The 1.5-hour wave is only seen in the FPI data in the intensities in the south and B directions, in KEOPS and Sodankylä (Figure 4.8). As it is seen in the STARE data (Figure 4.20) only above KEOPS, at the south of the STARE field of view, this implies it has a mostly southward propagation direction. That the wave is only seen in the EISCAT N_e values above 300km also corroborates this, as the EISCAT beam was directed to be at 240km directly over the KEOPS site. The electrojet altitude of 110km would be at shorter range-gates and therefore less far from the transmitter at Tromsø, i.e. north of KEOPS. Conversely, the 300km range gate and above would be further from the transmitter at Tromsø and so further to the south of KEOPS. The wave is only seen in the high altitudes, implying the wave was travelling southward.

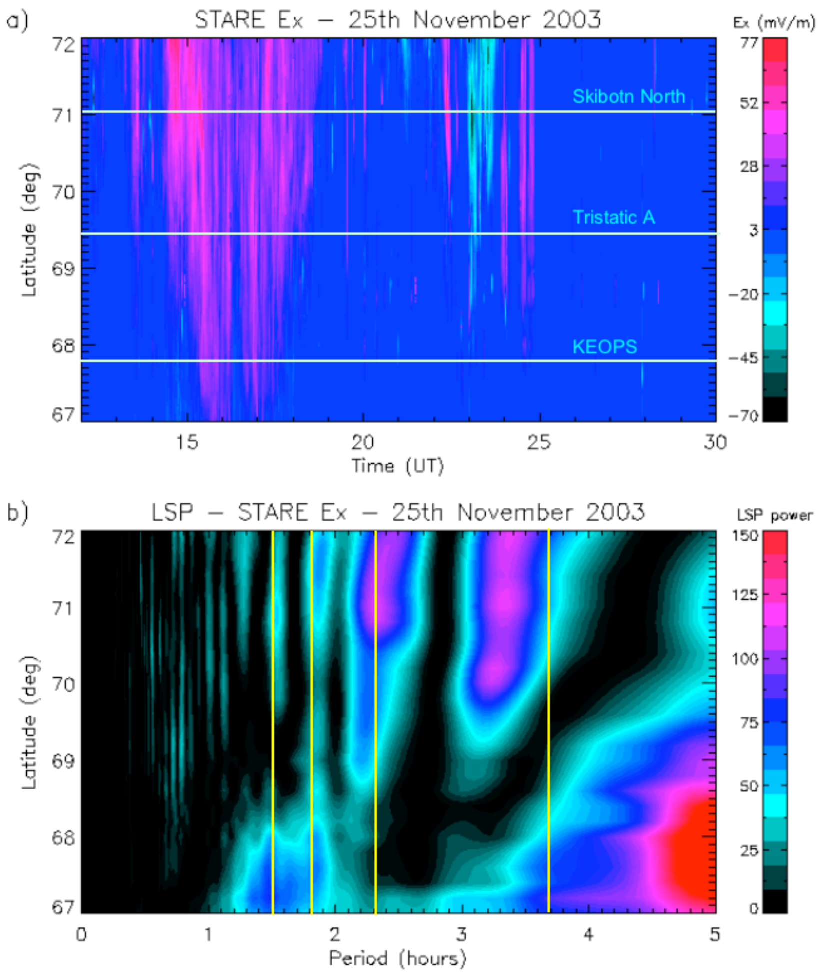

Figure 4.20 *STARE electric field (northwards component) for the 25th November 2003 (top) and LSP (bottom), at a longitude of 18.0°E. Horizontal lines (top plot) show the latitudes of KEOPS, the tristatic A point, and the northernmost extent of the FPI data. Vertical lines in the LSP highlight some of the waves present.*

The 1.8-hour wave, seen in both the FPI intensities and temperatures, is also seen in the N_e LSP (Figure 4.18) in the E and lower F region heights (up to about 290km). This 1.8-hour peak is not seen in the highest altitudes, and not above about 300km. This period is also present in the electrojets (Figure 4.17), equivalent current densities (Figure 4.16), and electric fields (Figure 4.20b) all of which are at E-region altitudes. This therefore agrees with an electrojet source of the wave, which has dissipated by the time it reaches higher F-region altitudes, higher than the FPI observing height.

The fact that the 1.2-hour wave seen in the N_e LSP is not observed in the FPI data, but the 1.8 hour wave is, which is also of an electrojet source, could imply that they have different limits to their propagation distances. The amplitudes of the two waves (not shown) are very similar. The STARE data (Figure 4.20b) shows the 1.2-hour wave in the lowest and highest latitudes, but not very strongly at the FPI data latitudes (shown on Figure 4.20a).

There are two overlapping periodicities in Figure 4.18 from 2.5 to 3 hours. The higher altitudes, from 305km peak at 2.6 hours, but below this, the peak moves to nearer 2.9 hours. This discrepancy cannot be explained by uncertainty in the data, as it is a larger gap than the errors in the LSP (0.1 hours). The equivalent current density LSP has a peak at 2.4 hours, and the STARE electric fields have peaks at 2.3 hours at high latitudes, from the tristatic A point and further north (Figure 4.20), and the position of the electrojets (Figure 4.17b) shows a periodicity of 2.3 hours. Both the FPI intensities and temperatures have periodicities that differ in the different look directions, ranging from 2.3 to 2.6 hours. The tristatic A point temperatures has a small peak at 2.8 hours, but this is below the 70% confidence level. Whether these are superimposed waves or one wave with a varying source is unclear.

The 3.7-hour wave only occurs at the higher altitudes in the EISCAT N_e data (Figure 4.18), from 300-500km. As the 3.7-hour wave is not seen in the lower altitudes, this shows that it is unlikely to be a wave of electrojet source, as this is at lower heights and the wave would be expected to be stronger here. This periodicity is not seen in the magnetometer or STARE electric field data at E-region heights. The magnetometer data will show the height integrated currents, and so will be dominated by the E-region. This therefore implies an auroral precipitation source region, at high altitudes, and therefore from lower energy particles. This agrees with ASC data shown in section 4.4.1.

4.4.4 SUMMARY OF RESULTS

Gravity waves detected in the high latitude thermosphere in neutral winds and temperatures as well as emission intensities. Previous work on observations of thermospheric gravity waves (e.g. de Deuge et al., (1994) and Innis et al., (2001)) used photometer data, so only had 6300Å intensities. Innis and Conde, (2001) observed gravity waves in vertical thermospheric winds from satellite (DE2) data. The data presented here is therefore believed to be the first detection of gravity waves in upper thermospheric temperatures. Their work was also in the Southern polar cap region, whereas these data are from northern latitudes and from the auroral oval region. The Scandinavian region also has the advantage of being covered by a range of other instrumentation.

Gravity waves of a variety of periods have been detected using a Lomb-Scargle analysis, from tens of minutes to a few hours. The gravity waves were seen in thermospheric intensities, winds and temperatures from Fabry-Perot interferometer measurements of the 6300Å atomic oxygen line. The amplitudes of the waves show that the data are consistent with an auroral source region.

Two mechanisms have been proposed (de Deuge et al., (1994)) for the formation of thermospheric gravity waves – Joule heating from particle precipitation and heating from electrojet currents. A case study has been presented here, for the 25th November 2003 from the Scandinavian sector, when waves of several periods have been found.

There is evidence that these waves have different source mechanisms. It has been shown that for this data set that heating from particle precipitation is the likely source mechanism for a wave of period 3.7 hours. Two periods of precipitation are seen in the FPI intensities, also separated by 3.7 hours. Although the first of these periods of precipitation occur after waves are seen in the temperatures, an auroral arc is seen in all sky camera images 3.7 hours before this. This therefore suggests that these periods of precipitation are driving the wave, as proposed by Williams et al., (1993). The amplitudes of the waves are greater to the north of the region, which is where the auroral activity is observed. This too supports an auroral source region.

For the other periodicities seen in these data, of 1.2, 1.5, 1.8 and 2.5 hours, it is proposed that the source of the gravity waves is Joule heating from electrojet activity. The Lomb-Scargle analysis performed on equivalent current densities, as well as the position of the electrojet, calculated from magnetometer data from the IMAGE chain, reveal periodicities similar to those found in the FPI data. Periodicities seen in the FPI temperature data are very well matched by the periods found from the J_x component of the equivalent current densities, showing that the strength of the electrojet is the cause of the gravity waves observed. Cross correlations of the neutral temperatures were used to find that the phase speed of the 1.8 hour wave to be 250ms^{-1} with a phase angle □ of 300°, which gives a horizontal wavelength, λ_h of 1600km. This is both consistent with previous TID measurements and with an auroral oval source region. The phase speeds and angles could not be determined for the other period waves as they could not be easily distinguished in the data sets.

The gravity waves found in the FPI thermospheric data have also been compared with the TIDs found in radar data. EISCAT incoherent scatter data provides electron densities and ionospheric wind speeds and temperatures. STARE coherent scatter radars provide electric field data, which can be used to identify gravity waves at E-region altitudes. These data sets confirm the source of the waves with periodicities of 1.2, 1.5 and 1.8 hours to be from Joule heating from electrojet currents, and the 3.7-hour wave to be of higher altitude in origin, and therefore from particle precipitation localised heating. These complementary data sets therefore show that on the 25th November 2003 waves of several different periodicities were present, and that different source mechanisms produced the different waves.

4.5 HIGH-RESOLUTION KEOPS DATA

In April 2004 the KEOPS measurement cycle was set to take constant exposures in the tristatic A position, with the rest of the look directions only observed approximately every 16 minutes. At this time, the Andor detector was installed on the KEOPS FPI. As this is a back-illuminated electron-multiplying CCD (see section 2.2.4), the Andor detectors have very high quantum efficiencies. This allowed short integration times of 10 seconds. This, combined with the fact that the mirror does not need to be moved if only one direction is observed, allowed a time resolution of the data of 15 seconds. The disadvantage of this is that the context from the other look directions is lost, which is why a full cycle was taken at intervals, including a calibration lamp exposure. This high time resolution is useful for the study of gravity waves as all periods down to the Brunt-Väisälä period should be observable. This is taken to be approximately 12 minutes at the 6300Å emission height. If gravity waves are not seen down to this period, then this is likely to be due to the limits of the variability of the thermosphere.

There were four clear nights when data were collected in this way: the 3rd, 5th, 10th, and 11th April 2004. The all sky camera keogram from Muonio, which is between the latitude of the KEOPS site and the tristatic A position, is shown in Figure 4.21, for the 3rd April 2004. Aurora can be seen throughout the night, which shows that the sky was clear. Peak intensities show that the auroral oval was overhead of Muonio for most of the night, but northwards between 20 and 21UT, and overhead to southwards from 22UT. The keogram also shows that the Moon passed over the zenith of the camera at 20UT.

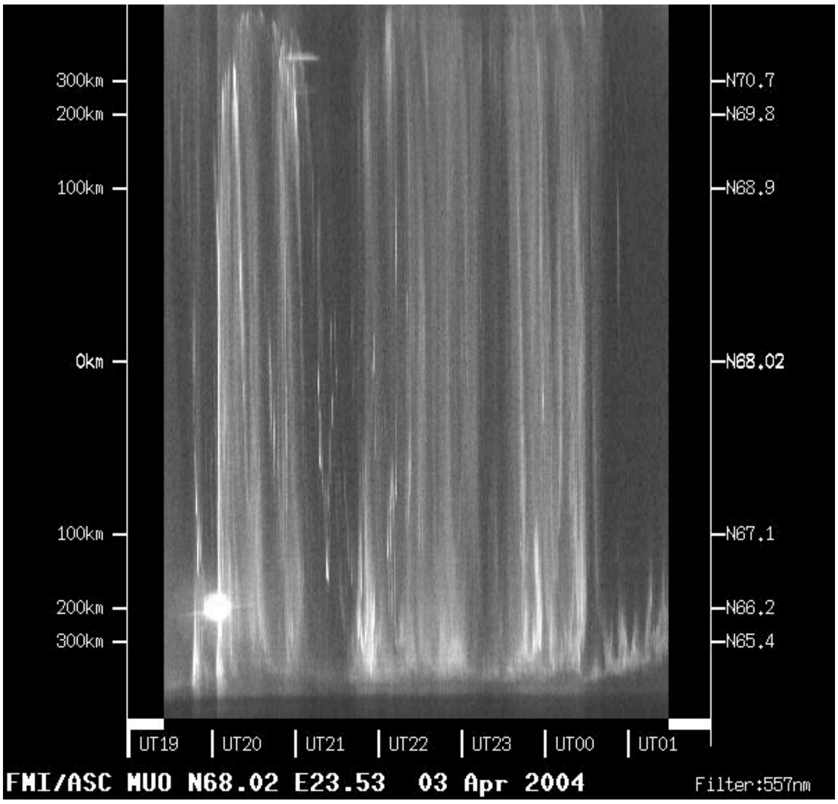

Figure 4.21 *ASC keogram from Muonio on the 3rd April 2004.*

Results of the Lomb-Scargle analysis of the data from these nights are shown below. The look directions other than the tristatic A data have poor time resolution, and the short length of the night at this time of year gives a short data set. Therefore, waves cannot be found in these with enough confidence to calculate phases and lags, and so calculate the wave propagation direction and wavelength. The other issue with the poor resolution of the cardinal look directions is that they produce time series with a small number of data points, which is what is used to calculate the number of independent frequencies sampled. As the program is run on each look direction in a loop, this does not give a sufficient sampling frequency for the tristatic A data. Therefore, for these nights, the sampling frequency used is calculated as twice the Nyquist frequency of the tristatic A data, which for 10-second resolution data is 0.05Hz.

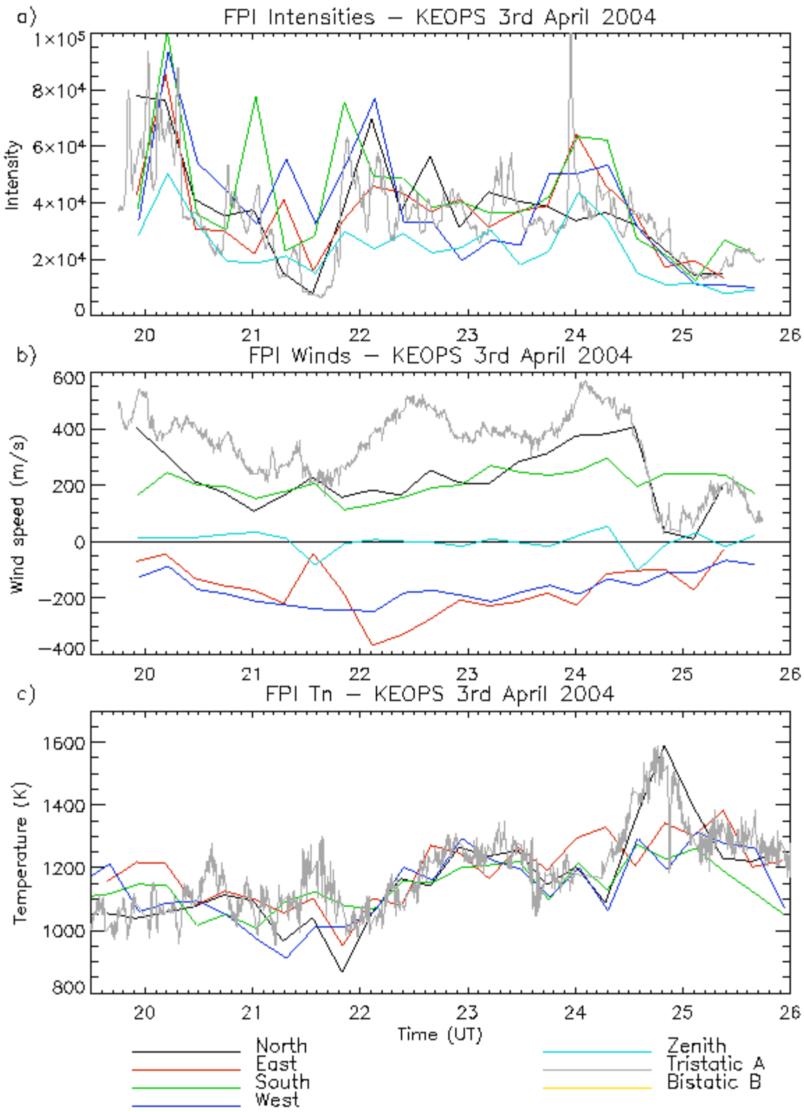

Figure 4.22 *FPI data from KEOPS on the 3rd April 2004 (a) intensities, (b) winds, and (c) temperatures.*

The FPI data for the 3rd April 2004 are shown in Figure 4.22. The high time resolution of the tristatic A data clearly show much greater variability than the other look directions. This is especially true for the intensities in Figure 4.22a, which vary on shorter scales than the winds and temperatures. The tristatic A winds in Figure 4.22b follow the general pattern of the north winds, which are not far from the A position, but show much more structure. All the temperatures match reasonably well on a large scale when the changes are slow, for instance between 22 and 24UT in Figure 4.22c. However, when there are jumps in the temperatures, such as in the intervals 20 – 22UT and 24 – 25UT, the low resolutions of the other look directions do not see much of the variability of the tristatic A data.

To see the variability of this high-resolution data more clearly, Figure 4.23 zooms in on an hour section of the data from 21 – 22UT on Figure 4.22. Only the tristatic A and north look directions are shown for clarity. The overlaid purple line is an interpolation of the data at 0.1-hour (6 minutes) resolution. This is the approximate resolution of the data when a normal cycle of measurements is made, and any one look direction is measured just once in a cycle of seven measurements and a calibration lamp exposure. This shows the information that is lost on reduction of the time resolution. For example, the peaks in the intensities in Figure 4.23a, at 21.05UT and 21.15UT, are completely missed out in the interpolated data. The three reversals in the wind direction around 21.2UT in Figure 4.23b are also not seen, nor is the fast variability from 21.6UT to 21.7UT. This is also true of the temperatures in Figure 4.23c through much of this period.

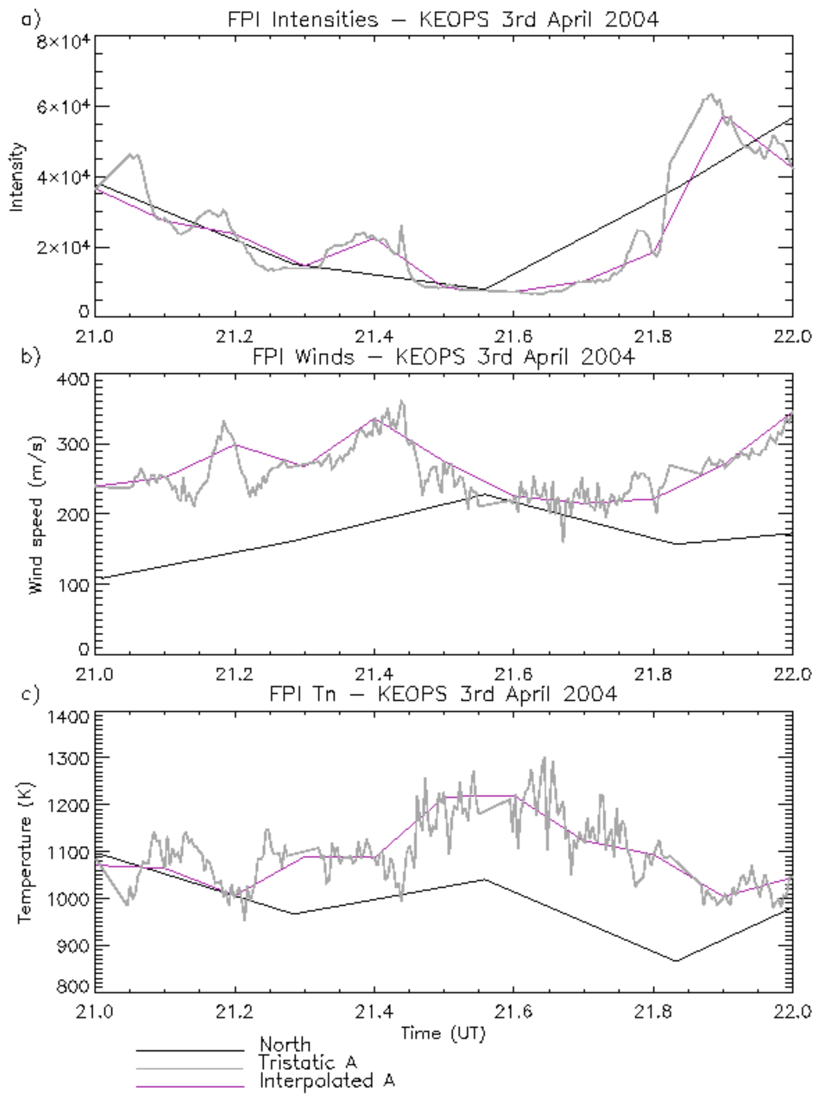

Figure 4.23 *1 hour of high resolution FPI data from KEOPS on the 3rd April 2004, including 0.1hr (6min) interpolated tristatic A data: (a) intensities, (b) winds, and (c) temperatures.*

To show that these deviations from the interpolated data are real and not just noise, Figure 4.24 shows the data zoomed in once more to a 6-minute period (0.1 hours). For this plot, the north data has been removed, and the individual data points have been marked with crosses for further clarity. All three plots show structure in the high-resolution data not seen in the interpolated data. However, there are more data points than are needed here as the lines produce relatively smooth curves, so most of the structure in these data could be seen with data approximately every 0.02 hours, or data with 1 minute resolution. This implies a limit to the scale on which the thermosphere can vary, of approximately 1 minute. This is much shorter than the scale of 1 hour on which the thermosphere is usually assumed to vary.

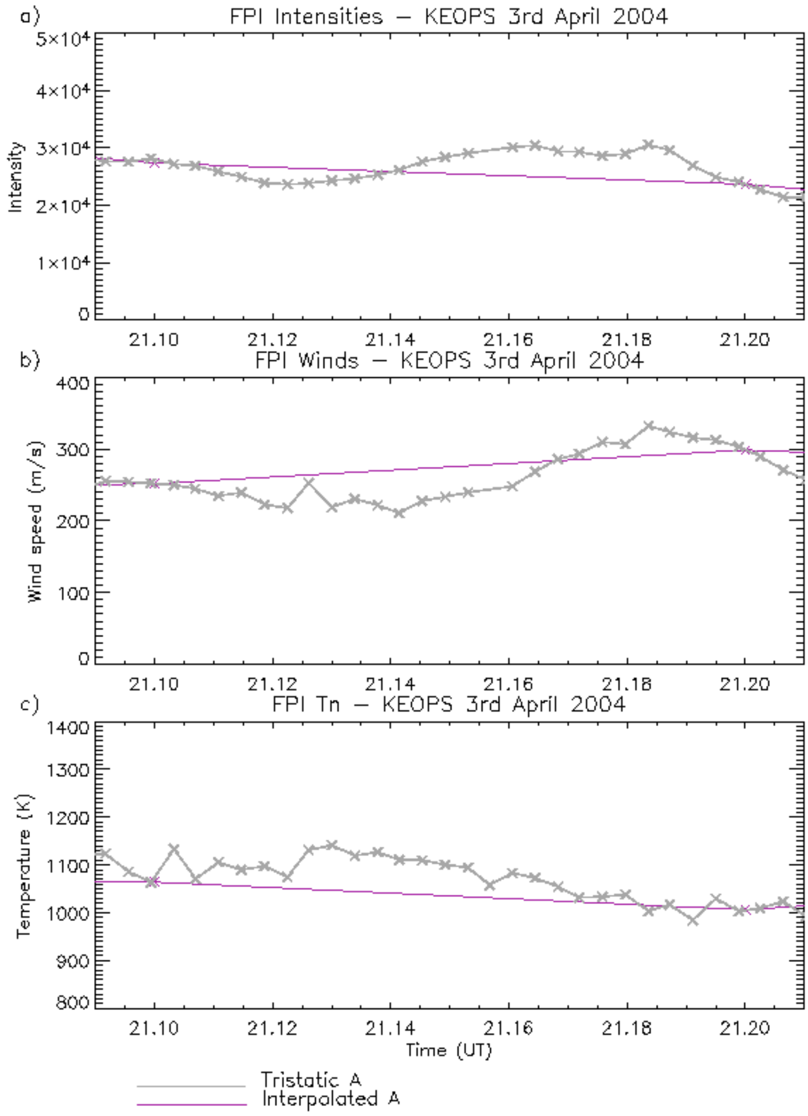

Figure 4.24 *10 minutes of high resolution FPI data from KEOPS on the 3rd April 2004, including 0.1hr interpolated tristatic A data (purple line): (a) intensities, (b) winds, and (c) temperatures.*

The Lomb-Scargle periodogram of the above data is shown in Figure 4.25. The spectral power is dependent on the time between data points, so the power is much larger for the tristatic A data than the other look directions. Therefore, the other directions have been multiplied by a factor of ten so that all the directions can be seen on the same plot. As the confidence levels are dependant on the time resolution of the data as well as the length of the night, a separate confidence level has been calculated for these other directions. The 70% level for these is shown with a dot-dot-dash line. The 90% level for the tristatic A data has also been removed for clarity.

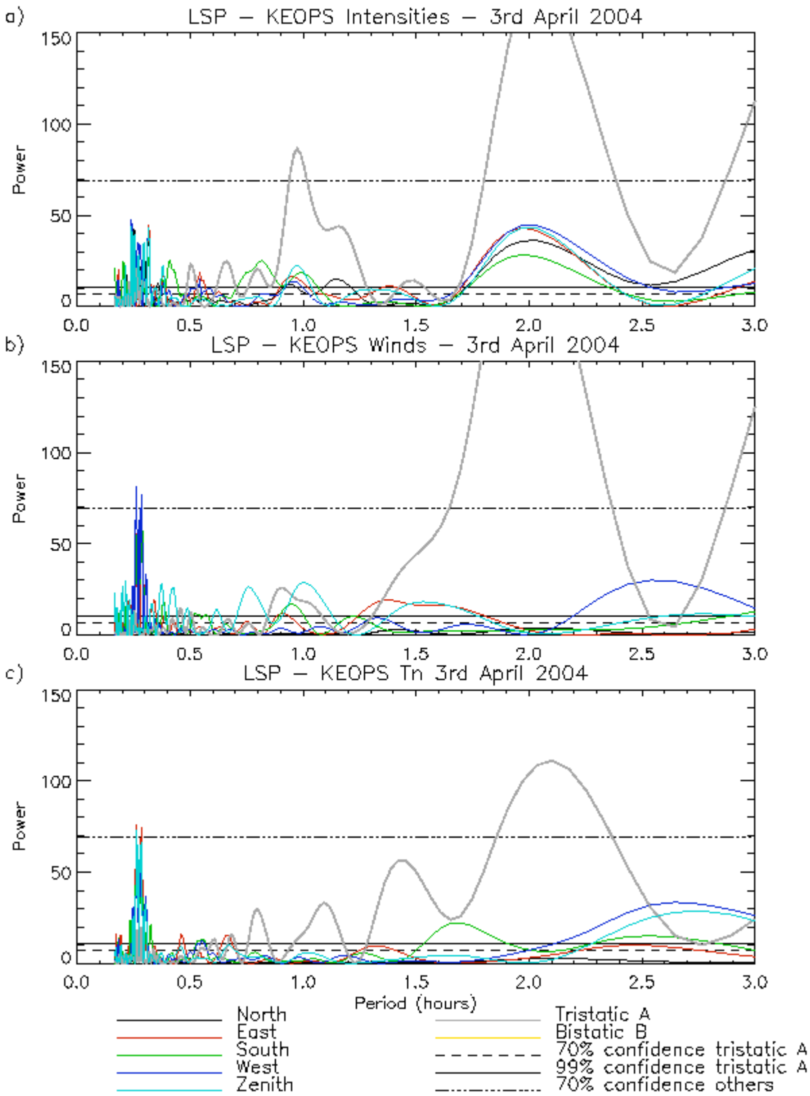

Figure 4.25 *Lomb-Scargle Periodogram of FPI data from KEOPS on the 3rd April 2004: (a) intensities, (b) winds, and (c) temperatures. The low resolution look directions are scaled by a factor of ten for comparisons.*

Several periodicities can be seen in all the parameters through the whole range of periods, from half the length of the night, down to the Brunt-Väisälä frequency (0.2 hour period). There is a strong wave of 2-hour period in all the parameters, within 0.1 hour. There appears to be a wave at 1.5 hours in all parameters, though for the winds, the power blends with the 2-hour wave. There are also periods around 1 hour, and several at shorter periods.

To see these short period waves more clearly, periods up to 1 hour are re-plotted on a different scale in Figure 4.26. For this plot, the low time resolution data (i.e. all but the tristatic A data) have been scaled by a factor of 4. Many peaks above the 99% confidence level can be seen at several periods down to 0.23 hours (13.8 minutes) in the intensities, 0.26 hours (15.6 minutes) in the winds and 0.27 hours (16.2 minutes) in the temperatures. Uncertainties for these high-resolution data are approximately ± 0.01 hours, which is roughly twice the time resolution of the data. Although few of the peaks in the other look directions reach the 70% confidence level, due to the poor sampling of the wave with the time resolution available, some of the periods seen in the tristatic A data are seen at lower powers in the other directions. Clear exceptions to this are the peaks at periods of 0.26 and 0.28 hours (15.6 and 16.8 minutes respectively) in the temperature data. These two periods are seen clearly in all of the look directions, and above the 70% confidence level in the east and zenith data.

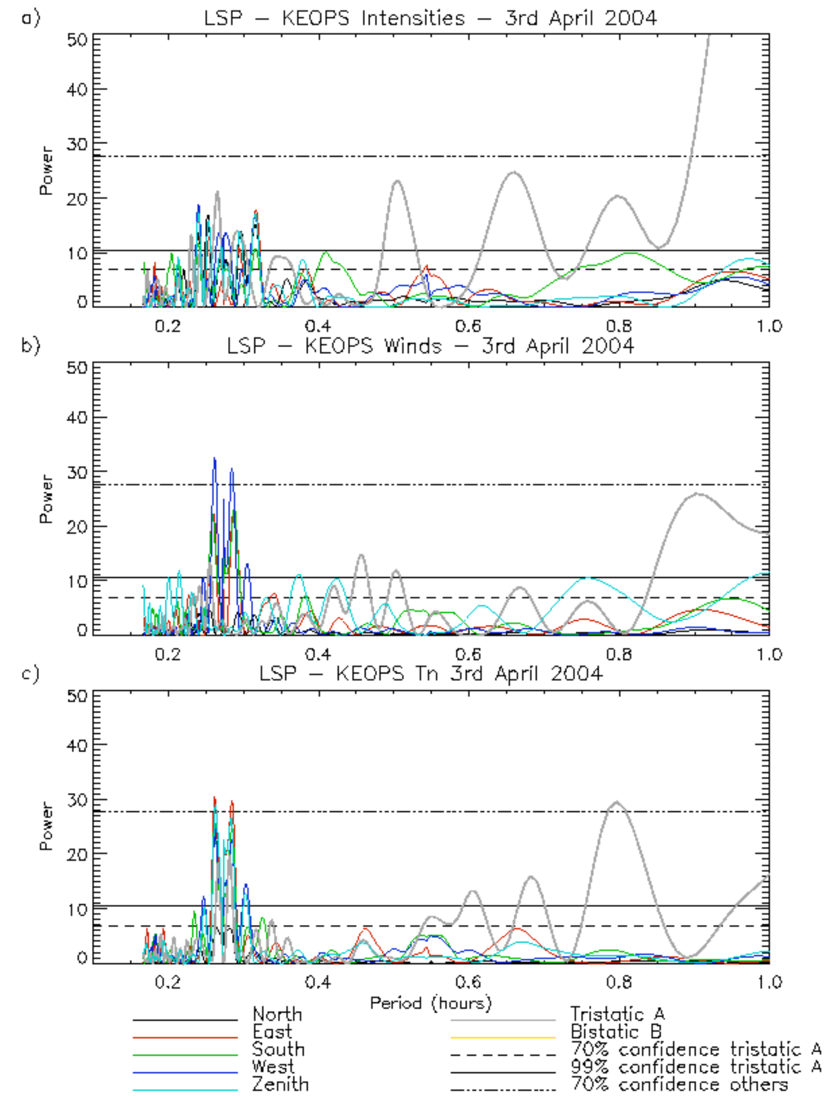

Figure 4.26 *Lomb-Scargle Periodogram up to periods of 1 hour, from KEOPS data on the 3rd April 2004: (a) intensities, (b) winds, and (c) temperatures. The low resolution look directions are scaled by a factor of 4.*

Many other periods are seen from these periods up to an hour, some of them are seen in more than one parameter, but several are only seen in one parameter. To better compare the three parameters, the tristatic A data are plotted on the same graph in Figure 4.27. These data are also smoothed, to remove the longer scale trends, and so to increase the spectral power available for the shorter periods, as was done for the second tristatic campaign data in section 4.4. Intensities are shown in red, winds in green, and temperatures in blue. A running smooth value of 120 points was removed from these data, which corresponds to approximately 30 minutes with this high time resolution data.

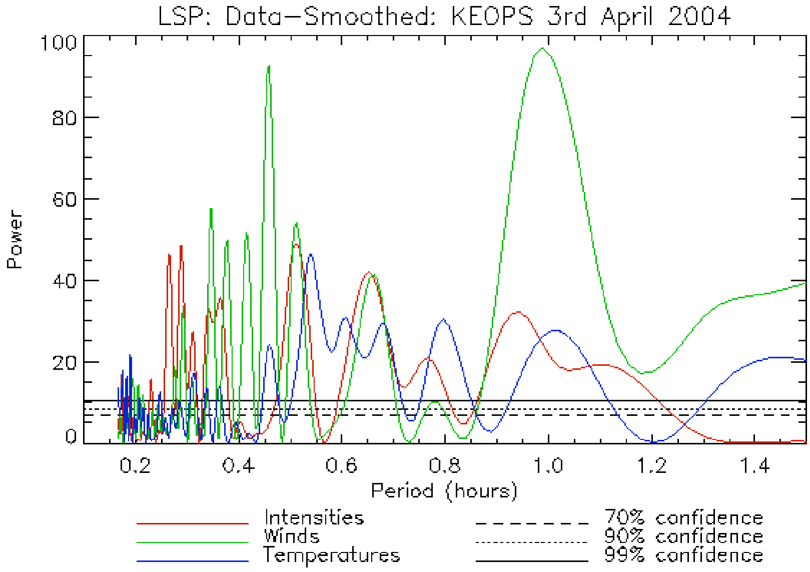

Figure 4.27 *Lomb-Scargle Periodogram of smoothed intensities, winds, and temperatures, from KEOPS data on the 3rd April 2004, using a 30-minute running smoothing value.*

Several periods have high spectral powers in two of the three parameters, such as at 0.46, 0.50, 0.66 and 1.0 hours. However, there are few periods where the peak is above the 70% confidence level in all three parameters. The only instance of this in this data set is at 0.32 hours (19.2 minutes). As this is the only occurrence out of several peaks, this is possibly just a coincidence, rather than a mechanism that drives all three parameters in the same way. From the results from the second tristatic campaign in section 4.4, it would be expected that the intensities and temperatures would produce peak powers at the same periods. The winds would not necessarily show these same periods however, due to the winds having a more complicated relationship with the gravity wave parameters, for example because there is a feedback between the velocity of the wave and the velocity of the background wind field. However, this effect is not clearly seen in Figure 4.27, as the periods that are seen in two of the parameters are not always the same pair, and are various combinations of the three parameters.

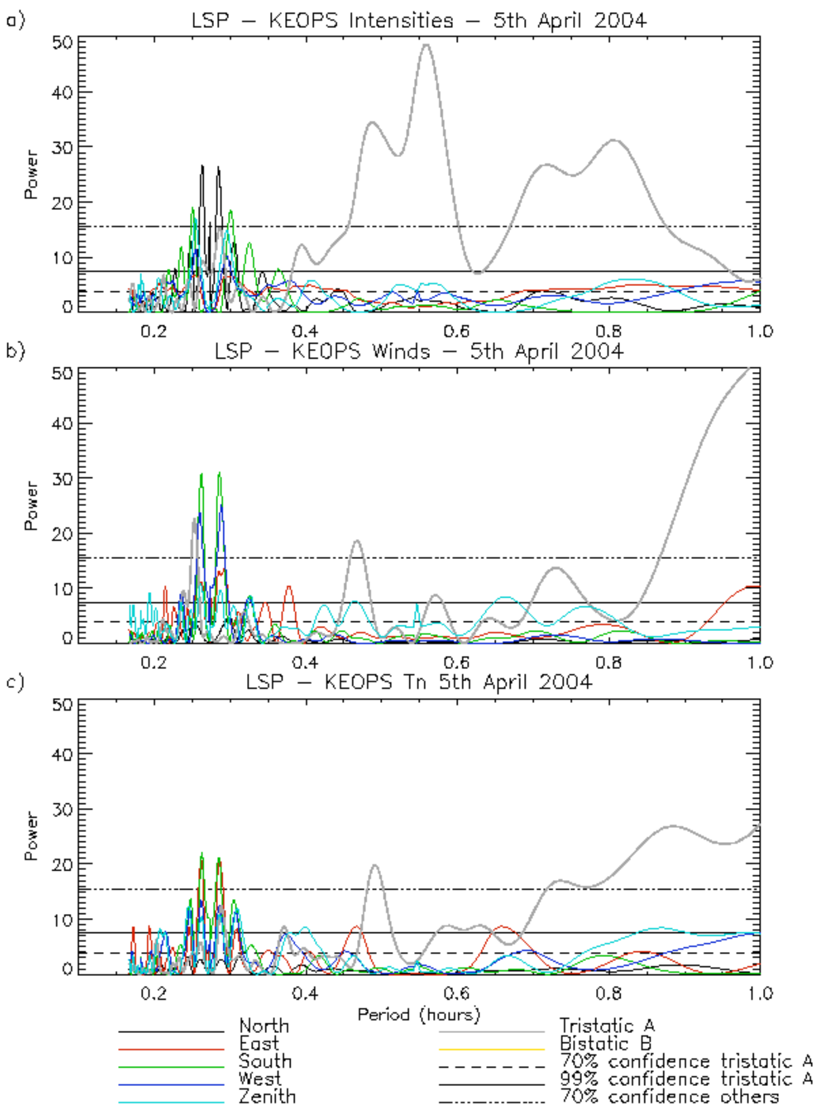

Figure 4.28 *Lomb-Scargle Periodogram up to periods of 1 hour, from KEOPS data on the 5th April 2004: (a) intensities, (b) winds, and (c) temperatures. The low resolution look directions are scaled by a factor of 4.*

Lomb – Scargle periodograms from the other clear nights during this experiment are shown in Figure 4.28, Figure 4.29, and Figure 4.30, from the 5th, 10th and 11th April 2004 respectively. Periods up to 1 hour are shown, and all the nights show several peaks with periods less than 1 hour. On the 5th April 2004, the intensities show a peak at 0.29 hours in the tristatic A data, above the 99% confidence level, that is also seen in the north intensities. This period is also seen in the winds and temperatures on this night: in the south, east, west, and zenith directions in the winds, and in all look directions for the temperatures, though only above the 70% confidence level in some of the directions. The peak at 2.6 hours is also seen in several of the look directions in each of the parameters. Amongst the other peaks throughout the data, periods of 0.48 and 0.57 hours are present in all three parameters.

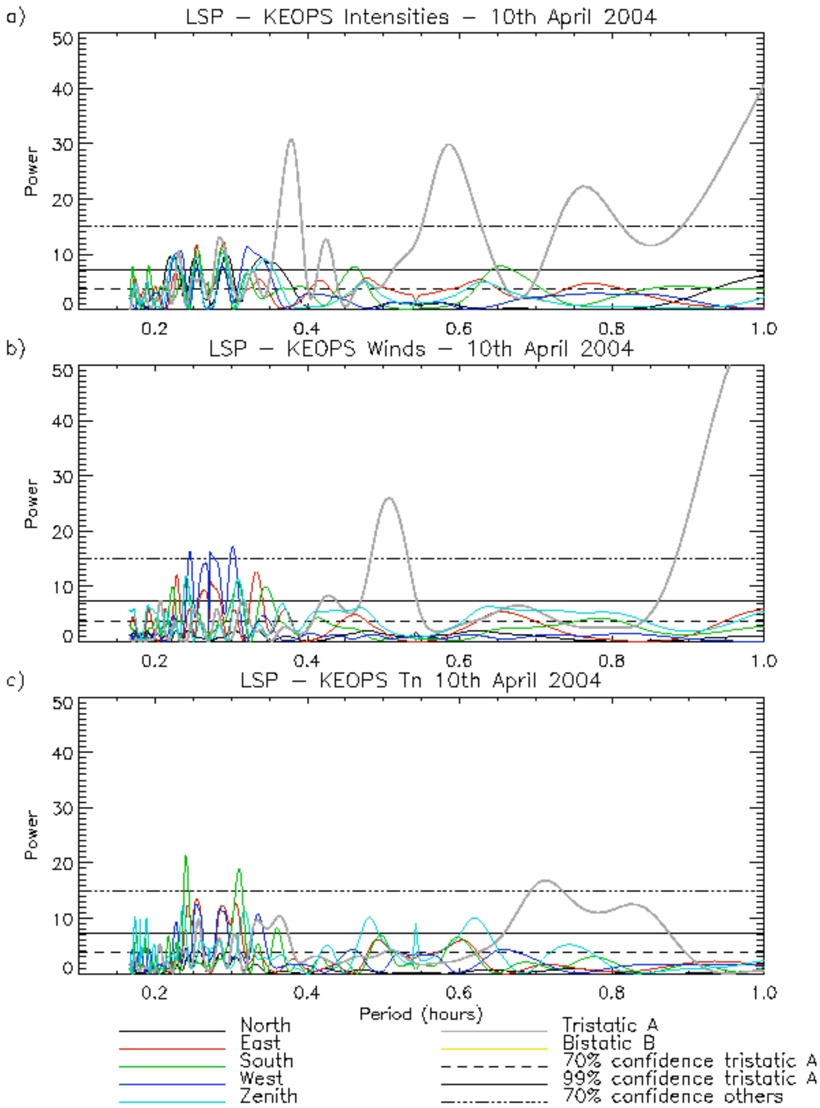

Figure 4.29 *Lomb-Scargle Periodogram up to periods of 1 hour, from KEOPS data on the 10th April 2004: (a) intensities, (b) winds, and (c) temperatures. The low resolution look directions are scaled by a factor of 4.*

The 10th April 2004 data in Figure 4.29 again show many short period gravity waves. The winds and temperatures however show slightly less waves than other nights at high confidence levels. For instance, there are only two peaks in the temperatures above the 70% confidence level, in the south direction. There are peaks in the intensities in all the look directions at 2.3, 2.5, and 2.9 hours, though again only above the 70% level in the high time resolution tristatic A data. Several periods are also seen in the data of the 11th April 2004 in Figure 4.30, but again there is little correlation between the three parameters for most of these short period waves.

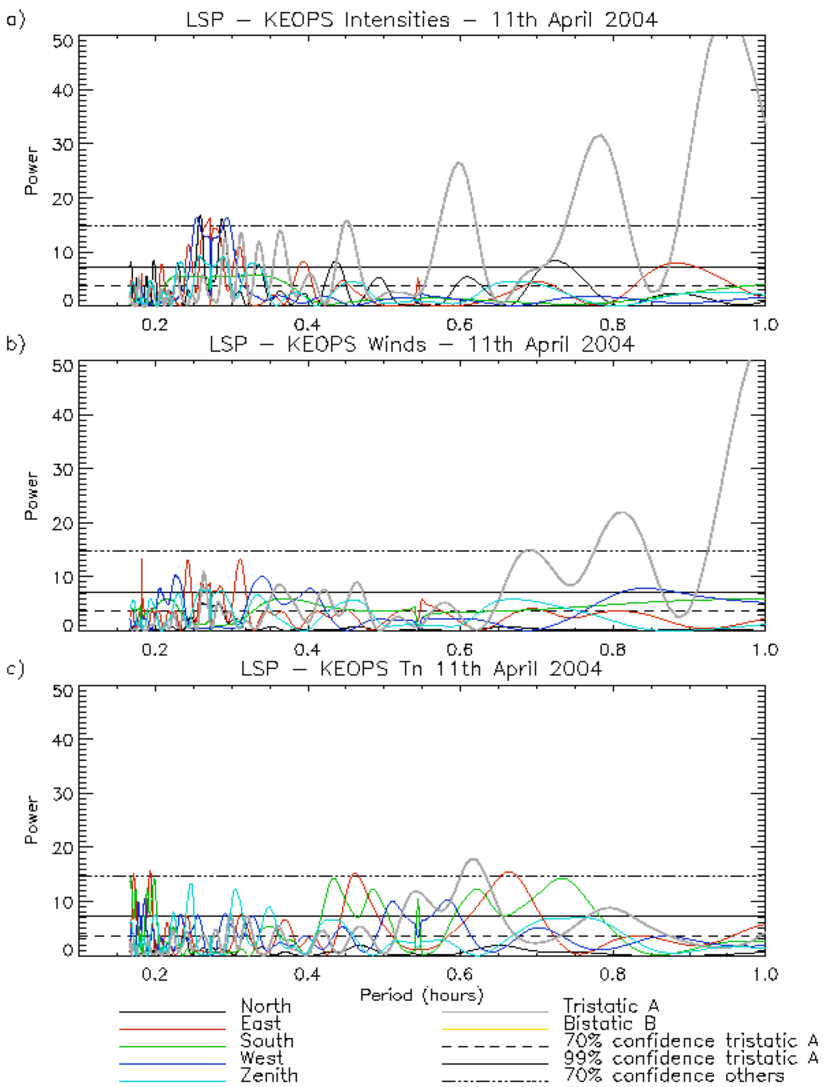

Figure 4.30 *Lomb-Scargle Periodogram up to periods of 1 hour, from KEOPS data on the 11th April 2004: (a) intensities, (b) winds, and (c) temperatures. The low resolution look directions are scaled by a factor of 4.*

These four nights of high resolution data show that short period gravity waves are often seen in the FPI data, and waves are seen down to the order of the Brunt-Väisälä frequency, the limiting frequency at which waves can be sustained. That several short period waves are seen in all of these nights implies that they are present reasonably often, and the nights shown are not unrepresentative of the typical state of the thermosphere. Comparisons of structures seen in the high-resolution data with the interpolated data show a limit to the scale on which the thermosphere can vary of approximately one minute.

4.6 Svalbard and Mainland Comparison

Despite the amount of data collected from each of the FPIs (see Chapter 5), there are few numbers of nights where data are available from both of the mainland sites (Sodankylä and KEOPS), and from Svalbard. This is mostly due to problems with the detectors at one or other of the sites. There was approximately a week in September 2002 when data from all three sites are available, however, at this time of year there are only approximately three hours of dark enough night to collect data at Svalbard. As some of the gravity waves are likely to have periods longer than 1.5 hours, this is not a sufficiently long data set to observe them. There was a week in December 2003 with clear skies in Svalbard and one of the mainland sites on most nights. Some of these data are discussed further below. There were more coincident clear nights at both locations in the 2005 – 2006 season, after the new Andor was installed on the Svalbard FPI at the beginning of December 2005. Much of this winter was cloudy, in all three sites, reducing the possible number of nights considerably. Many of the nights during a clear spell in Svalbard were during a period of low geomagnetic activity, resulting in low intensities so that the interferometer peaks could not be fitted for much of the night. There were however some dates in 2006 that were clear in two or even all three sites. However, to compare the gravity waves found in the FPI data with the SIF data, a night from the 2003 season is used for the analysis below, as the clear nights available in all three sites were after the time when SIF was shut down for the summer.

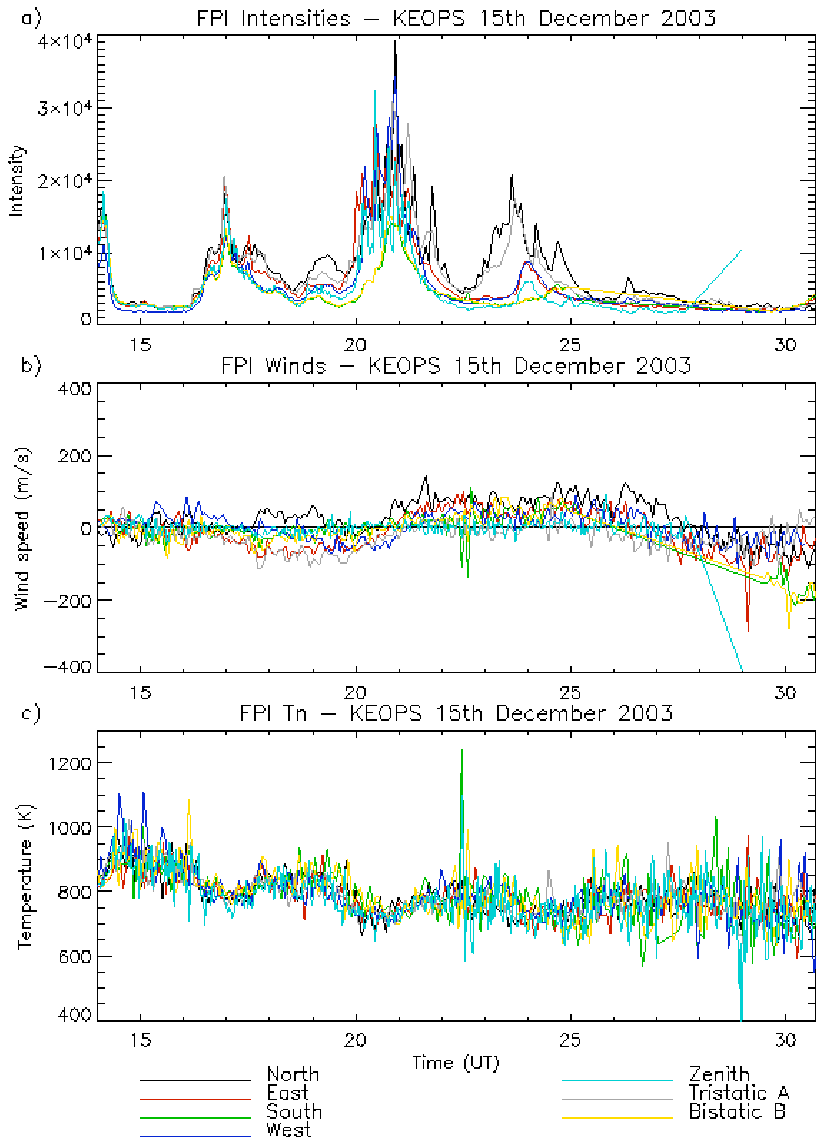

Figure 4.31 *KEOPS data from the 15th December 2003: a) intensities, b) winds, and c) temperatures.*

Data from the 15th December 2003 from KEOPS are shown in Figure 4.31. Four periods of precipitation can be seen in the intensities in Figure 4.31a, generally stronger in the more northern or central look directions, and weaker in the south and B positions. These bursts of particle precipitation are all in evenly spaced periods, approximately 3.5 hours apart. This is also reflected in the temperatures in Figure 4.31c, which show a clear wave structure in all of the look directions. The spike in the south (green line) data at approximately 22:30UT is not likely to be real as it is just one data point, and is likely to be due to the fitting of poor fringes. Contrary to the more noisy data at the end of the night of the 25th November 2003 in section 4.4, which was due to cloud cover, the noise in the temperatures in Figure 4.31c were just due to the low intensities, which create poor interference fringes that are hard to fit temperature profiles to.

The periodicities that are clearly seen in the data are also detected by the Lomb – Scargle periodograms. In addition, there are long period waves, especially in the winds and temperatures, and for instance there is a strong period at 6 hours in the temperatures, which is likely to be an atmospheric tide. Therefore, in Figure 4.32 the LSPs from smoothed data are plotted to remove this 6-hour wave and any other longer period waves, which increases the spectral power in the shorter periods. A 60-point running smooth was used, which corresponds to a smoothing of approximately 4.3 hours. This again was chosen as a compromise between removing the long period waves but still being able to detect the period of interest at 3.5 hours. A wave with period 3.5±0.1 hours is clearly seen in the intensities in Figure 4.32a. Several other periods are able to be seen in the zenith data at shorter periods, above the 70% confidence level. These are not seen as strongly in the other directions as the zenith data still had twice the time resolution at this time, following the second tristatic campaign on the 25th November 2003. The winds do not see this period so clearly due to the same complications mentioned earlier. The temperatures also detect a periodicity at 4.4 hours, which in most of the directions merges with the 3.5-hour wave.

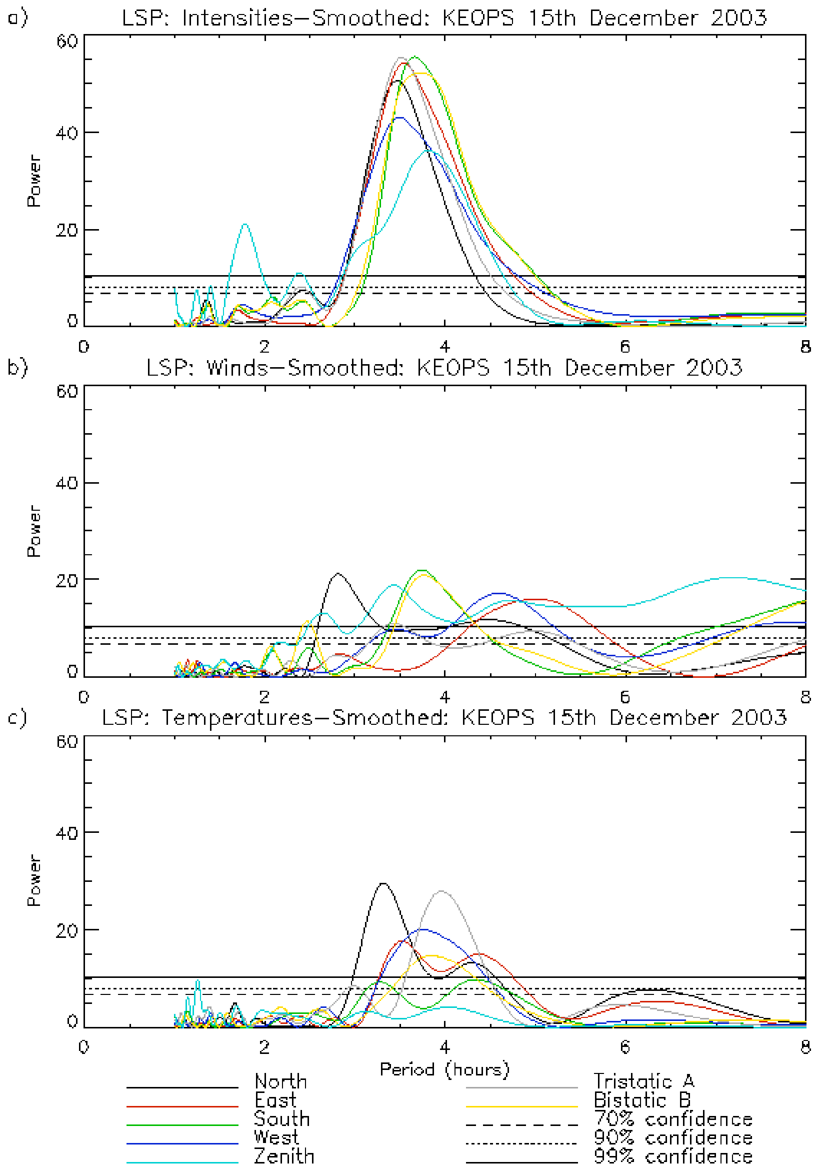

Figure 4.32 *Lomb – Scargle periodograms from KEOPS data: a) intensities, b) winds, and c) temperatures, from the 15th December 2003.*

The purpose of choosing this night of data to study was that Svalbard data are also available from this time, as FPI data are shown in Figure 4.33. This FPI had an Astrocam detector, which had a much poorer time resolution of approximately 10 minutes, compared with the 4.3 minutes of the KEOPS data, provided by the Andor detector. This is reflected in the quality of the data in Figure 4.33 compared with Figure 4.31. In addition, the low intensities particularly at the start of the night have resulted in some larger gaps in the data.

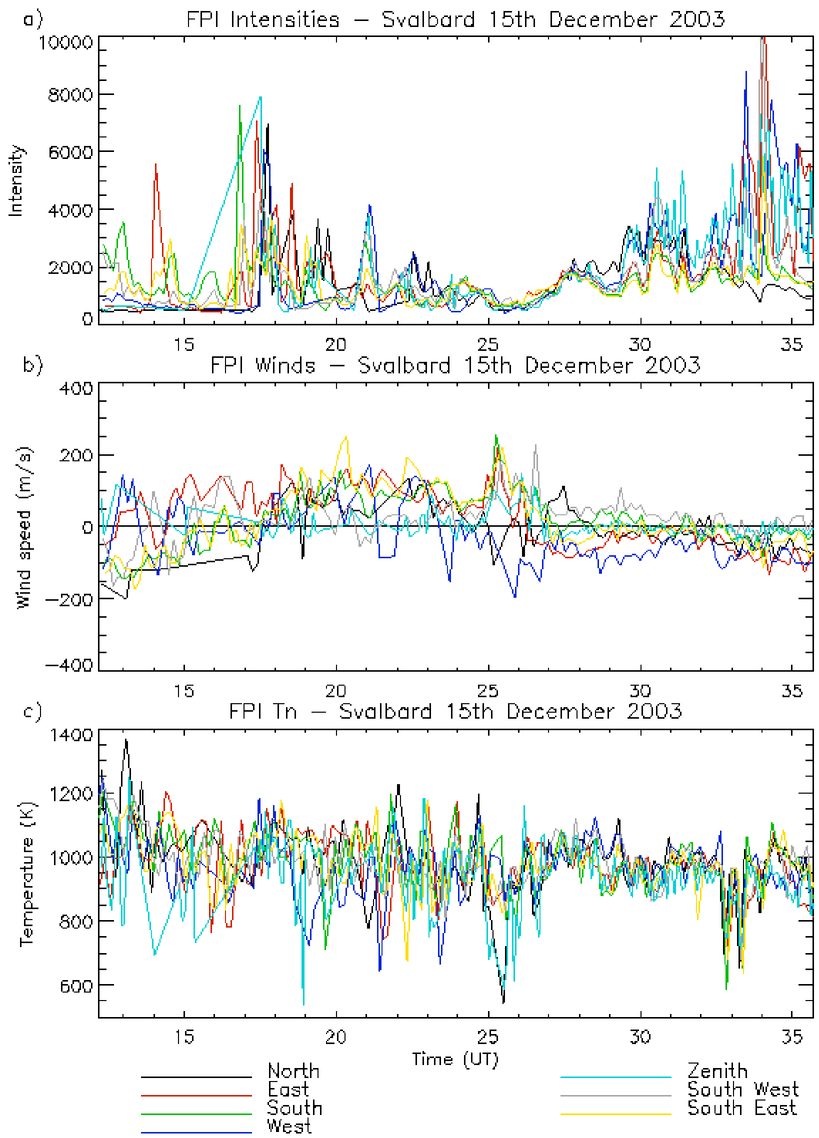

Figure 4.33 *Svalbard data from the 15th December 2003: a) intensities, b) winds, and c) temperatures.*

The Lomb – Scargle periodogram in Figure 4.34 shows that waves seen at KEOPS were present in the Svalbard data as well. The data have been smoothed as for the KEOPS data, with a 4.3-hour smooth, which for this poorer time resolution data corresponds to a 26 point running smoothing value. The 3.5-hour wave is not as strongly seen in the Svalbard data as the KEOPS data, but it is present in most of the look directions in the intensities and temperatures. Due to the decreased time resolution, the minimum error on the period determination is 0.3 hours, so although the peak periods do vary between the look directions, they match the 3.5-hour wave of the KEOPS data within the uncertainties.

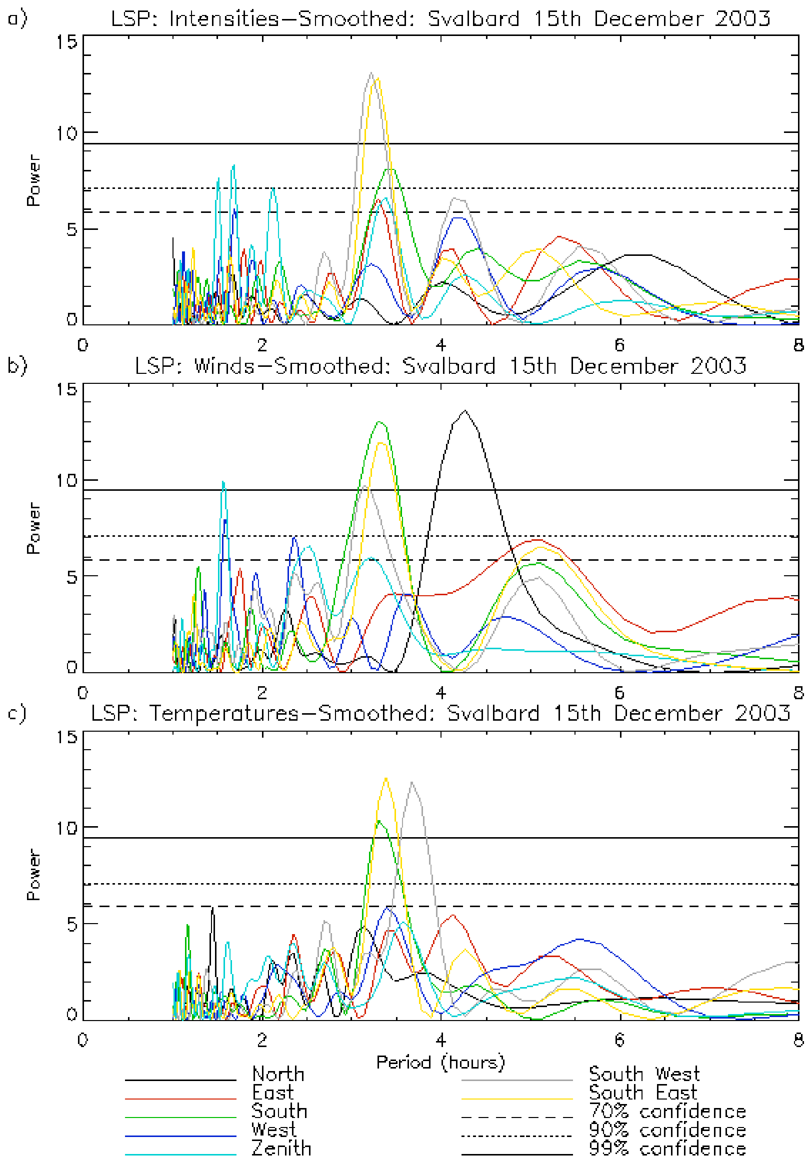

Figure 4.34 *Lomb – Scargle periodograms from Svalbard data: a) intensities, b) winds, and c) temperatures, from the 15th December 2003.*

The 3.5-hour wave is seen in the Svalbard intensities in Figure 4.34a most strongly in the south-east and south-west directions, and is only detected above the 99% confidence level in these directions. The wave in the south is between these two directions so would be expected to be of similar power, and it is likely to be less due to the south peak being wider than the south-east and south-west, resulting in the power being spread over a wider period range. The east and zenith directions have similar peak widths to the south-east and south-west, but have smaller powers, and the north and west directions are well below the 70% confidence level. The wave is therefore stronger to the south of the site and weaker to the north, implying a source location to the south of Svalbard. Similar ratios are also seen in the winds and temperatures for this night in Figure 4.34b and Figure 4.34c respectively.

The 15th December 2003 was also chosen as a case study night for gravity waves as the Spectrograph Imaging Facility was running during this period. SIF is at the same location as the Svalbard FPI so provides information on the same waves, but it views different emission lines, which originate from different altitudes, so some information on the vertical distribution of the wave can be obtained. The photometer data are simple time series so can be easily analysed for gravity waves. Figure 4.35 shows the data from the two photometers from this night, the blue line is the $H\beta$ at 4861Å, and the green line is an N_2^+ line, at 4652Å. The FPI data end just after 35 hours (11UT on the 16th December) and before this shows the same general increase in activity levels. Spikes in the nitrogen (green) line at for example 17UT, 21UT, 24UT, and 31 hours show auroral arcs passing over the field of view (the magnetic field line) as these nitrogen emissions are formed from electron aurora and so spike above the more diffuse proton aurora background.

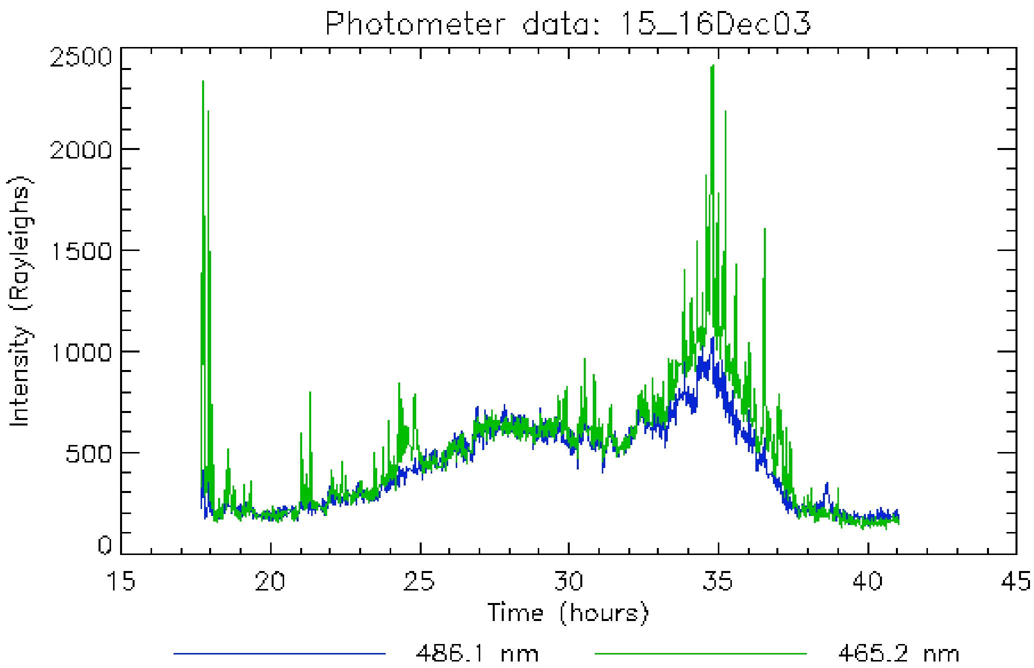

Figure 4.35 *SIF photometer data from the 15th December 2003, Hβ at 4861Å, blue line and N_2^+ at 4652Å, green line.*

The LSP of the photometer data in Figure 4.35 is shown in Figure 4.36 below. The data were first smoothed using a 260-point running smooth, which for the 1-second resolution data displayed in Figure 4.35 corresponds to the same 4.3 hour smoothing of the two FPI data sets. Waves, with periods up to 8 hours, are shown for the two photometer channels separately, for clarity. Both photometers show a clear wave with a period of 3.5 hours, above the 99% confidence level for the data set, as well as several other waves. The waves at approximately 4.3 hours and 2.1 hours are also seen in the Svalbard FPI data (Figure 4.34).

The photometer data have a much higher time resolution than the FPI data, and have integration times of 200ms. For this analysis, 60-second data are used due to the large array sizes in calculating the LSP. The 3.5-hour wave on the 4652Å line photometer data (Figure 4.36b) is at 3.42±0.02 hours. This therefore gives a more accurate period to the wave, which is also within the errors on the FPI data.

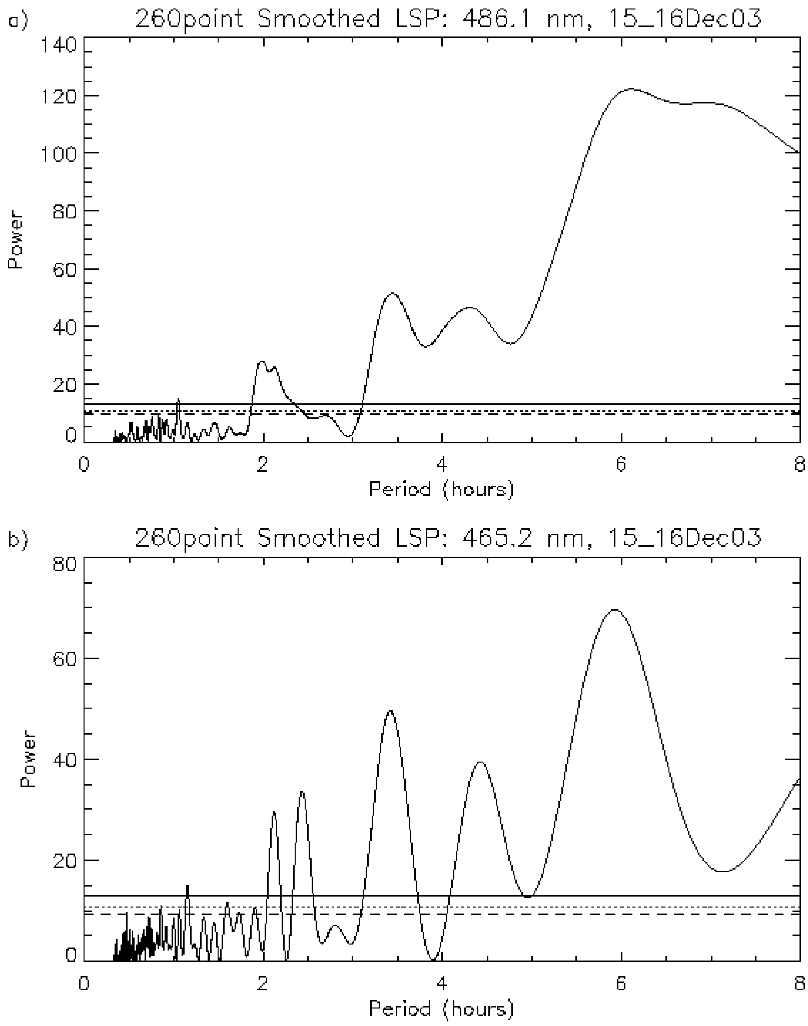

Figure 4.36 *LSP of smoothed SIF photometer data from the 15th December 2003, a) Hβ at 4861Å, and b) N_2^+ at 4652Å.*

The peaks are less well defined for the Hβ than the 4652Å N_2^+ line, which is due to the Hβ emissions coming from proton aurorae. These are more diffuse than electron aurorae due to the interactions of the protons with the neutral hydrogen atoms in the atmosphere, which are not constrained to the magnetic field lines, allowing the emissions to spread horizontally as the proton enters the atmosphere. Both the Hβ and the 4652Å line emissions occur at lower altitudes than the 6300Å line, but they are not from as well defined emission layers. N_2^+ emission originate from 90km to 125km (Chamberlain, 1961), and Hβ emissions generally occur from around 100 – 120km altitude, but the altitude will vary with the energy of the incoming protons and will be higher for less energetic protons. This does however show, as the 3.5-hour wave is seen strongly in both lines, that the waves cover the altitude region between all three emission lines.

The photometers are aligned along the magnetic field line (see section 2.3), which is just to the south of the Svalbard FPI zenith position. The FPI detect the wave with a high power in the southerly directions, but with a lower power in the zenith, then even lower towards the north direction. A possibility is that rather than the wave just dissipating by this latitude, it may also be propagating with a downwards vertical component, as the wave is seen at the lower altitudes at the Svalbard observatory location, but is weaker in the higher FPI zenith data. This however cannot be confirmed as the emission heights are not exactly known, and the FPI intensities are not calibrated, so the wave amplitudes cannot be directly compared with the photometer data. This would be interesting to investigate further, when the FPI intensities are calibrated (see Chapter 6) so that FPI datasets can be compared, and these can be compared with other instruments such as the photometers.

An auroral source region can be further confirmed, other than from the periods of precipitation seen in the KEOPS intensities, and the decrease in power towards the northern Svalbard directions, by comparing with other data sets. Figure 4.37 shows the CUTLASS coherent scatter plot from the Pykkvibaer radar in Iceland. The top plot shows the backscatter power, and the bottom is the ionospheric velocity. There are enhancements in the power at the latitude of KEOPS (67.8°N) between 13-14UT, 16:30-18UT, and 20-21UT, that are all approximately 3.5 hours apart. There are other enhancements too which may have contributed to the other periods detected in the FPI data. These enhancements are all between 65°N and 69°N, with very little activity at the latitude of Svalbard, 78°N. This again corroborates an auroral source region, at KEOPS latitudes.

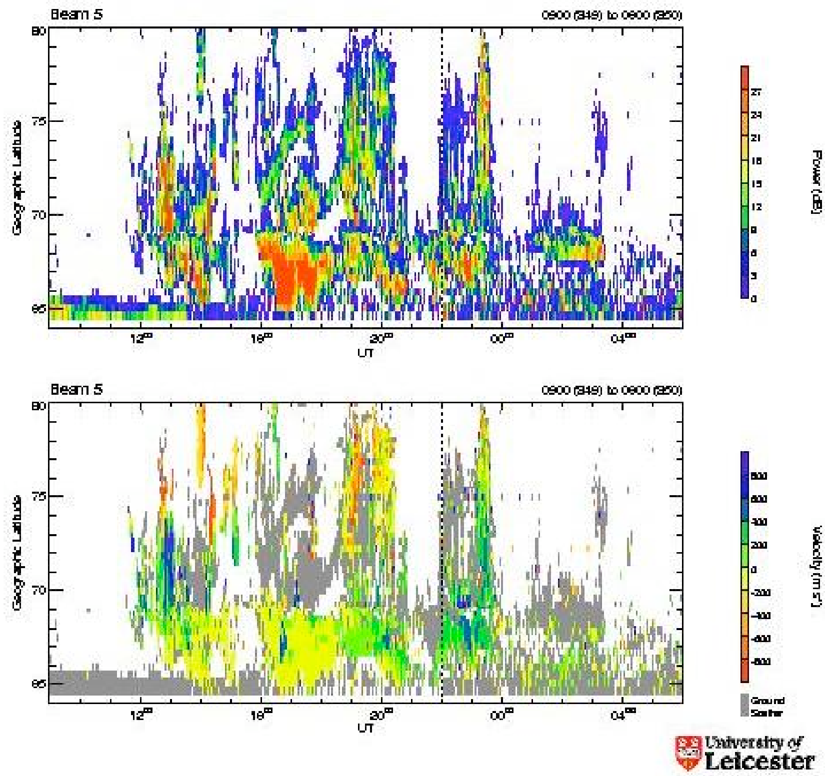

Figure 4.37 *CUTLASS Pykkvibaer beam 5 power (top) and velocity (bottom), on the 15th December 2003.*

A source of the gravity waves of Joule heating from particle precipitation is also supported by the results of the Lomb-Scargle analysis on the magnetometer data for this night, shown in Figure 4.38. The 3.5-hour wave is not seen in the Kiruna B_x component of the magnetometer data, which is affected by the electrojet activity. It would be expected to be seen if heating due to electrojet currents was the proposed mechanism for the production of the 3.5-hour wave, so as it is not detected, particle precipitation is likely to be the production method.

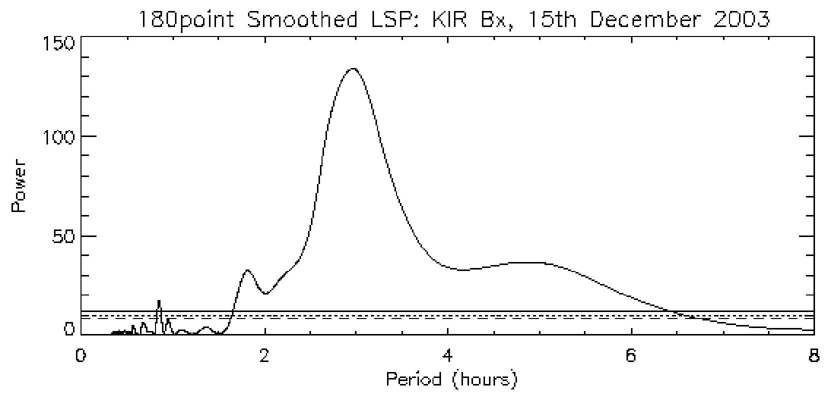

Figure 4.38 *LSP of IMAGE B_x component data from the Kiruna magnetometer, on the 15^{th} December 2003.*

To obtain information on the velocity and wavelength of the gravity wave detected, the smoothed temperatures used to create Figure 4.32c are plotted in Figure 4.39, but with an additional 15-point (~1 hour) running smoothing removed. As for the second tristatic campaign, this allows the lags between the different look directions to be seen. Only the four cardinal directions are shown as the meridional and zonal lags are needed to determine the velocity and propagation direction. Only data from the start of the night until midnight are shown as the data become noisy at the end of the night, and the shorter scale allows the lags to be seen clearly.

A clear lag can be seen between the north (black) and south (green) directions in Figure 4.39 in all of the cycles of the wave shown. The lag, of the south behind the north temperatures, is on average 1.0 ±0.1 hours. There is no discernable lag between the east and west data. The zonal lag is therefore τ_x = 0 hours, and the meridional, τ_y=1.0 hours. Using these lags and a distance between the look directions of 480km, and putting these into equation [38] gives a phase speed v of 130±20ms^{-1}. As there is no zonal lag for these data, the propagation direction is southerly. This is consistent with an auroral oval source region, as from for example all sky camera data, the auroral oval was to the north of KEOPS.

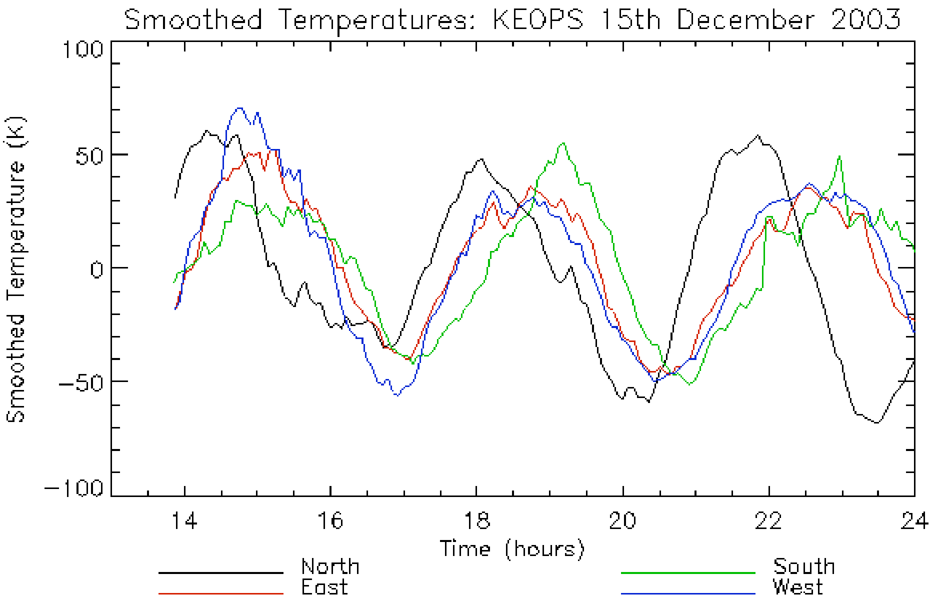

Figure 4.39 *60-point smoothed temperatures from KEOPS from the 15th December 2003, with a further 15-point running smooth removed to show lags between directions.*

The speed of the wave can also be used to determine the horizontal wavelength, which for this gravity wave λ_h is 1680±200km. The Svalbard data are too noisy to determine the lags clearly, but the presence of the wave at Svalbard indicates that the KEOPS wave also had a northerly propagation direction as it reached the Svalbard site. The wave therefore originated in the auroral oval and propagated out poleward and equatorward to the two FPI sites. The values calculated here are of the order of magnitude expected from previous work, and are similar to the values of the second tristatic campaign, in which a 1.8 hour wave had a velocity of 250ms^{-1}, a south-south-eastward propagation direction and a horizontal wavelength of 1600km. This shows that these are not unusual cases and they are likely to be typical properties of the thermospheric high-latitude waves.

5 ATMOSPHERIC GRAVITY WAVES: STATISTICAL ANALYSIS

5.1 INTRODUCTION

In depth studies of individual gravity wave events are useful to ascertain information on the origin and source of gravity waves, and to determine their properties. However, from only a few nights of data, it is not certain whether the waves seen are the usual behaviour of the thermosphere, or whether there are atypical conditions. Therefore, the data presented in this chapter makes use of the large database of FPI measurements available from operating the instruments at various sites over many years. This can be used to statistically analyse the data to determine how frequently gravity waves are present, what periods they have, and what factors influence their amplitude or frequency of occurrence.

Determining the proportion of nights where gravity waves are present will obviously be limited by the data available. The airglow and auroral emissions are only detectable over background solar emissions at night times, hence data can only be collected during the autumn, winter and spring at the latitudes of the instruments, inside the Arctic Circle. This will limit studies of the seasonal variability of gravity wave occurrence, as the day lit polar summer cannot be observed at all, and the lengths of the season and nights in the autumn and spring are much less than the complete darkness of the winter polar nights. Fritts & Alexander (2003) report in their review of gravity waves a greater number of gravity waves in winter, and less in the summer, in the high latitude northern mesosphere using ionosonde and radar data. This effect however was observed at lower altitudes where the source of the gravity waves was likely to be tropospheric. Ogawa et al., (1987) also found that there was a maximum in gravity wave activity in winter and a minimum in summer in medium scale TIDs from satellite data. However, summer/winter effects are not detectable with optical data as is used here. Ogawa et al., (1987) also found that the majority of the waves propagate equatorward, and that the number of waves detected did not increase with geomagnetic activity.

Data are not necessarily collected every night through the observing season, due to instrument failures, computer crashes, power cuts, and similar interruptions. When the instruments are running correctly and collecting data, the next obvious hindrance to collecting useable data is the amount of cloud cover. This significantly reduces the amount of data available for the statistical survey, and will bias the results. However, as cloud cover does not influence high altitude gravity waves, the percentage of nights where gravity waves are observed should be representative of the whole observing season.

Operational periods have varied between instruments over the years that APL has been collecting data. Data availability from the various sites are given in section 5.2. The Sodankylä and Svalbard FPIs have only been running for a few years, compared with the long run of the KEOPS FPI that has collected data since 1981, although this was at a nearby location nearer the Kiruna FPI for the majority of this period. The Kiruna dataset has therefore been accumulating for nearly two solar cycles, so the dependence of the number of gravity waves formed on solar energy input to the thermosphere can be investigated with this data set. There has not been much progress on inferring the climatology of global TIDs (Fritts & Alexander (2003)) because of the limitations of each observational data set, in terms for example of location, altitude and frequency range of observations, and period of the dataset. APL's FPI data is limited in the same way, in just observing one altitude in the high latitude northern hemisphere, during hours of darkness. The climatology of gravity waves within this limit is still of importance however and can provide significant new information. Additional Solar affects through geomagnetic activity levels can be studied with all of the sites, by binning geomagnetically quiet, moderate, and active nights separately.

Differences in the gravity waves present in data from each of the sites can also be investigated. For example, would the recent KEOPS data with its higher time resolution data due to the Andor detector detect more waves than Sodankylä, Svalbard, and Skibotn with their lower time resolutions? Another important comparison is between the mainland sites with Svalbard, as the mainland sites are in the auroral oval area, while Svalbard is in the polar cap region. For gravity waves with auroral sources, it may be expected that the mainland sites would show more gravity waves as they are nearer to the source region, and the waves might not propagate in the direction of Svalbard in the polar cap. However, due to problems such as detector failures, there are relatively few nights where Svalbard and mainland data are both available. There are only concurrent data in one week in September 2002, approximately three weeks in December 2003, and four months from December 2005 to March 2006. The nights in Svalbard in September are too short to be able to detect many gravity waves, so this period is discounted. Of the other periods, many of the nights were cloudy at one of the sites. The only nights that were clear at all three sites were the 16th December 2003, the 14th February 2006, and the 6th March 2006. There are, however, several nights over the 2005 – 2006 season where two of the sites were clear, for comparisons.

Whether any particular periods are preferred over others or not should be determinable by collating the periods of waves observed over a statistically significant number of nights. The periods of the gravity waves observed are limited by the length of the data set, i.e. the length of the night, and of the time resolution of the data, so this will bias the results by excluding periods outside of this range.

It is important to know the dependence of the analysis on background intensity levels, when comparing different data sets, both data from different instruments, data from one instrument with different detectors, and at various points in a detector's lifetime. This was discussed fully in section 4.3. The Lomb – Scargle analysis programme normalises the time series using the mean of the data, which means that various data sets from different sites and under different conditions can be directly compared.

5.2 DATABASE

The first priority in performing a statistical survey of the data for gravity waves is to determine which nights of data can be used. This first means identifying when data are available at each site, and then which of these nights have good quality data and are cloud free. A list of the data available for KEOPS, Sodankylä, and Svalbard is given in Table 5.1 below.

Determining cloud cover levels for each of these nights is a time consuming task due to the number of nights and the various sources of information on the weather status. The simplest method of quickly determining if a night is cloudy or not is to use all sky camera keograms. These provide a quick view of the whole night, and there is a large database of keograms for many years available from the MIRACLE website, for several sites. There are two problems however with using keogram data. Firstly, keograms are not always available when FPI data are, due either to the instrument not installed or working at that time, or to the camera not running that night. The latter is often a problem with the Svalbard (Longyearbyen) ASC, as it has an intensified camera that is very sensitive and will saturate with too much background light, so it is turned off further in time from dusk and dawn, and for over a week around the full Moon period. ASCs are generally switched off earlier than the FPIs, which are still able to obtain data during twilight hours. Additionally, the Kiruna FPI is not run as part of the MIRACLE network, and so keograms are not made. Therefore, the new APL ASC data are used for KEOPS data, again for which only individual images throughout the night are available. The second problem with the keograms is that they are made with the green line (5577Å) images. The easiest way from a white light ASC image to determine if it is clear or not is whether stars are present in the image. Stars however are not seen in green line keograms. However, if structured aurora can be seen, then it can be concluded that the sky is clear. If there is diffuse aurora then this cannot be distinguished from aurora seen through a thin cloud layer. If there is no aurora at all, the image will be black, and cloud conditions cannot be determined from the keogram.

If the cloud cover levels cannot be determined from the keograms, another method needs to be found. If white light ASC images are available (as is the case for the Sodankylä and KEOPS ASCs), then these can be inspected. Sodankylä images need to be requested individually from FMI, so this is not practical for large numbers of nights. Keograms are not made from KEOPS ASC images, so inspecting the individual images is the usual method for determining cloud cover levels at this site. There are keograms at several sites around the FPI viewing volumes, e.g. Kevo, Muonio, and Kilpisjärvi, as well as Sodankylä. If all these ASCs show overcast skies, it is likely that the whole region is covered in cloud, so KEOPS can be assumed cloudy too. In addition to the ASCs, weather satellite images are available at various intervals on most nights. The infrared images from the European weather satellite, METEOSAT, provide night time weather information. An additional method is to use the FPI data itself to see if a night is cloudy, as discussed in section 2.2.6.

The nights of data are divided into geomagnetic activity levels. Quiet is determined as having all 3-hourly Kp indices over the 24 hour period of a night (from midday to midday on the following day) as being below 3. Moderate has levels between 3 and 5 for the entire night, whereas active nights need to have a Kp value greater than 7 at some point during the night. As the number of quiet nights is much greater than that of active nights, only the quiet nights have a sufficient statistical sample for any one observing season to be able to make statistically significant conclusions. An additional problem however with the geomagnetically quiet nights is that the signal to noise ratio is often very poor. The parameters, in particular the temperatures, cannot be sufficiently accurately determined, so these nights are also not included in this analysis. Therefore, to test for geomagnetic activity dependences, the whole database is used from all available years from each site. Many of the nights will not fit into these categories, and will be partially quiet but with periods of Kp between 3 and 7, and so the total number of clear nights are separately analysed as well, to increase the statistical sample.

Season		***KEOPS***	***Sodankylä***	***Svalbard***
2005 – 2006	Dates of data	27/8/05 – 15/3/06	8/9/05 – 23/3/06	7/12/05 – 20/3/06
	Detector	Andor	Astrocam	Andor
	Cycle time	3.5 minutes	7.5 minutes	4.0 minutes
	Clear nights	22 2 2 35	29 0 1 41	12 0 0 22
2004 – 2005	Dates of data	16/9/04 – 27/4/05	6/11/04 – 1/4/05	no data
	Detector	Andor	Astrocam	
	Cycle time	3.5 minutes	6.5 minutes	
	Clear nights	12 4 1 48	3 0 1 12	
2003 – 2004	Dates of data	11/9/03 - 21/12/03, 12/2/04 – 27/4/04	1/9/03 – 21/9/03, 13/11/03 – 30/4/04	30/11/03 – 18/12/03
	Detector	Andor	Astrocam	Astrocam
	Cycle time	3.5 minutes	11.0 minutes	10.0 minutes
	Clear nights	11 4 2 21	13 8 0 57	1 4 0 11
2002 – 2003	Dates of data	31/1/03 – 15/4/03	15/11/02 – 15/4/03	19/9/02 – 4/12/02
	Detector	Andor	Astrocam	Astrocam
	Cycle time	3.5 minutes	12.5 minutes	14.0 minutes
	Clear nights	4 4 0 27	3 2 0 23	2 0 1 21
2001 – 2002	Dates of data	20/1/02 – 30/4/02	no data	19/9/01 – 30/3/02
	Detector	Astrocam		Astrocam
	Cycle time	14.5 minutes		15.0 minutes
	Clear nights	13 1 1 23		43 2 3 78
2000 – 2001	Dates of data	15/8/00 – 27/4/01	no data	29/10/00 – 20/3/01
	Detector	IPD		Astrocam
	Cycle time	8.5 minutes		15.0 minutes
	Clear nights	43 7 8 95		34 0 0 53
Total	Clear nights:	105 22 14 249	48 10 2 133	92 6 4 185

Table 5.1 *FPI data availability for each site: dates data were collected; the detector; the typical cycle times of the FPI; and the numbers of clear nights for geomagnetically quiet, moderate, and active nights, and the total clear nights.*

Table 5.1 shows the dates where data were being collected, for KEOPS, Sodankylä and Svalbard FPIs, along with the number of clear nights where data are available, and the proportion of these with quiet, moderate or active geomagnetic activity levels. The total clear nights is more than just the sum of the quiet, moderate, and active nights as some nights have levels that vary throughout the night, as discussed above. The total number of nights used, for each activity level, over all the observing seasons, are also given in Table 5.1. A total of 567 nights of data are used in this analysis.

The detectors that were on the FPIs each season are also listed. These were discussed in section 2.2.4. The new Andor detectors are the most sensitive and have high quantum efficiencies, and allow shorter integration times than the Astrocam CCDs, with chips from e2v Technologies. The increase in the time taken for a cycle of look directions decreased at Sodankylä between the 2003 – 2004 and the 2004 – 2005 seasons, and at Svalbard between the 2002 – 2003 and the 2003 – 2004 seasons. This was due to the Astrocams being upgraded with new chips, with higher sensitivities. The integration time used is not only affected by the efficiency of the detector, but also of the typical activity levels. The normal exposure times used may be lengthened or shortened on any particular night depending on circumstances. The typical total cycle times are also given (to the nearest 0.5 minutes) for each site and season, and are dependant on the exposure length, the time taken to move the mirror between look directions, and the choice of directions viewed in each cycle.

5.3 RESULTS AND DISCUSSION

To view the data clearly for the many nights in each observing season, this section shows histograms of the occurrence of gravity waves for different periods. To plot data as histograms, the periods need to be binned. Bin sizes of 0.2 hours (12 minutes) are used. Periods up to 8 hours are shown. Plots are therefore of the number of wave periods found in each period bin, and the tolerance on the periods is 0.2 hours.

Data are shown from all three sites: KEOPS, Sodankylä, and Svalbard, in the next three sections (5.3.1 – 5.3.3), for all the data available from 2000 – 2006. Section 5.3.4 shows the data for these three sites summed up over all the years of data presented in the first three sections. This has the advantage of increasing the number of nights of data that are analysed, which increases the statistical sample.

The figures in this section show the peak periods in histograms for the intensities, winds, and temperatures. The various Kp levels, as described above, are plotted together in separate colours. One count is shown for each period bin in a night, i.e. if there is a peak at any bin period in any of the look directions, then it is counted once, rather than a separate count for each look direction. There is a general point to note about this type of plot. There is a bias to the periods found due to the way the data are collected. No very short periods will be seen, below twice the time resolution of the data, and there is a limiting value on the periods of the gravity waves due to the Brunt-Väisälä frequency (approximately 12 minutes). However, once above a few tens of minutes, all the periods should be detected equally. There is a bias at the longer periods due to the maximum period observable being half the lengths of the data sets. Therefore, only waves up to this maximum limit have been included in these plots. Only the mid-winter months will have enough hours of darkness to include waves with periods up to 8 hours, while nights at the beginnings and ends of the observing seasons will have hours of darkness as short as four or five hours, depending on when the instruments are closed down for the summer. This will increase the proportion of shorter periods over the longer ones. In addition, as the shorter waves could have more wavefronts detected within the length of the night, and the longest may only see one cycle, the confidence of the detection will be greater for the short periods where the spectral power is reinforced by successive waves.

5.3.1 KEOPS DATA

Data from the 2000 – 2001 season is shown in Figure 5.1. This period was before the Sodankylä FPI had been installed for the tristatic FPI configuration, so the additional look directions, other than the cardinal and zenith directions, were to the north-east and north-west rather than tristatic A and bistatic B. These were chosen to give better coverage of the auroral oval, which is typically overhead or to the north of Kiruna. This season was good both in terms of detector performance and cloud cover levels, consequently there were 95 nights of data over the observing season. The distributions of the waves across the period range show the decrease in numbers of the shortest and longest period waves due, respectively, to the time resolution of the data set and the limited length of the data sets for some of the nights.

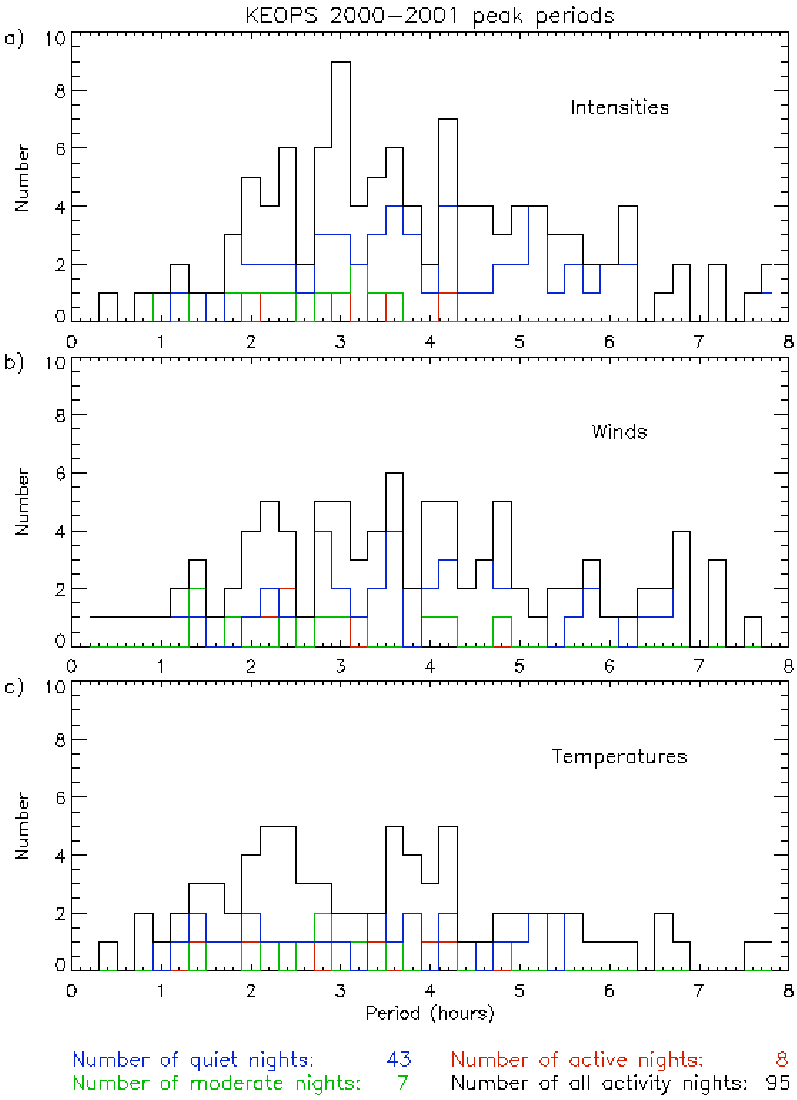

Figure 5.1 *Histogram of gravity wave periods for all nights and all activity levels, from KEOPS data over the winter season of 2000-2001: a) intensities, b) winds, and c) temperatures.*

As there were data available for over eight months on a continuous basis during the 2000 – 2001 season, this is the best data set to investigate seasonal dependences on the numbers of gravity waves detected. Figure 5.2 shows the periods of the peaks in the LSP power for each day, plotted against the day of year number, starting at the beginning of 2000 (i.e. 365 is the start of 2001). Intensity, wind, and temperature data are all plotted. The main feature of this plot is the limit of the periods of waves detected due to the length of the data set, which is determined by the number of hours of darkness varying through the winter. This approximate limit is shown by the solid line, which is half the length of the night, determined from civil twilight times (when the Sun is 6° below the horizon). This is the approximate time that the FPIs start collecting data, which is when the zenith angle is 98°. The small number of many waves with short periods (less than ~1 hour) due to the time resolution of the data is also apparent. The apparent increase in the number of waves on day numbers between 440 and 470 is due to the increased number of clear nights in March and April, rather than an increase in the number of waves present. For example, there were 18 clear nights in March, but only 6 in December.

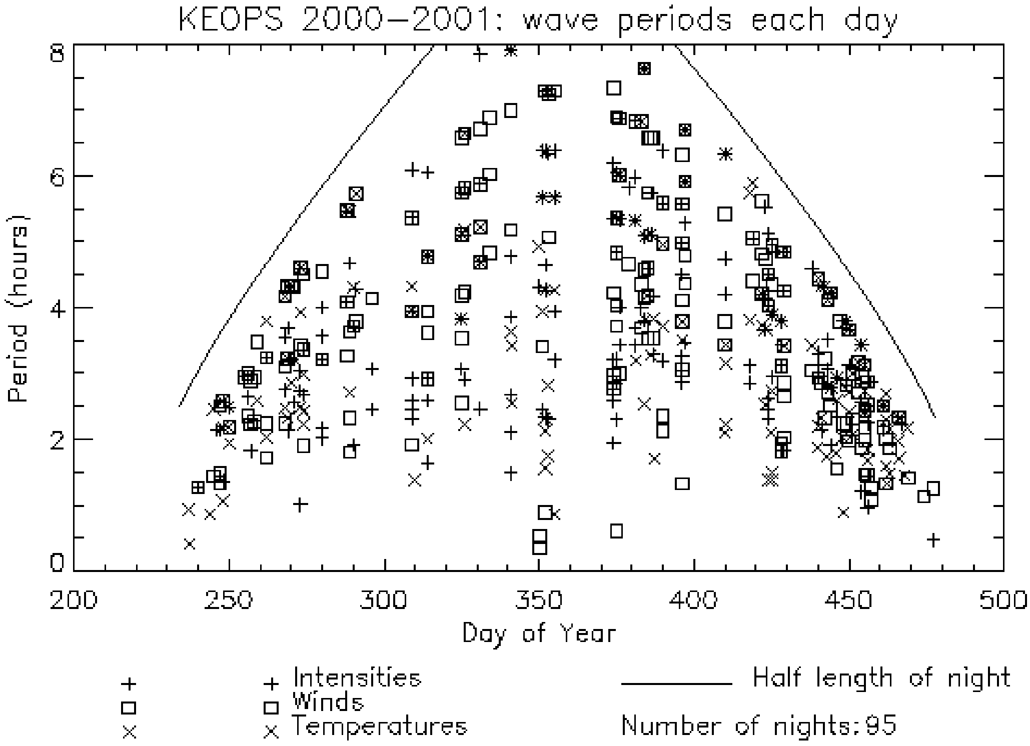

Figure 5.2 *Gravity wave peak periods for each night over 2000-2001, from all KEOPS intensities (plus symbols), winds (squares), and temperatures (crosses). Solid line is half the night length from civil twilight times.*

The peak periods in Figure 5.2 show a noticeable pattern, particularly at long periods in the middle of the winter, where the waves with periods greater than 4 hours appear in lines parallel to the solid line showing the length of the night. This can be seen more clearly in Figure 5.3 below, where the peak periods for each day of the year, in Figure 5.2, are divided by the length of the night for that day. In this plot, the symmetry of the season around the winter solstice has been used to increase the statistical sample. The days from the second half of the winter, after the winter solstice, have been plotted with the first half of the winter, effectively folding the year around the solstice (day of year 355). Horizontal lines are seen in the ratio of peak periods per night length for values longer than approximately 0.25, and after day of year number 260. This corresponds to the 16th September or day 85, the 25th March, so approximately the equinoxes. Peaks in the ratios are seen at roughly 0.27, 0.31, 0.34, 0.39, and 0.43. The error on this is less than 0.01 on the nights nearer the solstice, but approximately 0.02 towards the equinoxes. The differences in these are roughly 0.04±0.02, implying that the gravity wave periods are at harmonics of the length of the night.

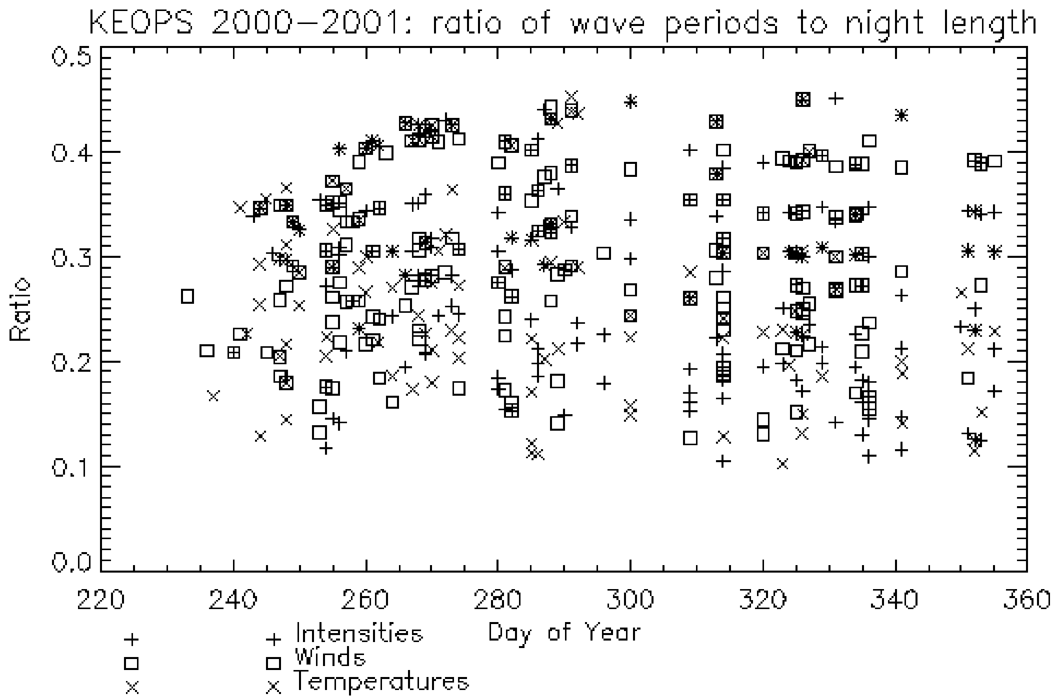

Figure 5.3 *Ratio of gravity wave peak periods to night length for KEOPS 2000-2001, intensities (plus symbols), winds (squares), and temperatures (crosses). The season has been folded around the winter solstice to improve statistics.*

An explanation for this effect could be due to the differences in electron densities and conductivities between the daytime and night time, as there are larger electron densities during the daytime due to photoionisation, and the terminator acts as a boundary between the two. Waves that have periods that are harmonics of the length of the night will survive, through constructive interference, whereas other wave periods will not. The length of the night will vary through the year, so the periods observed will vary too. Another reason for not seeing these harmonics for the nights past the equinoxes is probably due to the short length of the night, as only waves less than 4 hours can be detected, and for instance 30% (an approximate harmonic) of this is 1.2-hours. At the beginnings or ends of the season only 2-hour period waves can be detected, and 30% of this is 0.6 hours. At these short periods, the time resolution of the data becomes important, as is discussed further with other data sets below.

To better show the distribution of the waves across the year, the 2000 – 2001 data are grouped together in months in Figure 5.4. Only waves with periods up to 4 hours are included, as otherwise the varying length of the night creates a bias towards the longer winter nights, as can be seen from Figure 5.2. The dotted line shows the number of nights of data that were used for each month. This shows that there were an increased number of clear nights in September, March, and April compared with the other months. The solid line shows the number of gravity waves detected in each month, over 4 hours, and above the 70% confidence level, divided by the number of nights used in that month. An increased number of waves are seen in March where the equinox occurs, in the intensities and winds. The vernal equinox in September sees an increase in the number of waves in the winds and temperatures, but this is shifted to November in the intensities. Nights in August and much of April are too short to detect many waves. The number of waves seen within these limits are too low to draw any statistical conclusions, however an increase in the numbers of waves around the equinoxes is possible.

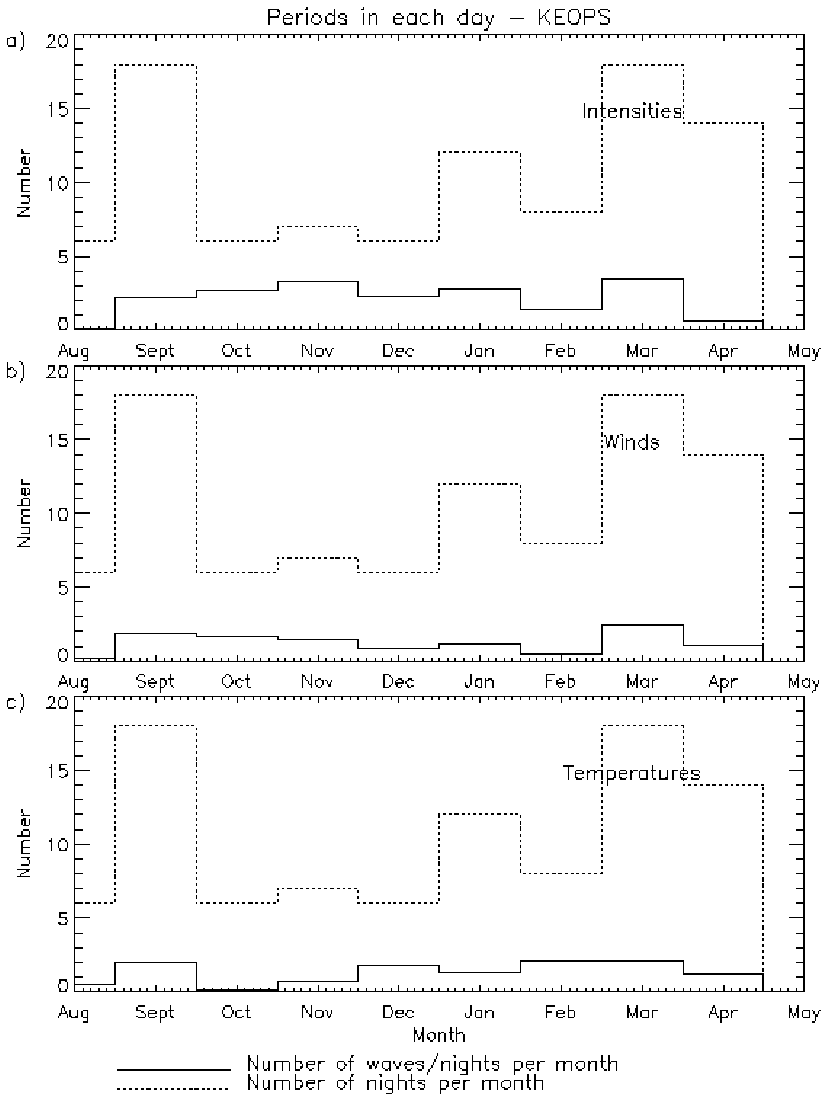

Figure 5.4 *Monthly counts of gravity waves per number of clear nights, with periods from 0 – 4 hours, from KEOPS during 2000-2001: a) intensities, b) winds, and c) temperatures.*

At the beginning of the 2001 – 2002 season, the IPD detector that was operating on the KEOPS FPI began to fail. Data were collected, but due to the decrease in sensitivity of the failing IPD, integration times had to be increased, so the resolution of the data was as poor as half an hour. This is not sufficient to sample many of the waves seen in the other seasons data, so these data are not used in this study. The detector failed completely by the end of November, so it was replaced with a CCD from a green line instrument by the 20th February 2002. There are consequently only three months data for this observing season, of which only 23 nights were clear. This season was also before the tristatic campaigns, so the look directions included north-east and north-west rather than tristatic A and bistatic B.

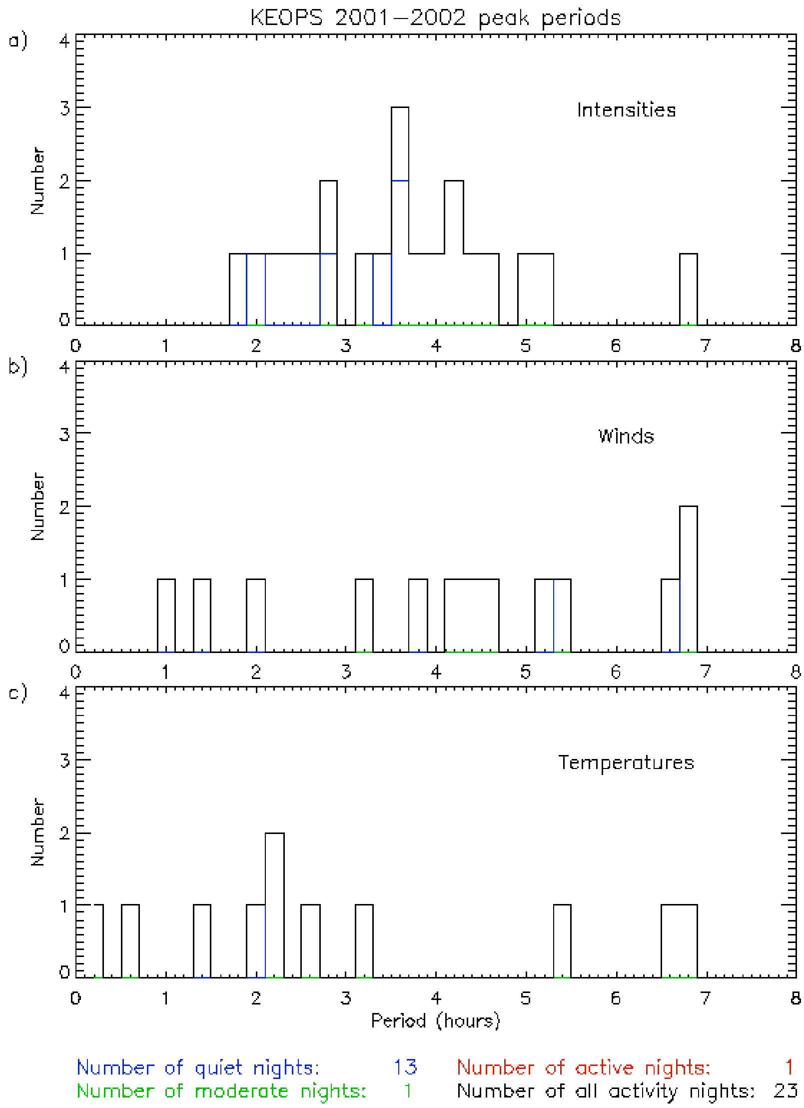

Figure 5.5 *Histogram of gravity wave periods for all nights and all activity levels, from KEOPS data over the winter season of 2001-2002: a) intensities, b) winds, and c) temperatures.*

With a total of only 23 clear nights of data the number of gravity waves in any one period bin are very small. In addition, many of these are geomagnetically quiet nights, which often have low 6300Å intensities, and if the intensities are too low, observed profiles will not be well enough defined to fit the parameters. This can result in gaps in the data, possibly for more than an hour, and this will affect the number of gravity waves that are detectable. Figure 5.5 shows the waves that are detected for this season. The winds and temperatures see waves distributed across the period range, but there is mostly only one wave in any one period bin. The intensities show more gravity waves and do show waves with the same period in more than one night for three period bins. All but one of these waves has periods in the range 1.7 – 5.3 hours. The statistical sample for this season is small, however, this data set had the same time resolution as Svalbard data in the same year, so comparisons can be made where time resolution does not affect the results. This is discussed in section 5.3.3.

Data from KEOPS from the 2002 – 2003 season are shown in Figure 5.6. Due to a detector failure at the beginning of the season, there were no data for the majority of the winter, and data were only collected from the beginning of February to the middle of April. This resulted in only 27 clear nights for this observing season. The first of the tristatic campaigns with the EISCAT radars was during this period on the 27th and 28th February 2003, discussed in Chapter 3. For this season onwards therefore, the additional look directions to the cardinal and zenith directions were the tristatic A and bistatic B positions, rather than the north-east and north-west.

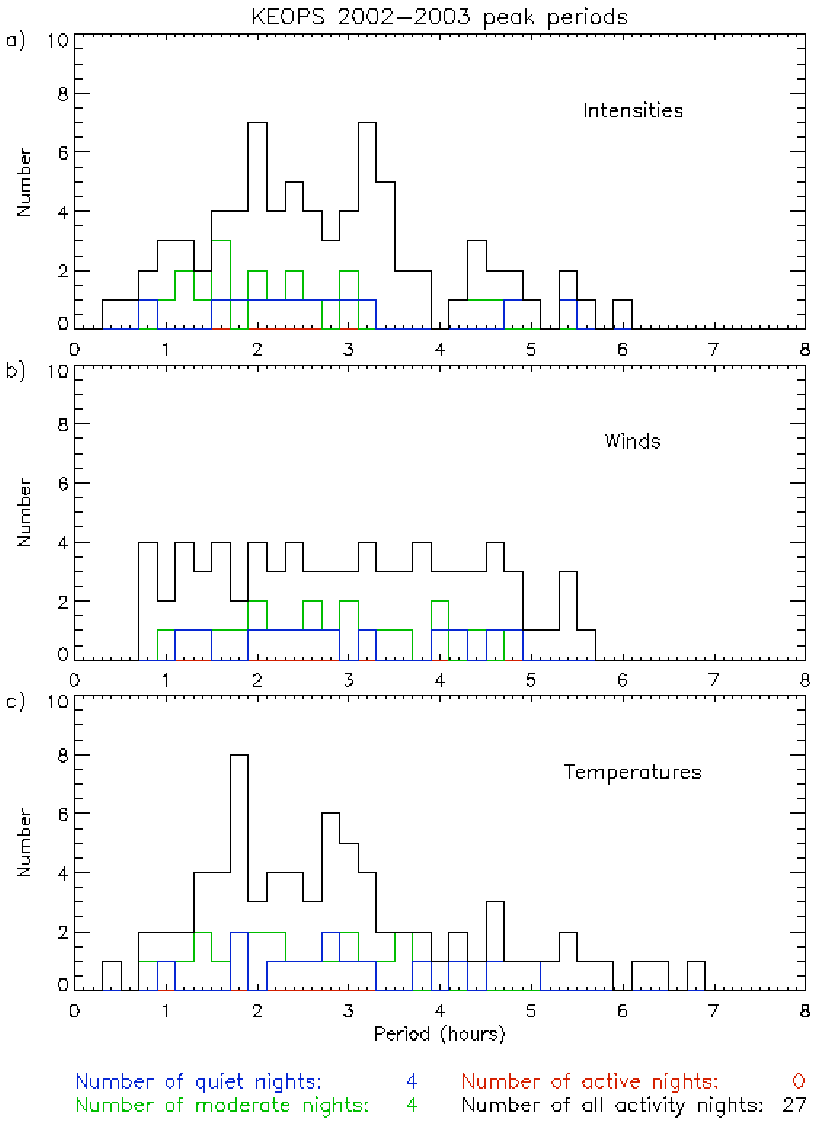

Figure 5.6 *Histogram of gravity wave periods for all nights and all activity levels, from KEOPS data over the winter season of 2002-2003: a) intensities, b) winds, and c) temperatures.*

Due to the limited number of long winter nights in this data set, there are few waves with long periods in all of the plots in Figure 5.6. The winds in Figure 5.6b show a fairly uniform distribution as with the previous data sets, but there are more shorter period waves in the intensities (Figure 5.6a), with a similar distribution for the temperatures (Figure 5.6c). The median periods for the intensities and winds are approximately an hour shorter than in the 2000 – 2001 data. This is likely to be due to the time resolution of the data, which was over twice as high in 2002 – 2003 with the Astrocam detector as the 2000 – 2001 season data, with the IPD. Longer integration times of the data, and hence longer cycle times around all of the look directions, results in lower time resolutions. This means the shorter period gravity waves will not be sufficiently sampled to create spectral powers over the 70% confidence level, and so they are not included in this analysis and these plots. This is discussed further below where data sets with similar time resolutions (KEOPS and Svalbard) are compared with Sodankylä data that has longer cycle times, for 2005 – 2006 data.

Figure 5.7 shows the 2003 – 2004 data, which again has relatively few clear nights of data. This is due to several reasons. It is in part due to large amounts of cloud cover, especially at the beginning of the season. The CCD failed in the middle of December, resulting in a loss of two months data, again in the middle of the season in the period with the longest hours of darkness. However, this detector was replaced with an Andor EMCCD in mid-February. The higher quantum efficiency of this detector allowed short integration times, which lead to the high – resolution study that was performed at the end of March and April of 2004, data from which were discussed in section 4.5. These data were, however, not used in this statistical study due to the poor time resolution of the look directions other than the tristatic A position, resulting in the loss of another month of data from the analysis. The distributions of the waves over period in Figure 5.7 are similar to those from the 2000 and 2004 seasons, with the majority of the waves in the intensity data having periods of 1.3 to 3.8 hours, the winds having a reasonably uniform distribution, and the temperatures having a higher tendency towards waves between approximately 1 and 4 hours.

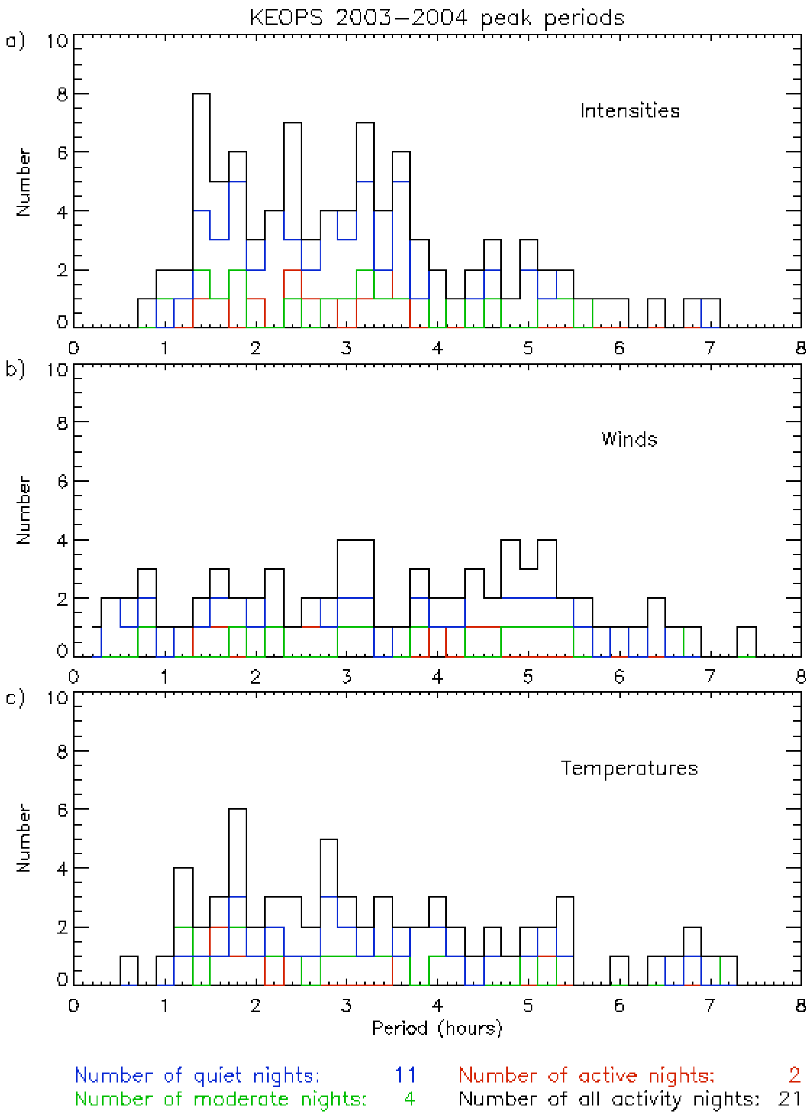

Figure 5.7 *Histogram of gravity wave periods for all nights and all activity levels, from KEOPS data over the winter season of 2003-2004: a) intensities, b) winds, and c) temperatures.*

Data from 2004 – 2005 from KEOPS are shown in Figure 5.8. In the intensities in Figure 5.8a, the increased proportion of waves with shorter periods can be clearly seen. The distribution of the peaks in the wind data (Figure 5.8b) is more evenly spread, and the majority of the waves have periods of 1 – 5 hours. The temperature data (Figure 5.8c) shows a similar shape to the winds, but with fewer counts. For this season, 71% of all the clear nights were during the periods when there were 16 or more hours of darkness. All of the geomagnetically quiet nights were during this period, so for these dates all the waves up to 8 hours should be detected.

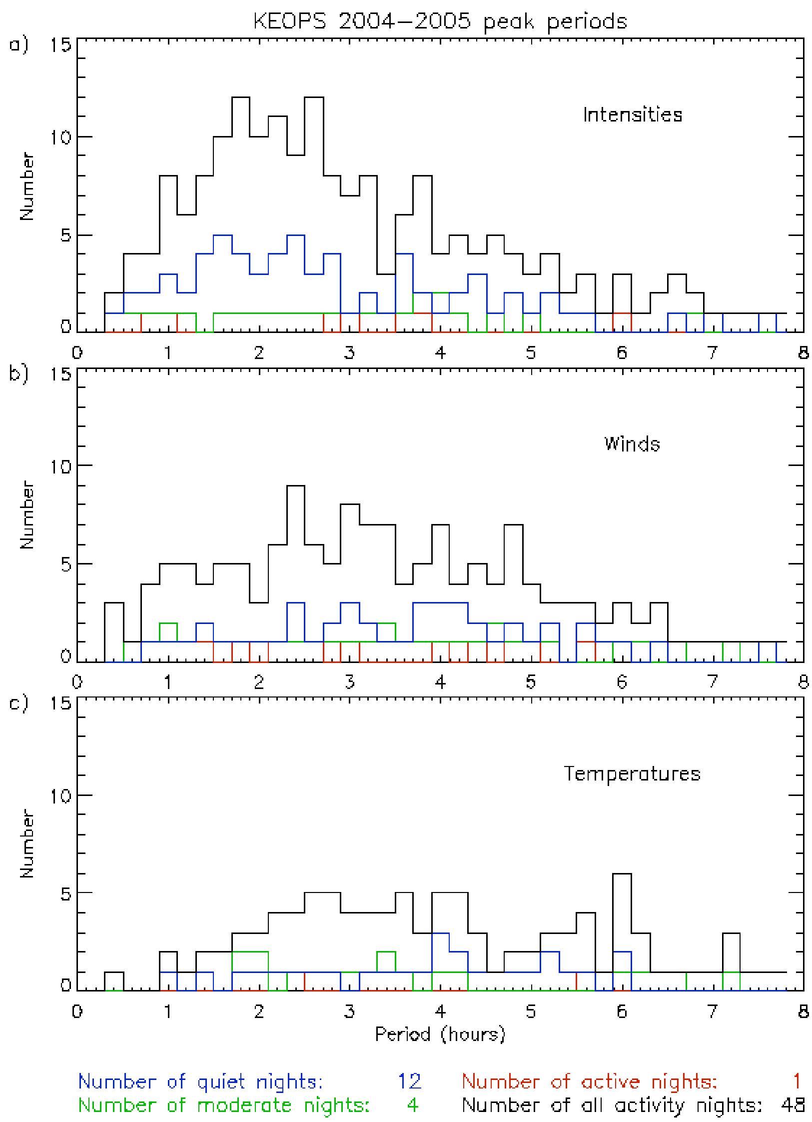

Figure 5.8 *Histogram of gravity wave periods for all nights and all activity levels, from KEOPS data over the winter season of 2004-2005: a) intensities, b) winds, and c) temperatures.*

There is a significantly larger number of waves seen in the intensities at shorter periods than longer, with the majority being between 1 hour and 3 hours. The lower limit on this may be due to the time resolution of the data, but periods longer than 3 hours are easily detectable, so this implies that longer period waves are not as common as short period waves. In comparison, the peaks in the winds and temperatures are more evenly spread across the period range, possibly with an increase in the numbers of waves with periods of a few hours, of around 2 – 5 hours. The intensities are quicker to react to changes in conditions, such as gravity wave forcing from particle precipitation, than the winds and temperatures, due to the inertia of the thermosphere. This would explain the difference in the distributions of the periods in the different parameters, as the short period waves are not able to form in the winds and temperatures, and the inertia of the thermosphere in effect acts as a low band pass filter.

Figure 5.9 for KEOPS 2004 – 2005 season shows one count for each look direction, so shows a number greater than one if a wave is seen in multiple look directions in a night. This can be used to show an indication of the proportion of nights where the gravity waves are seen in multiple look directions. However, this will exaggerate the apparent number of gravity waves, so is only used for this study. Only the north, east, south, and west data are used, as if all the directions are included there will be a bias as they are not uniformly distributed over the region. The distribution of the waves with period has a similar shape to that of Figure 5.8 where all look directions count as one per night. In addition to this shape though, there are an increased number of wave peak periods at 6.0 hours in the intensities and at 5.6 and 6 hours in the temperatures. The peak at 5.6 hours is due to the wave in the active night being observed in all the look directions. The 6-hour wave, in intensities and temperatures is likely to be due to atmospheric tides, caused by solar heating, at a harmonic of the 24-hour day, or to a harmonic of the two-cell convection pattern. This tide is probably not observed in the wind data due to the interaction of the tide with the background wind fields, for example from the two cell convection pattern. The tide would be expected to be stronger in the temperatures as it is caused by solar heating.

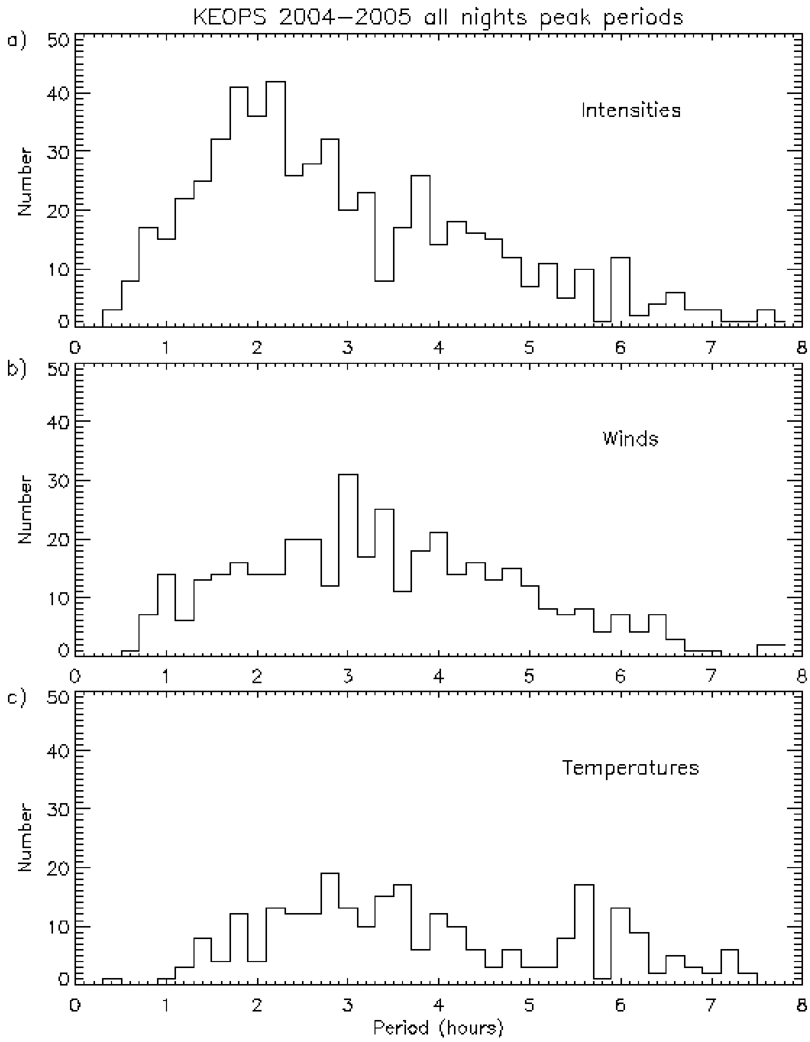

Figure 5.9 *Histogram of gravity wave periods for all nights for all look directions individually, from KEOPS data from the north, east, south, and west directions, over the winter season of 2004-2005: a) intensities, b) winds, and c) temperatures.*

Comparisons in the proportion of counts over the periods of waves between the histograms in Figure 5.9 and those in Figure 5.8 will show an indication of the proportion of nights where the gravity waves are seen in multiple look directions. The general shape of the histogram in Figure 5.9 and the total nights (black line in Figure 5.8) are the same, though the count rate is approximately 25% of the plot counting each look direction separately. By finding the ratio of these two plots, the average number of look directions in which a wave is observed can be found. This is plotted in Figure 5.10.

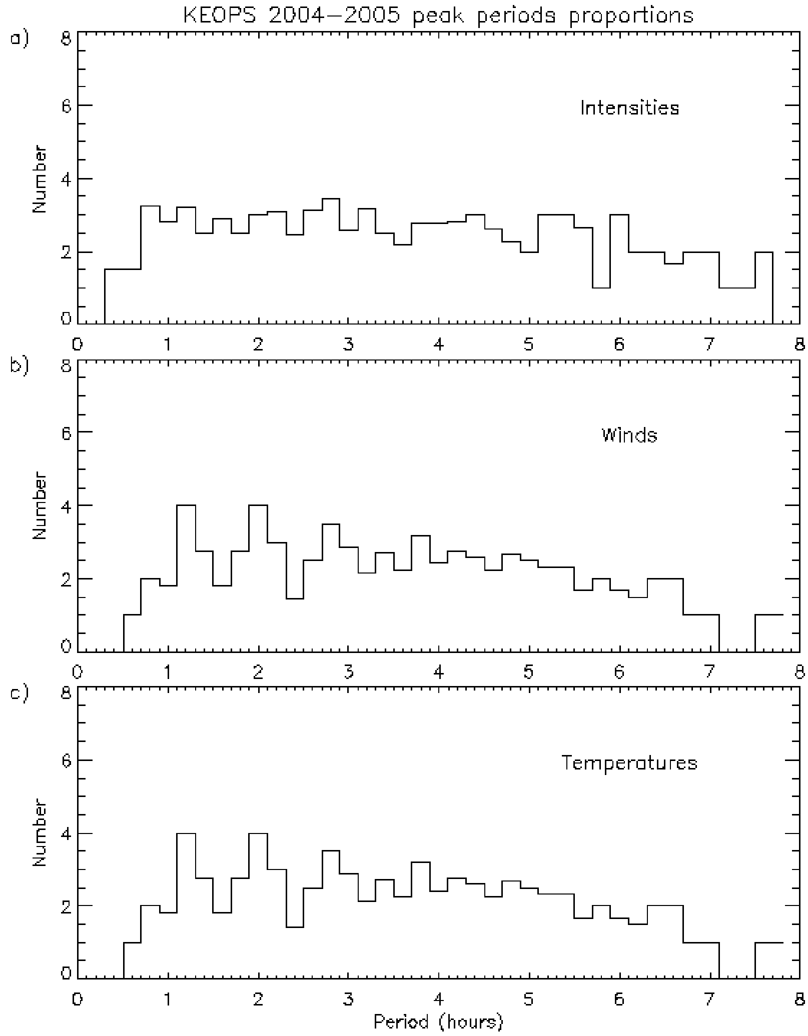

Figure 5.10 *Proportion of gravity wave periods, for the north, east, south, and west directions, compared with the number per night, from KEOPS data over the 2004-2005 season: a) intensities, b) winds, and c) temperatures.*

Apart from the decreases at very short and long periods, the distributions in Figure 5.10 are fairly uniform. The average, from periods between 0.6 and 7.0 hours are 2.6 for the intensities, 2.3 for the winds, and 2.2 for the temperatures. The standard errors on the means (□= □/ N, where □ is the standard deviation over N data points) are 0.10, 0.13, and 0.17 respectively, so these values are the same within errors. This means that, on average, a wave is seen at a confidence level above 70% in over half of the four look directions (on average 60%). From comparing with the case studies on gravity waves in Chapter 4, many of the strong waves that are seen with high confidence levels in one or two of the look directions, the waves are often detected in the other look directions but with smaller powers and lower confidence levels. If lower confidence levels were therefore to be considered, this number could be greater.

There are two possibilities for the waves not often being detected in all the look directions, with high powers. This could be an indication that the waves are only propagating across part of the field of view of the instrument, for instance along or across the auroral oval. Alternatively, the waves are dissipating before reaching all the look direction locations. To be able to detect the wave in one look direction with a high confidence level, the wave either needs to have a large amplitude over the background variations, or several wavelengths need to be observed. If the amplitude is large in one look direction, the wave is unlikely to have dissipated by reaching another look direction, which will only be a fraction of a wavelength away. Similarly, if several wavelengths are observed in one direction, the source is likely to be nearby and the wave should not dissipate over the distance covered by the field of view. It is therefore more likely that the waves are not seen in some of the look directions due to the propagation angle of the wave. This could be an indication that the waves do not propagate uniformly away from the source, but have a preferred direction. From the results of the case studies in Chapter 4, this is likely to be in an equatorward or poleward direction, perpendicular to the auroral oval.

Data from KEOPS from the most recent season, 2005 – 2006 are shown in Figure 5.11 below. Despite data being collected for seven months, there are relatively few nights of data available for this analysis. This is mostly due to poor weather conditions, as much of the winter was cloudy. This can also be seen in the Sodankylä and Svalbard data. The distributions of the number of waves with period again follow similar patterns to the other data sets, with an increase in the number of short period waves in the intensities, which is not seen in the other parameters. This is more pronounced for this data set, with the greatest number of waves between 0.7 and 2 hours.

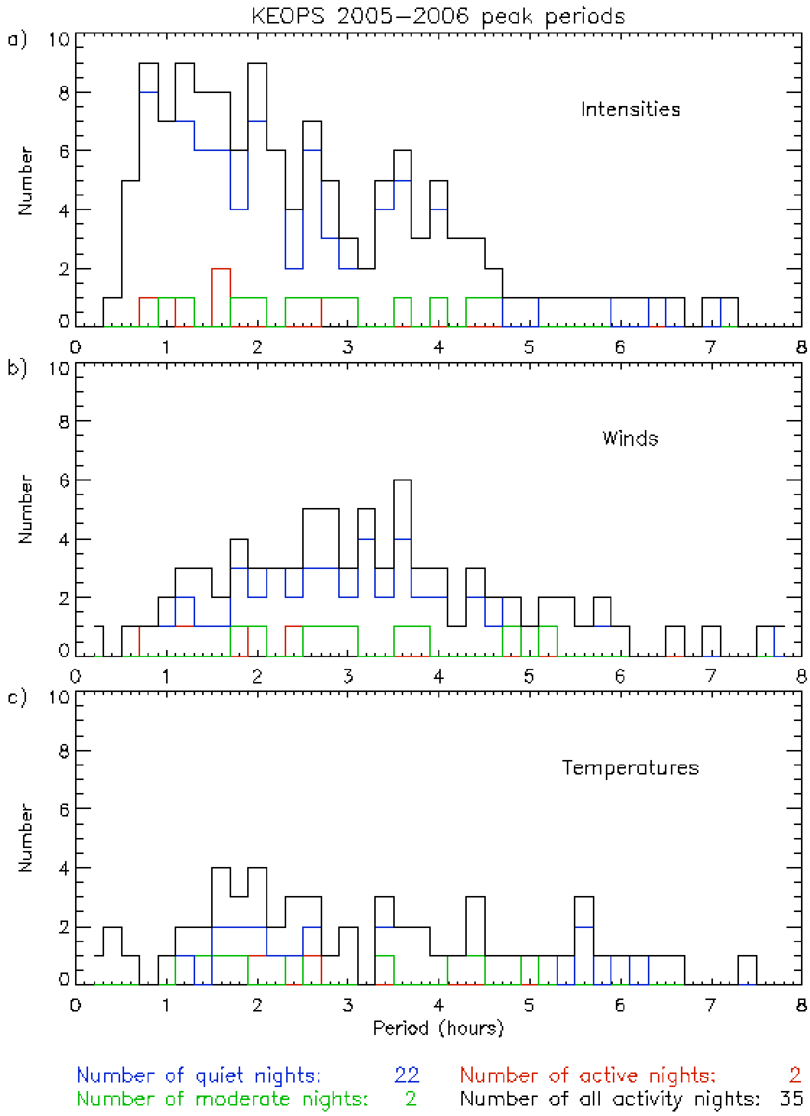

Figure 5.11 *Histogram of gravity wave periods for all nights and all activity levels, from KEOPS data over the winter season of 2005-2006: a) intensities, b) winds, and c) temperatures.*

Comparisons between the data in, for instance, Figure 5.11 from 2005 – 2006 and Figure 5.1 from 2000 – 2001 should show any solar cycle dependences. In the most recent years, the solar cycle has been approaching its minimum (predicted as 2007), and this can be seen from the increased number of geomagnetically quiet nights compared with the total in Figure 5.11. This effect can also be seen in the Sodankylä and Svalbard data that are shown later. However, despite there being more nights with moderate and high geomagnetic activity levels in the 2000 – 2001 data, there are no more gravity waves at any period in Figure 5.1 than in Figure 5.11. There is still generally only one wave in any one period bin, and at most two. Although the numbers of active nights are low so the results are inconclusive, it appears that geomagnetic activity levels has little effect on the gravity waves produced. This would be in agreement with, for example, Ogawa et al., (1987), who proposed that gravity waves are relatively easily formed in the thermosphere and large amounts of precipitation or energy input are not required to produce the gravity waves. This was however from ionospheric data from a range of lower thermospheric altitudes, so may not be applicable to the FPI data. This is investigated further in section 5.3.4, where the data from all the years together provide a larger statistical sample.

In summary, the time resolution of the data is the most important factor in the number of short period waves detected in the FPI data. The length of the night will limit the number of long period waves that are detected. However, by looking at a data set (2004 – 2005 geomagnetically quiet nights), where all the nights were longer than 16 hours, shows that there are fewer waves with periods greater than approximately 3 to 4 hours, implying gravity waves have preferred periods in the intensities shorter than 4 hours. For the winds and temperature data, the distributions with period are more even, and longer periods up to around 5 hours are more common. The differences are due to the response times of the parameters to particle precipitation.

Differences in the gravity wave properties due to geomagnetic activity levels cannot be easily determined from one season of data due to the low numbers of active nights, therefore several years of data will be used to determine this (section 5.3.4). Ratios of the number of waves detected to the number of nights in an observing season shows that gravity waves are seen on average in approximately half of the look directions, due either to the propagation direction of the wave or its dissipation rate. Seasonal variations were studied by counting the number of waves detected in each month of the year. This did show an increase in wave activity in the winter months around the solstice, decreasing towards the ends of the observable seasons at the equinoxes but this is due to the limited length of the night at the equinoxes. Comparisons of data sets at solar maximum and solar minimum do not show any change in the number of waves produced, therefore solar activity seems to have little effect on the production of gravity waves, possibly because gravity waves are relatively easily formed in the thermosphere and do not require large geomagnetic activity.

5.3.2 SODANKYLÄ DATA

Figure 5.12 below shows the histogram from the first season of observations from the Sodankylä FPI in 2002 – 2003. There are relatively few nights of data here compared with the number of clear nights over the period of data collection. This is mostly due to the intensities being too low to fit a temperature profile to the interference fringes. Only nights when temperature as well as winds and intensities are available are used in this study. The reason for frequently obtaining too low intensities is likely to be due to a too short an integration time used in collecting the data for the CCD that was used.

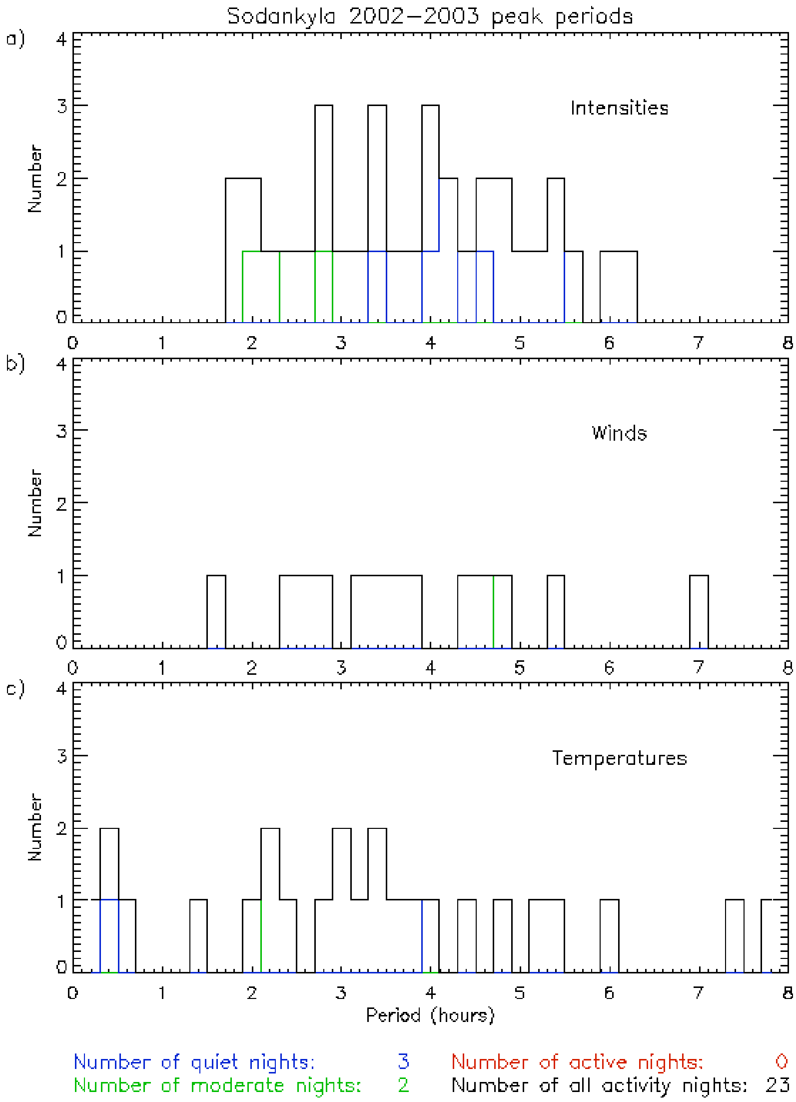

Figure 5.12 *Histogram of gravity wave periods for all nights and all activity levels, from Sodankylä data over the winter season of 2002-2003: a) intensities, b) winds, and c) temperatures.*

However, even for the 23 clear nights in the 2002 – 2003 season (Figure 5.12), there are relatively few waves counted. There are approximately half the number of nights than in the KEOPS 2004 – 2005 data set (Figure 5.8) but there are much less than half the numbers of peak periods detected. This data set has a maximum of 12 peaks at 1.8 and 2.5 hours in the intensities, whereas in Figure 5.12a, there are a maximum of only three counts in any one period bin, where approximately six would be expected for the size of the data set. This is even more apparent in Figure 5.13

below, which shows the histograms for the following season (2003 – 2004), and has more clear nights (57) but still has a maximum of only four peak counts in any one period bin, for the intensities.

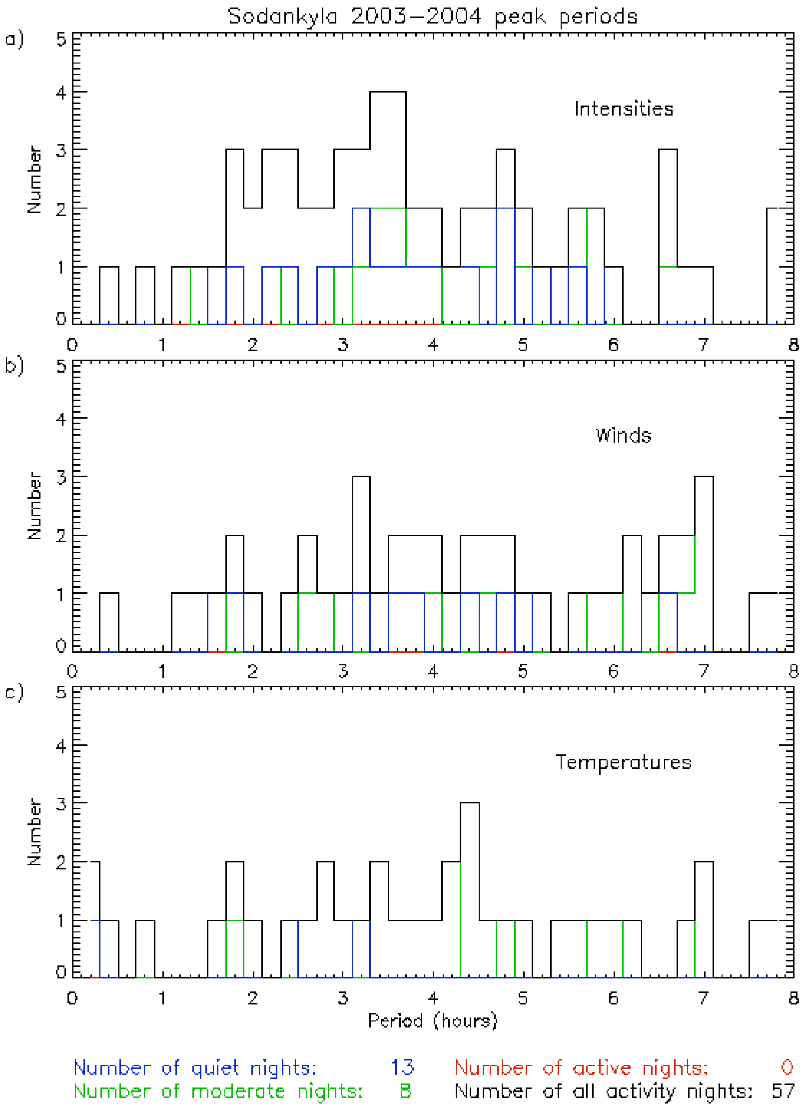

Figure 5.13 *Histogram of gravity wave periods for all nights and all activity levels, from Sodankylä data over the winter season of 2003-2004: a) intensities, b) winds, and c) temperatures.*

The most likely reason for this difference in numbers of waves detected is the time resolution of the data. It can be seen from Table 5.1 that the average resolution of the Sodankylä data for the 2002 – 2003 and 2003 – 2004 seasons were 12.5 and 11.0

minutes respectively, whereas for KEOPS data for these years, data were taken every 3.5 minutes. KEOPS therefore has more than three times better time resolution, so will sample oscillations at more points along the wavelength, resulting in a more confident detection of the wave, and a higher spectral power.

The data from the 2004 – 2005 season (Figure 5.14) had a higher time resolution of 6.5 minutes, but there were few clear nights to analyse. The instrument was running for a considerable period, however there were large amounts of cloud cover throughout the observing season, and this combined with software faults for the last month of the season, resulted in only 12 clear nights of data. However, even with this small data set, waves are seen in all parameters with a range of periods. The previous year's data from 2003 – 2004 in Figure 5.13 have a larger statistical sample, and show a similar distribution of waves at different periods as that of the KEOPS data in Figure 5.8, for the intensities and the winds. The peak in the numbers of waves, i.e. the median period, are shifted to longer periods in the Sodankylä intensity data (Figure 5.13a), with the mode period between approximately 2 and 4 hours, whereas the mode for KEOPS waves were between 1 and 3 hours. This is a bias in the measurements of the waves due to the poorer time resolution of the Sodankylä data, rather than any difference in the waves present at the two sites. This shift is also true of the winds, but there are insufficient numbers of waves in the temperature data for the distribution to be determined.

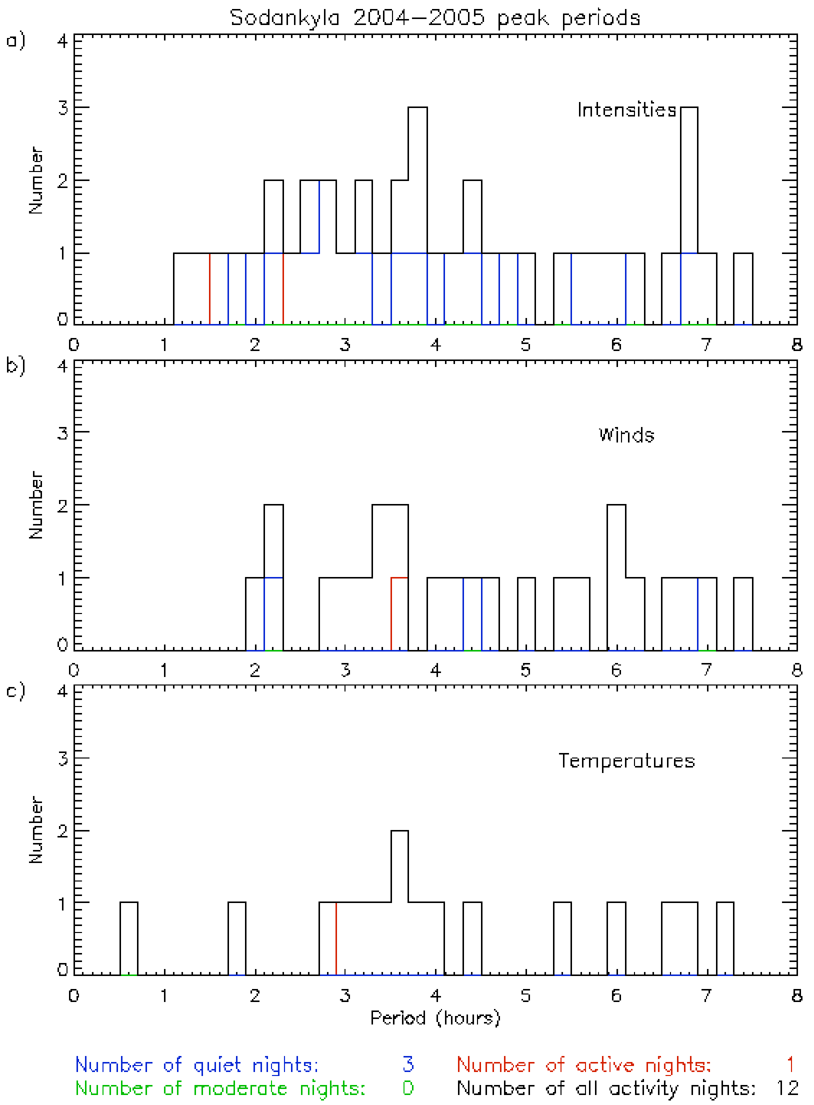

Figure 5.14 *Histogram of gravity wave periods for all nights and all activity levels, from Sodankylä data over the winter season of 2004-2005: a) intensities, b) winds, and c) temperatures.*

Figure 5.15 shows the histogram for the 2005 – 2006 Sodankylä data. The proximity to solar minimum can again clearly be seen by the large proportion of geomagnetically quiet nights, which is over 70% of the total clear nights for this season. This plot can be compared with Figure 5.11, which shows the KEOPS data for this same season. There are a comparable number of clear nights at each site, and the main difference between the two is the time resolution of the data sets. The KEOPS data at a cycle time of 3.5 minutes, but Sodankylä had over twice this at 7.5 minutes resolution.

This difference can be seen in the plots in the intensities, as the Sodankylä data (Figure 5.15a) has fewer waves at periods less than 2 hours than the KEOPS data (Figure 5.11a), due to the lower time resolution leading to reduced sampling of the waves. The numbers of waves with periods greater than 3 hours are comparable in each plot.

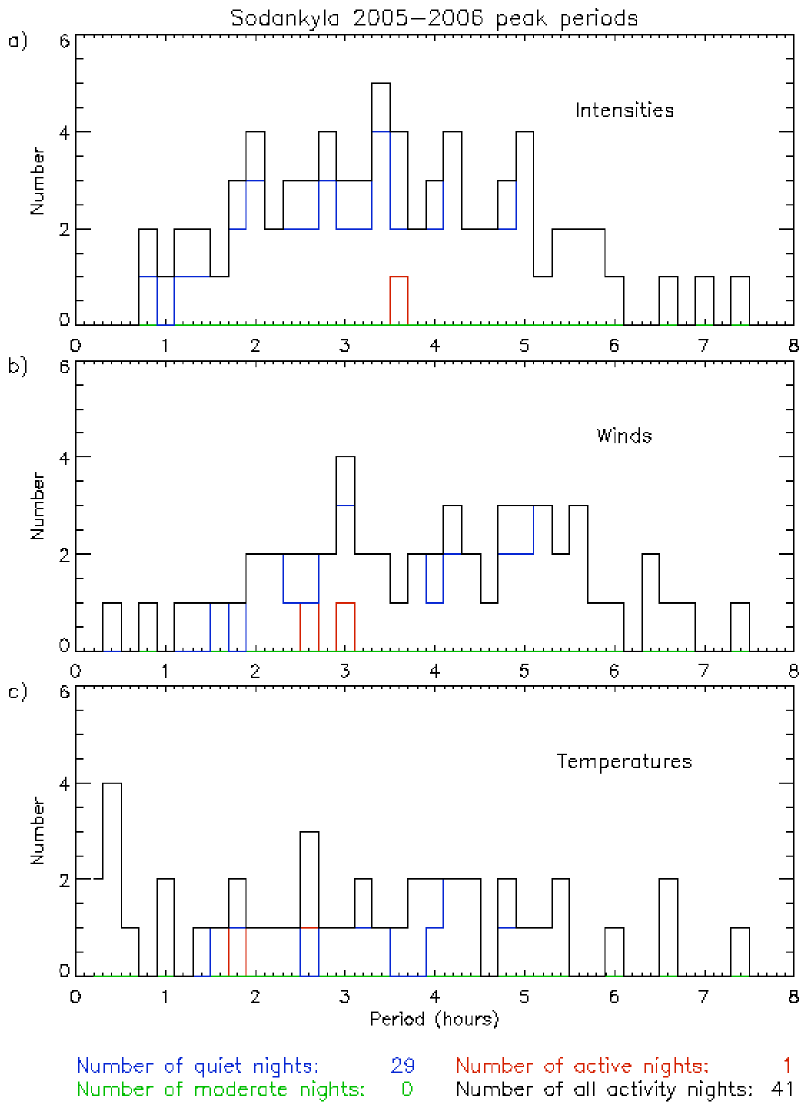

Figure 5.15 *Histogram of gravity wave periods for all nights and all activity levels, from Sodankylä data over the winter season of 2005-2006: a) intensities, b) winds, and c) temperatures.*

The winds from these two plots also show this time resolution effect, as there are fewer waves in the Sodankylä data (Figure 5.15b) than the KEOPS data (Figure

5.11b) with periods shorter than 4 hours. This is generally true of the temperatures too, with the shape of the distribution in Figure 5.11c not seen in Figure 5.15c. There are, however, four nights with waves of periods between 20 and 30 minutes in the Sodankylä data, which is greater than the number detected in the KEOPS data. However, for all of these waves, their power is only just over the 70% confidence level for the night, and they are only just over the minimum period detectable with the time resolution of the data, so they could be due to aliasing. The longer period waves, however, have similar distributions in the Sodankylä and KEOPS data for both the winds and the temperatures.

5.3.3 Svalbard Data

An FPI was installed in Svalbard prior to 2000. Good quantities of data were collected in the 2000 and 2001 seasons, however the detector broke down in December 2002, resulting in no additional data being collected that season, due to the difficulties in funding and the logistics of repairing or replacing a detector. A failure in the mirror rotation mechanism in December 2003 resulted in no data being collected for the majority of the 2003 – 2004 winter. A second detector breakdown at the beginning of the 2004 season resulted in no data being collected for all of the winter of 2004 – 2005. A new Andor detector was installed for the 2005 – 2006 season. A histogram of the wave periods found from the first season of data, in 2000 – 2001, are shown in Figure 5.16 below.

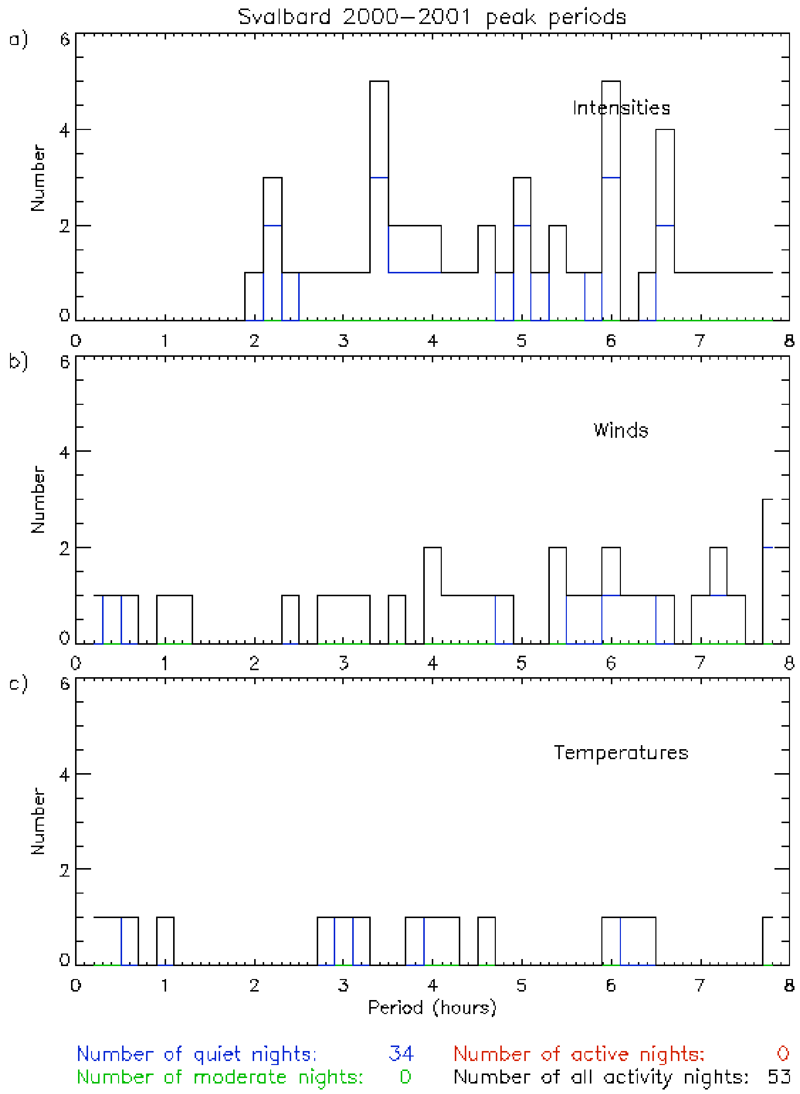

Figure 5.16 *Histogram of gravity wave periods for all nights and all activity levels, from Svalbard data over the winter season of 2000-2001: a) intensities, b) winds, and c) temperatures.*

There is a similarity in the difference between the plots from Svalbard and KEOPS to those between Sodankylä and KEOPS, in the number of gravity waves seen, over all periods. The Svalbard FPI detects fewer waves over all periods than the KEOPS data. The 2000 – 2001 season in Svalbard produced 53 clear nights with data and still produced a maximum of only five nights at any one period bin (see Figure 5.16), compared to, for example, 12 in KEOPS data in 2004 – 2005, which had only 48 clear nights with data in total (Figure 5.8). This is likely again to be due to the differences in time resolutions of the data sets, as exposures on the same look direction were as long as 15 minutes apart for the Svalbard data.

The Svalbard and KEOPS FPIs produced data with similar time resolutions in the 2001 – 2002 season, of approximately 15 minutes. Comparisons between these two data sets will therefore show differences that are not dependant on the time resolution. As they are also from the same year, there will be no solar cycle dependences, so any differences should be due to the differences in site. Svalbard in the polar cap might be expected to show a different distribution of waves to sites that are mostly under the auroral oval.

Figure 5.17 shows the Svalbard data from this 2001 – 2002 season. The main difference that can be seen between Svalbard and the mainland site is in the distribution of periods that are detected. For example, for the intensities from KEOPS in 2001 – 2002 (Figure 5.5) the majority of the waves had periods between 1.8 and 5.3 hours, but for Svalbard in this season (Figure 5.17) the greatest numbers of waves were between 2.8 and 8 hours. There are as many waves at short periods in each site, and the main difference is in the greater number of long period waves that are seen at Svalbard. Unfortunately, the detector on the KEOPS FPI failed at the end of November and could not be replaced until the end of January, so no data were collected during the longest winter nights. In contrast, Svalbard data were collected from October through to March, so all the polar nights data were used. This would have resulted in more long period data sets at Svalbard, and consequently an increased ability to detect the longer period gravity waves. The winds and the temperatures are reasonably evenly distributed across the period range in both sites, mostly due to the few waves detected in these parameters. There were only 23 clear nights at KEOPS this season, which produced few waves, whereas Svalbard had over 3 times the number of clear nights of data.

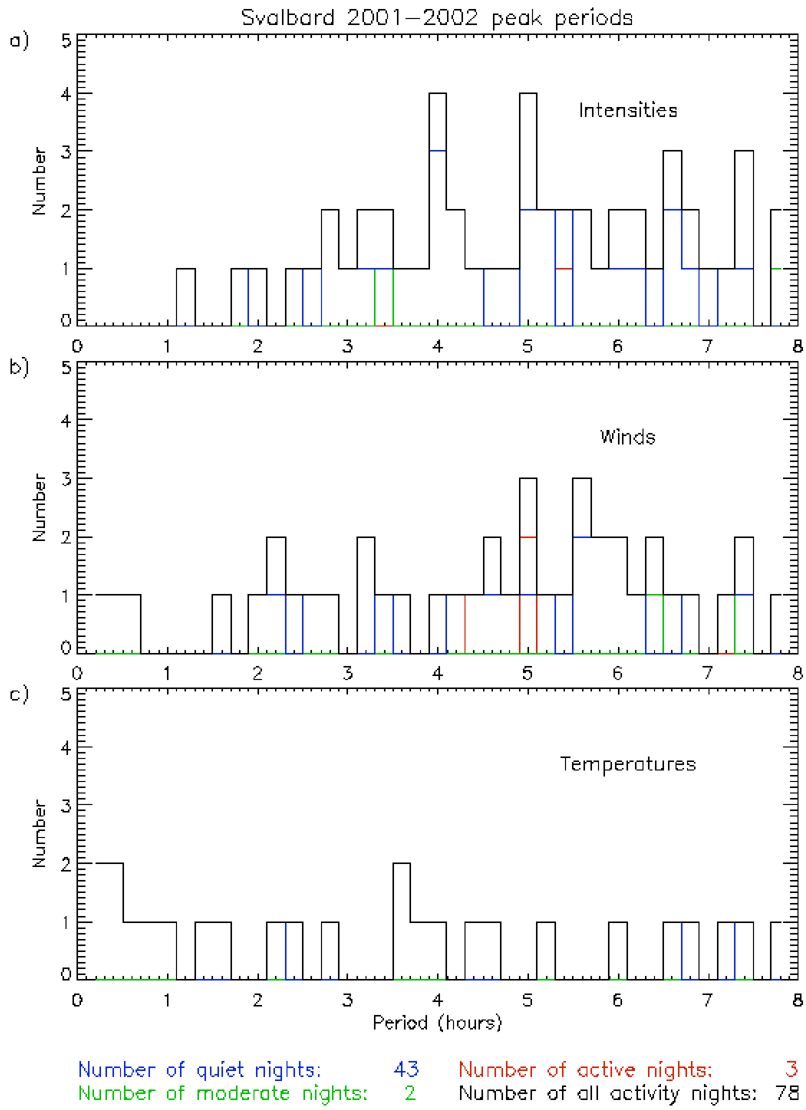

Figure 5.17 *Histogram of gravity wave periods for all nights and all activity levels, from Svalbard data over the winter season of 2001-2002: a) intensities, b) winds, and c) temperatures.*

Due to the failure of the CCD in December 2002, there is little data from the 2002 – 2003 season, with only 21 clear nights. Consequently, few waves are detected, and with the small statistical sample, their periods are fairly uniformly scattered over the period range, with mostly one and at most two waves in any one period bin. There certainly is not a sufficient statistical sample in each of the geomagnetic activity levels to analyse Kp dependences, therefore the data are not shown here.

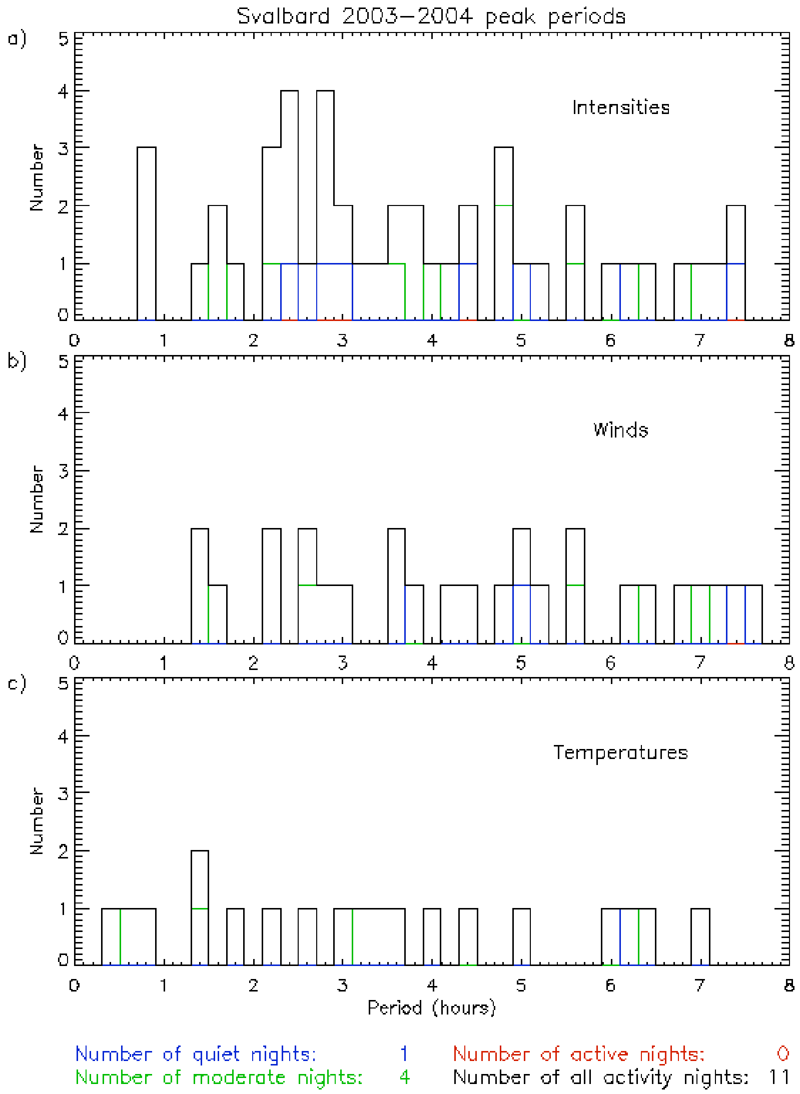

Figure 5.18 *Histogram of gravity wave periods for all nights and all activity levels, from Svalbard data over the winter season of 2003-2004: a) intensities, b) winds, and c) temperatures.*

Figure 5.18 shows the periods of waves from the 2003 – 2004 season in Svalbard. Data were first collected on the 30th November, and the mirror mechanism failed on the 19th December, allowing less than three weeks of data to be collected. Only 11 of these nights were clear. However, proportionally more waves were detected than in the 2002 – 2003 season, which had nearly twice as many clear nights. This is most likely due to the periods of data collection. The 2002 data period was from the 28th September to the 4th December. In Svalbard high in the Arctic Circle, there are only a few hours of darkness in September, but there is 24-hour darkness in November and December. In 2002, 8 of the 21 clear nights were before the first of the Polar nights, when the Sun no longer rises, on the 27th October. The 2003 nights were all in December, when there were sufficient hours of data to observe all waves in the period range used here. Figure 5.18 shows an even spread of periods for the winds and temperatures, but there are more waves in the intensities at periods of 2 to 3 hours, than at longer periods. Again, there are too few nights in the separate activity levels to draw any conclusions on activity dependence from this data set.

Data for 2004 –2005 do not exist due to the detector failure; however, Figure 5.19 shows the 2005 –2006 season after the Andor detector was installed at the beginning of December 2005. Cloud cover levels were considerable for this season, but despite the low number of clear nights available (22), and the proximity to solar minimum, a reasonable number of waves were detected, with up to 6 waves in any one period bin. This is most likely due to the improved time resolution of the data with the Andor detector, as cycle times of less than 4 minutes were achievable. The KEOPS FPI also has an Andor detector, and so has the same cycle time, therefore comparisons can be made between these two sites.

For the intensities at KEOPS in the 2005 – 2006 data (Figure 5.11), the majority of the waves have periods between 0.5 – 4.5 hours. However, for the Svalbard 2005 – 2006 data (Figure 5.19), the majority of the waves have periods between 2 and 5 hours. The maximum number of waves in any one period bin is nine for KEOPS but six for Svalbard, but this could be explained by the fewer number of nights of data (22 in Svalbard compared with 35 at KEOPS), as well as the broader spread of waves across the period range.

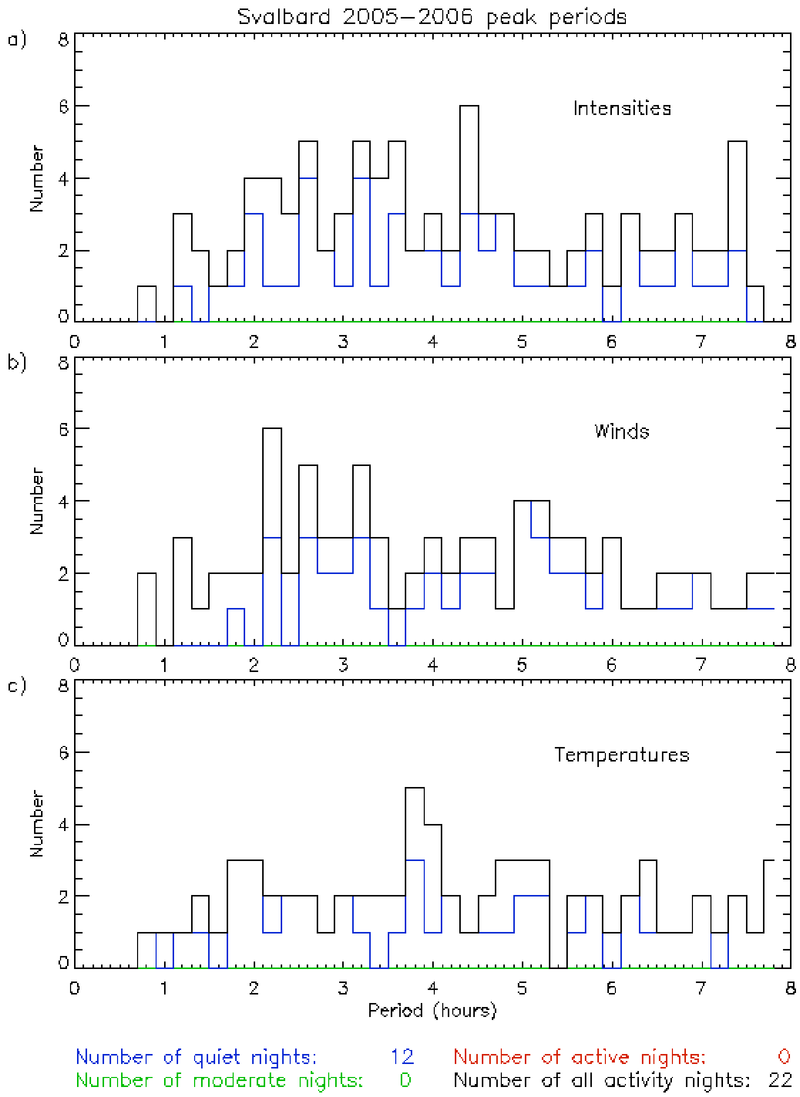

Figure 5.19 *Histogram of gravity wave periods for all nights and all activity levels, from Svalbard data over the winter season of 2005-2006: a) intensities, b) winds, and c) temperatures.*

There is only one wave with a period of less than an hour at Svalbard for this season, compared with 15 for KEOPS. As the time resolutions of the data sets are similar due to the same quality detectors, and the data sets are from the same year and so same point in the solar cycle, the main other difference between the sites is their location. As high latitude thermospheric gravity waves are thought to have a source in the auroral oval, and as KEOPS is located in the nominal auroral oval, it would be expected to be able to detect all waves that are formed there. In contrast, Svalbard is at a higher latitude, in the polar cap region, and so it is some distance from the auroral oval. Waves therefore have to travel some distance before being observed over Svalbard (there are over 1000km between the Svalbard FPI and the mainland sites). As the power and amplitude of a wave is dependant on frequency (e.g. Press & Rybicki, (1989) and Hocke, (1989)), waves with short periods will be attenuated more and they will therefore dissipate quicker and over shorter distances than longer period waves. Short periods cover more cycles over a given distance than long periods so dissipation effects act for further. This is the same effect as the Q-factor in for example electromagnetic waves in circuits or seismic waves. This accounts for the reduced number of short period waves seen in the Svalbard data in relation to the mainland data.

The Svalbard data also show many more waves with longer periods than the KEOPS data, especially over 5 hours. This again is likely to be due to the timing of the data rather than a real variation in the gravity wave parameters. The majority of the nights used for the Svalbard data set were in December and January, whereas the KEOPS data was from the end of August through to March, and the long winter nights were mostly cloudy, with the majority of the clear nights in February and March. This would have resulted in a greater number of nights in Svalbard with long enough hours of darkness to observe the long period gravity waves than in KEOPS.

Data from Sodankylä were also available from this season and were shown in Figure 5.15, and comparisons with the KEOPS data from 2005 – 2006 were discussed in section 5.3.2. The waves in the Sodankylä data show a distribution more similar to the Svalbard data (Figure 5.19) than the KEOPS data (Figure 5.11), in all of the three parameters. This shows that the dissipation of short period waves can have the same consequences for the detection of gravity waves as a decrease in time resolution.

Good quality detectors are therefore an important factor in obtaining a true distribution of gravity waves from the FPI data.

5.3.4 Site Comparisons

To increase the size of the statistical sample, the following plots show the results of adding the gravity waves found in each night over all of the seasons for each site. There were a total of 249 nights used from the KEOPS FPI, 133 from Sodankylä, and 185 from Svalbard. Figure 5.20 shows a histogram of the gravity waves found in each period bin for all the KEOPS data from 2000 to 2006, in the same format as those for the individual seasons. The distributions of the waves over the period range are very similar to those of for example the 2004 – 2005 data in Figure 5.8, for the intensities, winds, and the temperatures. The varying time resolution of the data over each of the years results in the number of the shortest periods of waves (less than an hour) being proportionally less than that, for example, the 2005 – 2006 data in Figure 5.11. The wind and temperature data again have broader peaks in the number distributions, and at longer periods than the intensities.

The main advantage of summing over all the years of data is that it increases the total number of nights with moderate geomagnetic activity (22 nights for KEOPS) and active nights (14). These are still not large statistical samples, but they are large enough to have more than one wave in each period bin, which was often the case with individual years of data. The only year with more than two active nights available was the 2000 – 2001 season, which was at solar maximum.

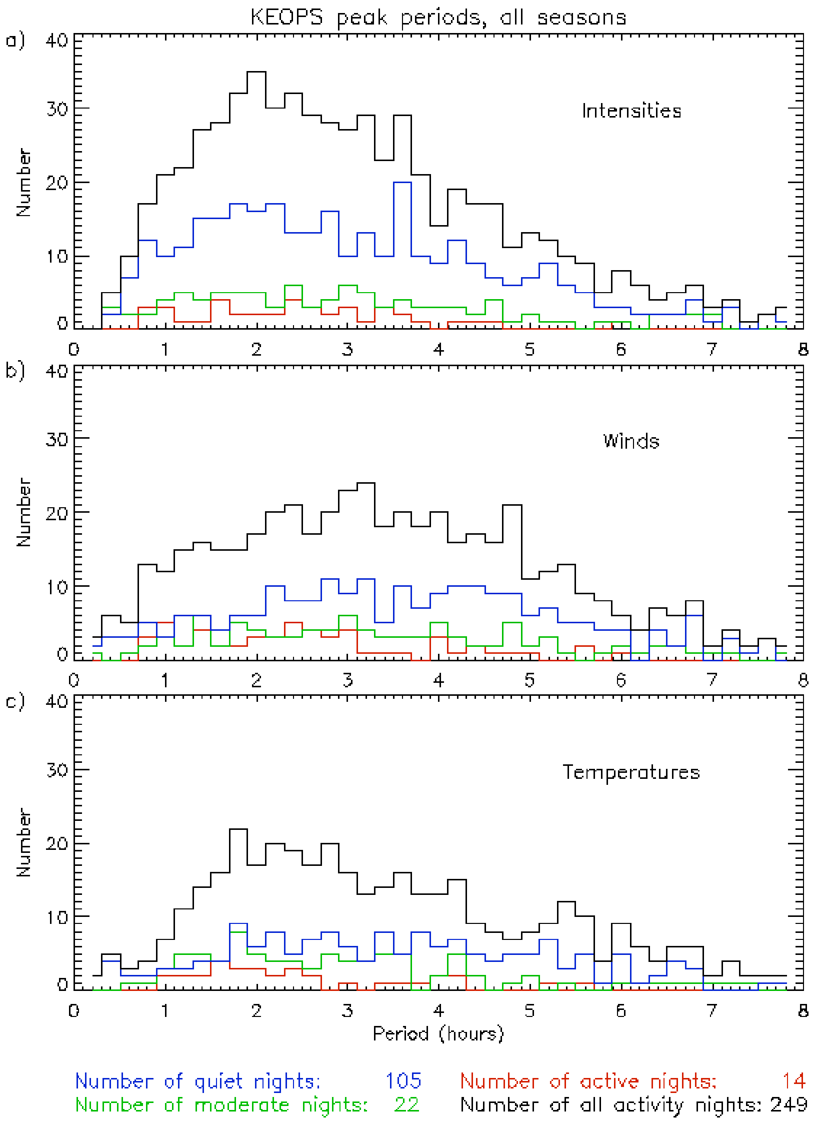

Figure 5.20 *Histogram of gravity wave periods for all KEOPS data (2000 – 2006) at all activity levels: a) intensities, b) winds, and c) temperatures.*

Figure 5.20 shows, particularly with the intensities, that the moderate and active nights have a similar distribution to that of the quiet nights, with a peak between approximately 1 and 4 hours. This is in agreement with the individual seasons' histograms and for example results from Ogawa et al., (1987) to indicate that the distribution of gravity waves is independent of geomagnetic activity. As the Sodankylä FPI was not installed until 2002, past the solar maximum, and there were roughly half the total number of nights available compared to KEOPS, there were only two active nights for the total of the Sodankylä data, shown in Figure 5.21. The moderate geomagnetic activity nights also only produce one or two waves in any one period bin, so also cannot be used to draw any useful conclusions.

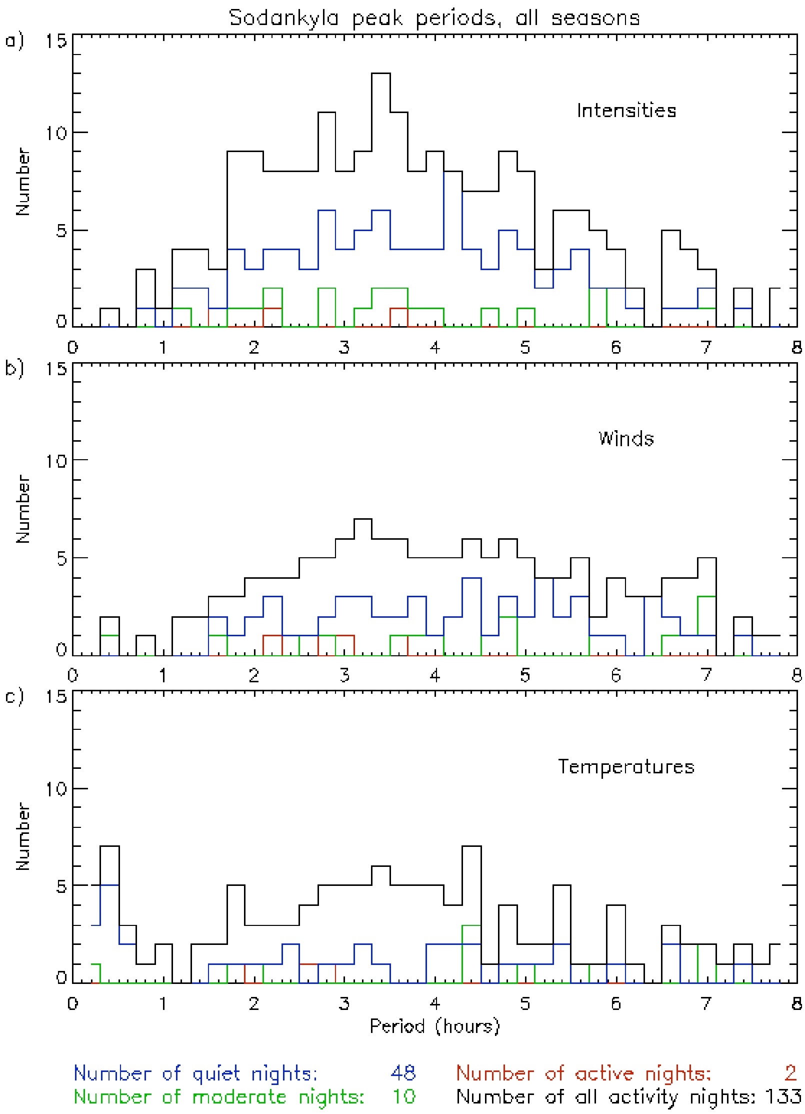

Figure 5.21 *Histogram of gravity wave periods for all Sodankylä data at all activity levels: a) intensities, b) winds, and c) temperatures.*

The total of the four seasons of Sodankylä data for all activity levels produce reasonable numbers of gravity waves. Few waves are observed with periods less than roughly 2 hours, which again is due to the poorer time resolution of the data, which was on average 3 to 4 times longer than for KEOPS. The highest number of waves in any one period bin in the intensities at Sodankylä in Figure 5.21a is 12, whereas for KEOPS, in Figure 5.20a the highest number is 35. Sodankylä therefore has only 35% of the peak number of waves of KEOPS, but it has 54% the total number of nights. However, as the peak at Sodankylä is shifted due to the reduced number of short period waves due to the lower time resolution, the numbers at the same period bin (3.4 hours) should be compared. This gives approximately 52% (±15%) of the KEOPS value, which is similar to the proportion of nights that were used for each data set. This is as would be expected, as the two sites are at similar latitudes, so are in similar parts of the auroral oval. They would therefore be expected to observe the same waves, especially as they have overlapping fields of view, within the experimental limits of the instrumentation.

Figure 5.22 shows the total of all the waves from the Svalbard FPI over the five seasons of data from 2000 to 2006. The most obvious difference between this plot and the two for KEOPS and Sodankylä is the increased number of waves with long periods, due to the increased number of nights with 24 hours of darkness at Svalbard at its higher latitude. This can be seen in the intensities, winds, and temperatures. The range of time resolutions of the data over different seasons at Svalbard is similar to that at Sodankylä, and the reduced number of waves with periods below 2 hours, similar to that at Sodankylä, reflects this.

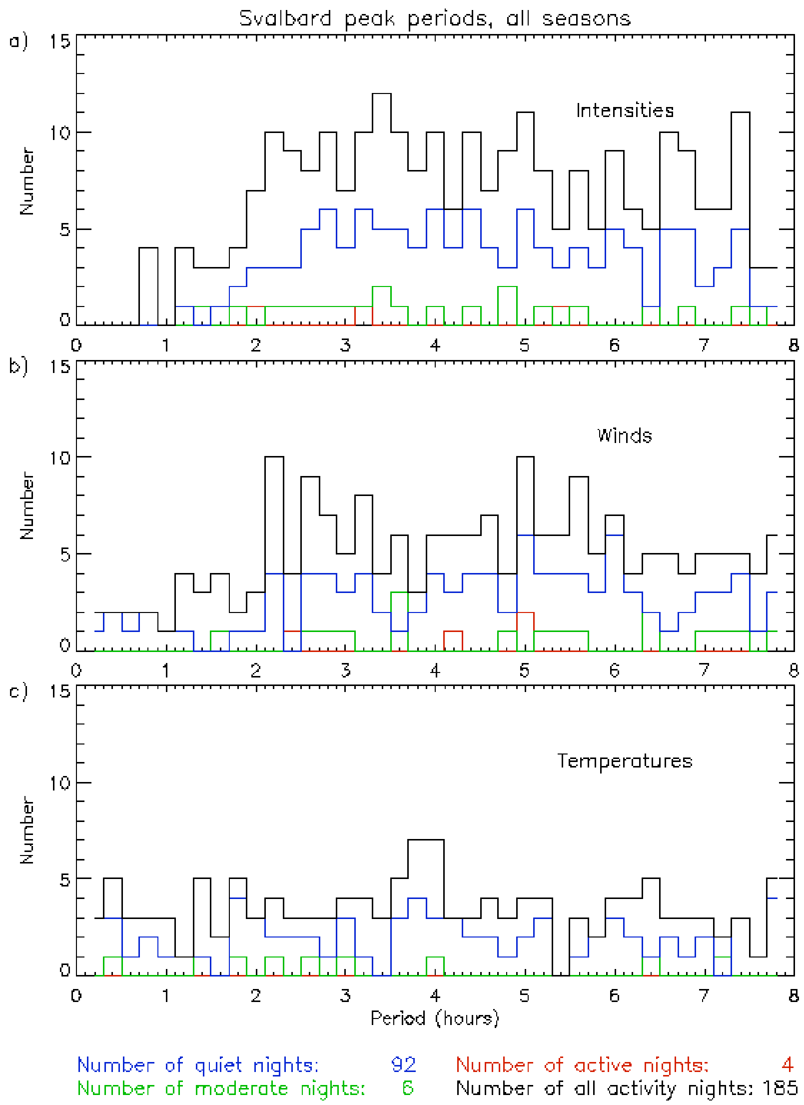

Figure 5.22 *Histogram of gravity wave periods for all Svalbard data at all activity levels: a) intensities, b) winds, and c) temperatures.*

The average occurrence per nights of gravity waves can be calculated from the total number of waves, over the period range used here, divided by the total number of clear nights used in the analysis, over all the seasons. The results of this for the three sites, for each parameter, are shown in Table 5.2 below.

	KEOPS		***Sodankylä***		***Svalbard***	
	Total waves	Occurrence	Total waves	Occurrence	Total waves	Occurrence
Intensity	614	2.5	208	1.6	256	1.4
Wind	485	2.0	139	1.1	196	1.1
Temperature	387	1.6	121	0.9	135	0.7

Table 5.2 *Occurrence of gravity waves per night, for each site, over all seasons of data for the intensities, winds and temperatures.*

The highest proportion of waves over the number of nights was in the KEOPS intensity data, with an average proportion of 2.5 waves per night detected. There are slightly fewer waves in the winds and temperatures. These numbers reflect the differences in different parameters, and the speed with which the intensities can respond to precipitation and energy input compared with the inertia that has to be overcome to move or heat the thermosphere. The occurrence of waves in the Sodankylä data is just over half as large as the KEOPS data. Svalbard data produce similar proportions to the Sodankylä data. This lower detection rate is most likely due to the lower time resolution of the data not sampling the waves sufficiently enough to give a spectral power over the 70% confidence level used here. Sodankylä and Svalbard data sets have on average approximately the same time resolutions, but Svalbard detects a lower proportion of gravity waves in the intensities and temperatures. This difference could be attributed to the site locations. This is as would be expected, as Svalbard is further from the typical source region of the gravity waves in the auroral oval, and so more waves would have dissipated by the time they have travelled the distance to the polar cap.

To show how the distribution of the number of waves varies with season, Figure 5.23 and Figure 5.24 show the total numbers of waves in each month of the year, for

KEOPS and Svalbard respectively. The Sodankylä data are not shown, as several of the months do not have enough nights of data for a significant statistical sample. These plots are similar to the 2000 – 2001 data in Figure 5.4, but the data are added across each of the years to increase the statistical sample. The dotted line shows the number of nights of data that were used for each month of the year and the solid line shows the weighted number of gravity waves detected in that month, by dividing the total number by the number of nights in that month of the year. Again, only periods up to 4 hours are included, to prevent the winter months being biased by the longer length of the night. These plots show the average proportion of nights in which gravity waves are detected.

The KEOPS intensities in Figure 5.23a show statistically significant peaks in the number of waves detected in November and February, and for the winds and temperatures this is nearer September and February. There is a minimum around December and January, at the winter solstice. Geomagnetic activity peaks around the equinoxes (e.g. McIntosh (1959)), due to the orientation of the Earth's magnetic field with respect to the Interplanetary Magnetic Field, so this could explain the increased number of waves here. However, compositional changes may also affect the number of gravity waves formed. For example, the winter anomaly produces a peak in electron density in mid-winter. This may explain why the peaks in Figure 5.23 are not at the equinoxes, as would be expected if it was just a geomagnetic activity dependence, but that the peaks are nearer the winter solstice.

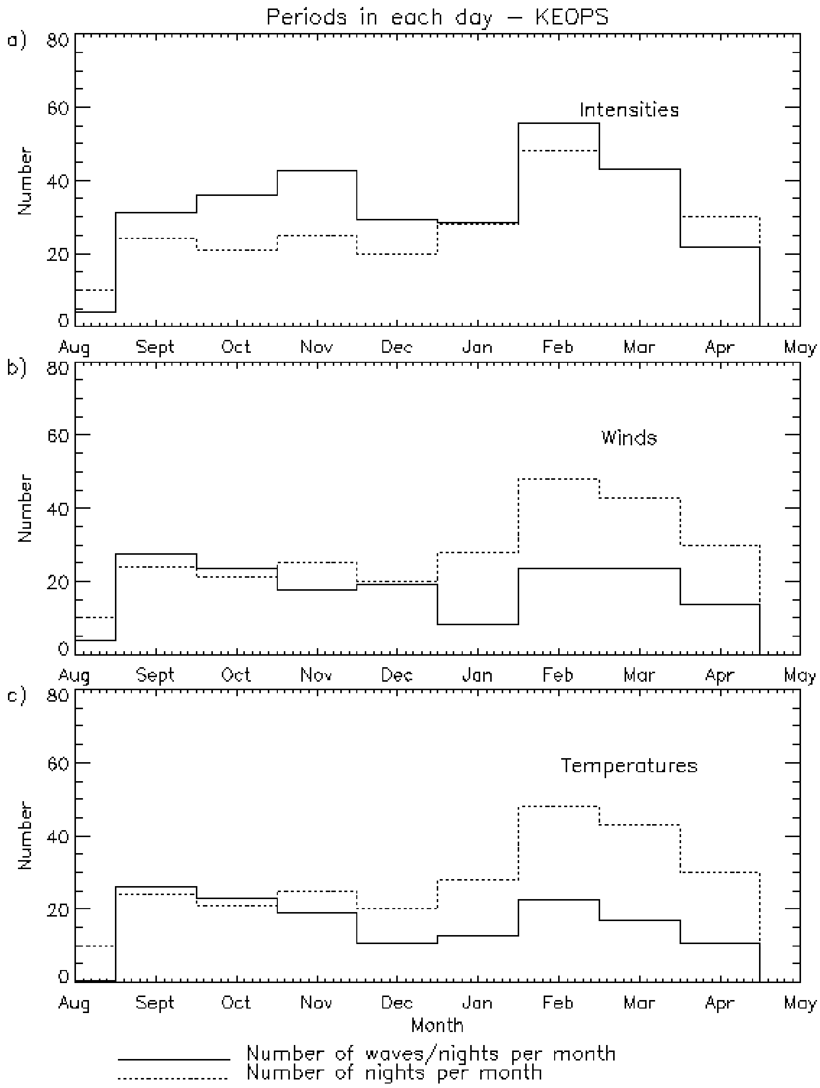

Figure 5.23 *Number of gravity waves with periods 0-4 hours detected in each month as a proportion of the number nights of data, for all KEOPS data: a) intensities, b) winds, and c) temperatures.*

The Svalbard data are shown in Figure 5.24, although there are too few nights in September to consider the data from this month. The lower proportion of waves per night in Svalbard compared with KEOPS from Table 5.2 is also seen in this plot. This lower detection rate is partly due to the lower time resolution of the data not sampling the waves sufficiently enough to give a spectral power over the 70% confidence level used here. Other differences could be attributed to the location of the sites. Svalbard is further from the typical source region of the gravity waves in the auroral oval, and so more waves would have dissipated by the time they have travelled the distance to the polar cap.

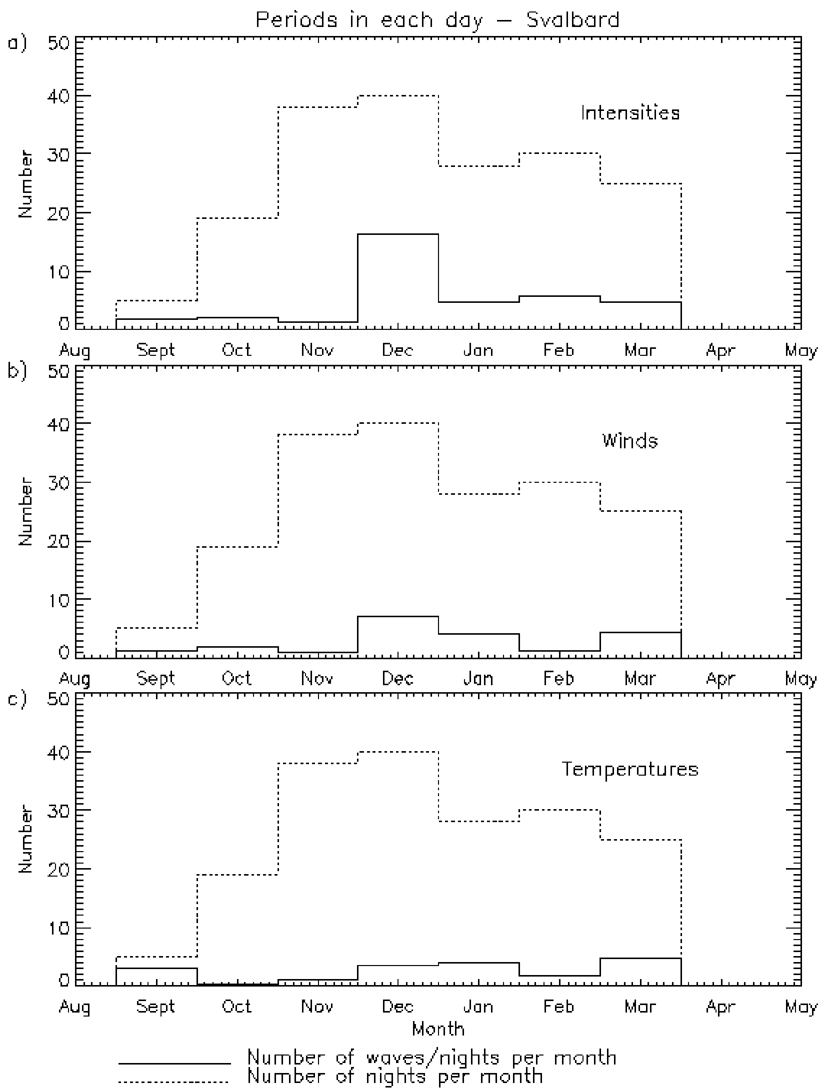

Figure 5.24 *Number of gravity waves with periods 0-4 hours detected in each month as a proportion of the number nights of data, for all Svalbard data: a) intensities, b) winds, and c) temperatures.*

Other than the decreased proportion of waves per night, there is a clear difference between the monthly distributions in this plot than with the KEOPS data in Figure 5.23. There is a large increase in the proportion of waves seen in December, rather than at the equinoxes, in the intensities and winds. The temperatures do not show this so strongly, but do show minima in October and February. The reason for this difference could be that the geomagnetic activity variations with season have a smaller effect on the number of waves formed. The changes in composition could be a greater influence, due to the longer period without any solar radiation, and the greater differences between the mid-winter and the equinoxes.

In summary, the main differences between both the distribution of gravity waves detected over different periods, and of the numbers of gravity waves observed, between the KEOPS data and the Sodankylä data, are due to the time resolution of the instruments. Fewer waves of shorter periods are observed in the Sodankylä data due to insufficient sampling of waves due to longer cycle times as a consequence of less efficient detectors.

The difference between these two mainland sites and Svalbard in the auroral oval is due to two reasons. The lengths of the nights effect means that Svalbard has a larger number of long period waves because it has more nights with long hours of darkness. The lack of short period gravity waves seen in the Svalbard 2005 – 2006 data compared with KEOPS shows that the waves are moving poleward. The same high-resolution integration times are used for both instruments, so the reason must be the dissipation of energy. The distance of Svalbard from the source of the gravity waves in the auroral oval means that the short period waves appear to have dissipated by the time they reach Svalbard. because of the large distance travelled. This is the first evidence of such a feature.

6 CONCLUSIONS & SUMMARIES

Small-scale structure in the high-latitude upper atmosphere has been investigated principally through data from Fabry-Perot Interferometers observing the atomic oxygen red line at 6300Å. Emission intensities, thermospheric winds and temperatures were obtained from this emission from 240km altitude from the northern high latitude region in Scandinavia. Data from the FPIs were complemented by data from instruments such as radars, all sky cameras, and magnetometers. The separations of the FPI look-direction volumes, both from a single FPI and between multiple FPIs, were used to study spatial meso-scale structures of the order of tens of kilometres. Small temporal scales of thermospheric variability down to tens of seconds resolution were measured due to improvements in detector capabilities. In this chapter, the main results are summarised, and the conclusions obtained from these data sets are discussed.

A summary of the relevant features of the Earth's atmosphere and atmospheric gravity waves were covered in Chapter 1. The instrumentation used for the measurements were described in Chapter 2. The results of the correlations between the intensities of the 6300Å emissions from Fabry-Perot interferometer data, with measurements of electron densities from tomography and radar data were shown in Chapter 3 and are summarised in section 6.1. Properties of atmospheric gravity waves obtained from case studies of individual nights that were obtained from Chapter 4 are summarised in section 6.2, and the results of statistical analyses of gravity waves in Chapter 5 are discussed in section 6.3. Some possible further directions for this work, and the future research work for the Atmospheric Physics Laboratory are outlined in section 6.4.

6.1 ELECTRON DENSITY CORRELATIONS

FPI emission intensities of the 6300Å atomic oxygen line were compared with electron densities obtained from the Sodankylä Geophysical Observatory tomography receiver chain, and from the EISCAT radars. If dissociative recombination is the dominant mechanism for production of the 6300Å emission that the FPIs measure, then the electron densities should be proportional to the FPI intensities.

Cross correlations between FPI datasets at different times and sites, and electron densities from the two tomographic reconstruction models, Chapman and IRI, varied considerably. It was found that the IRI model correlates better with the FPI intensities than the Chapman model. Although the Chapman model is more suited to active conditions, this is not always the situation in these data sets. The IRI model has a lower peak electron density height (hmF2) at 230km and as it has input from the IRI empirical model, it is time dependent and more sensitive to variations in conditions. This potentially allows it to show up more small-scale features.

The correlations showed that the small-scale structure of the thermosphere is important, as FPI data taken from the tomography satellite pass mid-point gave better correlations than the average intensities over the whole pass length. Since the satellite pass takes approximately 20 minutes, it showed that the variations are important on smaller scales than this time.

The correlation coefficients between the FPI intensities and electron densities from different altitudes were also compared. In some cases, the highest correlation was obtained from electron densities at 300km altitude. The 6300Å emission layer is typically at around 240km, but can change by over 50km depending on geomagnetic activity. When a larger statistical sample was used, a higher correlation was obtained when the electron densities were averaged over the 240km – 300km range, which is typical of the atomic oxygen emissions. This was as expected, because the activity level and hence emission height are likely to vary over the different passes. In addition, the width of the emission layer is broader than the 20km ranges, and so is likely to be an average of two or more model grid spacing ranges.

Electron densities at different altitudes were also studied using EISCAT radar data, from the first tristatic campaign held on the 27th February 2003. For the KEOPS and Sodankylä data, the largest value of correlation coefficient was obtained from lower altitude N_e data, at heights of 215km and 230km. The average N_e over the range of 215km to 330km also gave high values of coefficients for all FPIs. Of the three FPIs, this average N_e gave the highest value for Skibotn data. This has been shown to be consistent with the tomography data.

At the tristatic point, correlation coefficients from the different combinations of data sets were expected to be the same, as all six data sets (the three FPIs and three radar beams) should be viewing the same common volume. Although the coefficients are similar, they are not identical within the errors, implying that they were not looking at a common volume, which would be the case if the emission height was varying throughout the night.

Reasons for the correlation not being perfect, i.e. coefficients not reaching unity, include observational errors and uncertainties resulting in inaccuracies in the data. Differences in the timings or the locations of the emission peak, in altitude or horizontally, of the two data sets will mean the data that are compared may not be coincident. An additional, physical, reason for the correlation coefficients not being 1 could be due to the atomic oxygen being produced from the combination of atomic nitrogen with molecular oxygen. This accounts for 80% of the emissions below 200km and has been shown to be a source of $O(^1D)$ at thermospheric temperatures in the laboratory. Another reason could be the cascade from the higher energy atomic oxygen that produces the green line 5577Å emission. These cannot be confirmed without the ability to measure the nitrogen or NO levels. From current models, emissions below 200km are small due to quenching. The temporal and spatial variations discussed above are therefore the most likely cause of any further deviations in the correlation coefficients.

6.2 AGW CASE STUDIES

The majority of the previous detections of gravity waves have been from middle or lower atmosphere data. Observations of upper atmosphere gravity waves have mostly been inferred from observations of TIDs rather than true thermospheric gravity waves. This is due to the relative ease of measuring the ionosphere. The common assumption, especially in models, that gravity waves will only exist in the lower and middle atmosphere, from a tropospheric source, implies that the waves will have dissipated before reaching F-region altitudes. Gravity waves with sources already in the high altitude and high latitude auroral regions do not have this problem. Therefore, particle precipitation into auroral regions creating Joule heating, and auroral heating from electrojet currents are thought to be the mechanisms responsible for high latitude thermospheric gravity waves (de Deuge et al., (1994)).

Gravity waves have however been directly observed in the upper thermosphere over the southern polar cap by de Deuge et al., (1994) and Innis et al., (2001), through ground based photometer observations of 6300Å oxygen emissions. Gravity waves have been measured in neutral winds and temperatures only by satellite data e.g. by Johnson et al., (1995) and Innis and Conde, (2001). It was therefore reasonable to expect that gravity waves should be seen in the FPI data presented in this thesis. This is believed to be the first detection of gravity waves in upper thermospheric temperatures. These observations therefore increase the limited number of observations of gravity waves in the upper thermosphere and they provide a view of the gravity waves in the northern hemisphere to complement the previous measurements in the south. The Scandinavian region also has the added advantage of being covered by a range of other instruments, which provide data that have been used to identify the source of the gravity waves.

Lomb-Scargle analysis of the FPI data has provided evidence for thermospheric gravity waves with a variety of periods, from a few minutes to a few hours. Periods down to the Brunt-Väisälä frequency have been seen. The gravity waves were seen in thermospheric intensities, winds and temperatures, from Fabry-Perot interferometer measurements of the 6300Å atomic oxygen line. The amplitudes of the waves show that the data are consistent with an auroral source region. By looking at a few nights in detail, and comparing the FPI data with the analysis of data from other instruments, source regions and typical properties of the waves have been determined.

In the data from the second APL tristatic campaign, on the 25th November 2003, waves of several periods have been found. On this night, the EISCAT radars viewed the same common volume of sky that were viewed by the three FPIs, i.e. the tristatic A position. There is evidence that these waves have different source mechanisms. It has been shown that heating from particle precipitation is the likely source mechanism for a wave of period 3.7 hours. Two periods of precipitation were seen in the FPI intensities, also separated by 3.7 hours. Although the first of these periods of precipitation occurred after waves were seen in the temperature variations, an auroral arc was seen in all-sky camera images 3.7 hours earlier in the night, therefore suggesting that these periods of precipitation were driving the wave. This is further evidence to support a proposal by Williams et al., (1993). The amplitudes of the waves were greater to the north of the region, which is where the auroral activity was observed. This too supports an auroral source region.

For the other periodicities seen in the data of the second tristatic campaign, i.e. 1.2, 1.5, 1.8 and 2.5 hour waves, it is proposed that the source of the gravity waves was Joule heating from electrojet activity. The Lomb-Scargle analysis performed on equivalent current densities, as well as the position of the electrojet, calculated from magnetometer data from the IMAGE chain, revealed periodicities similar to those found in the FPI data. Periodicities seen in the FPI temperature data are very well matched by the periods found from the J_x component of the equivalent current densities, showing that the strength of the electrojet is the cause of the gravity waves observed.

The gravity waves found in the FPI thermospheric data have also been compared with the TIDs found in radar data. EISCAT incoherent scatter data provides electron densities and ionospheric wind speeds and temperatures. STARE coherent scatter radars provide electric field data, which can be used to identify gravity waves at E-region altitudes. These datasets confirm the source of the waves with periodicities of 1.2, 1.5 and 1.8 hours to be generated from Joule heating due to electrojet currents, and the 3.7-hour wave to be of higher altitude in origin, and therefore from low energy particle precipitation causing localised heating.

Cross correlations of the neutral temperatures were used to find that the phase speed of the 1.8 hour wave to be 250ms^{-1} with a phase angle □ of 302°, which gives a horizontal wavelength, λ_h of 1600km. This is both consistent with previous TID measurements and with an auroral oval source region. For the 1.5-hour wave, the STARE radar data only showed the presence of the wave at the bottom of its field of view in the region above KEOPS, and not to the north of it. In addition, the EISCAT data only showed the wave at high altitudes, which, due to the geometry of the instrumental set-up with the radar pointing southwards, indicates that the wave must have had a propagation direction that was mostly southwards. This is also corroborated by the wave only being detected in the FPI data in the south and bistatic B positions of the KEOPS and Sodankylä intensities.

These complementary data sets therefore showed that on the 25th November 2003 waves of several different periodicities were present, and that different source mechanisms produced the different waves.

Results from a high time resolution study of KEOPS data in April 2004 showed that waves with shorter periods are in fact present in the high latitude thermosphere. The tristatic A position was viewed at approximately 15-second resolution due to a new highly sensitive detector, and because a full cycle of measurements were taken only every hour, the effect of slow mirror rotation time was removed from the data. On the 3rd April 2004, periods down to 14 minutes were seen in the 6300Å intensities, with similar periods in the winds and temperatures. Gravity waves with these short periods were seen on several nights, and on all of the three other clear nights when this high-resolution experiment was performed.

On inspection of any random 10 minute period of the high-resolution data from KEOPS, features in the data can be seen to contain many data points. This implies a limit to the scale on which the thermosphere can vary, as more detail is not obtained when higher time resolutions were used. From the data viewed here, this limit appears to be approximately one minute. This shows that the thermosphere can vary on scales down to one minute, not just the often-assumed value of approximately one hour.

Comparisons were made between FPI data from KEOPS in the auroral oval, with Svalbard in the polar cap region. Lomb-Scargle analyses confirmed the source region of a 3.5-hour period gravity waves as being from the auroral oval, as well as showing that the wave partially, but not completely, dissipated over the approximately 1000km distance between the sites. Photometer data from the Spectrograph Imaging Facility showed that the wave had some propagation in the vertical direction, as the emissions observed with the photometers were from species at lower altitudes. The high time resolution of the data also allowed the period of the wave to be identified more accurately as 3.42±0.02 hours.

CUTLASS data confirmed the source region as being at latitudes around KEOPS rather than at Svalbard. As IMAGE magnetometer data did not observe the wave, this supports the source of the waves as being from heating due to particle precipitation rather than electrojet currents. Smoothing the temperature data from KEOPS allowed lags between the zonal and meridional directions to be determined, so that a phase speed of the wave could be calculated as 130±20ms^{-1}, the horizontal wavelength as1680±200km and the propagation direction to be southerly. The wave therefore originated in the auroral oval, which was to the north of KEOPS and propagated equatorward to KEOPS and poleward to Svalbard. The properties of the wave on this night were similar to those of the second tristatic campaign, which implies these could be typical properties of high-latitude thermospheric gravity waves.

6.3 AGW Statistical Survey

Data from the FPIs at KEOPS, Sodankylä, and Svalbard on all the clear nights from the years 2000 to 2006 have been analysed for gravity wave activity. A total of 249 nights were available from KEOPS, where data were collected through the night and the skies were clear. The Sodankylä and Svalbard FPIs provided 133 and 185 nights of data respectively. A Lomb – Scargle analysis was performed on each of these nights, as for the case studies described in Chapter 4 and section 6.2, to identify the periods of any wave activity during the night. Comparisons between many nights of data allow the general characteristics of the waves that are present in the high latitude upper thermosphere to be determined.

Within data from one year at a particular site, comparisons can be made between the different parameters, the atomic oxygen intensities, the thermospheric winds and temperatures. For each parameter, the distribution of frequencies of the waves was determined. Also, the nights in each year have varying geomagnetic activity levels, so these were analysed separately to determine the effect of geomagnetic activity on the number and properties of the waves determined. All the FPIs have had different detectors at various times, producing different cycle times and consequently different time resolutions of the data. Therefore, comparisons between the different years showed how the time resolution determines which waves are observed. This can also be determined from comparisons of data from different sites. The period covered by the data that were used for this study spanned half a solar cycle, and with data from previous years, the effects of solar activity on the production of gravity waves could be determined.

Comparisons between the data from FPIs at different sites showed the dependences both of the length of the dataset each night, and of the geomagnetic location of the sites, with respect to the source region of the gravity waves. The size of the statistical sample was increased by summing the data over the years. This gave further information on geomagnetic activity dependences as well as seasonal variations in the numbers of waves detected. The results of these comparisons are summarised below.

No particular individual frequencies are preferred in the thermospheric gravity waves detected by the FPIs; though a wave with a period of 6 hours was seen is some datasets which was either an atmospheric tide or due to the two cell convection pattern. The distribution of the number of waves across the period range however was not uniform. The time resolution of the data was the most important factor in the number of short period waves detected, and slow cycle times determine both the shortest period that can be detected and reduce the spectral power of the lowest period waves. The number of long period waves that are detected are limited by the length of the night. However, in a dataset (2004 – 2005 geomagnetically quiet nights), where all the nights were longer than 16 hours, there were few waves with periods greater than approximately 3 to 4 hours, implying gravity waves have preferred periods in the intensities shorter than 4 hours.

For the wind and temperature data, covering all seasons, there is a more even distribution with respect to period, of the numbers of waves than for the intensities. Also longer periods, up to around 5 hours, were more common. The differences were due to the response times of the different parameters to particle precipitation. The intensities seem to react very quickly to the energy input from bursts of precipitation, but there is an inertia within the winds and temperatures due to the bulk of the thermosphere, decreasing the number of short period waves formed.

Comparing data from different sites from the same year, and so at the same point in the solar cycle, but with different time resolutions, allowed comparisons of how the resolution affects the periods of waves found. The distribution of periods from low-resolution intensities had a closer correlation to that of the wind and temperature data than that of the high-resolution intensity data from later years. This implies that the intensities give a closer indication of the distribution of periods than the winds and temperatures, as these do not show the short period waves.

Comparisons of datasets from KEOPS at solar maximum and solar minimum did not show any change in the number of waves produced. This implies solar activity has little effect on the production of gravity waves, possibly because gravity waves are relatively easily formed in the thermosphere and do not require high geomagnetic activity.

However, differences in the gravity wave properties due to geomagnetic activity levels cannot be easily determined from one season of data, due to the low numbers of active nights. Therefore, several years of data were used to determine this dependence. Significant statistical samples were nonetheless only created with the KEOPS data. The total nights were separated into geomagnetically quiet ($Kp \leq 3$ all through the night), moderate activity levels ($3 \leq Kp \leq 6$ all night), and active nights ($Kp > 7$ at some time in the night). This, however, did not allow any difference to be shown between the different activity levels in the distribution of the gravity waves over the period range for any of the sites for the intensities, winds, or the temperatures.

Seasonal variations were studied by counting the number of waves detected in each day of the year, for a year when there was a reasonably even distribution of clear nights of data across the year (at KEOPS in 2000 – 2001). A relationship was found between the gravity wave periods detected and the length of the night through the year. These apparent harmonics in gravity wave periods might be formed by waves reflected from the solar terminator boundary, owing to the difference in conductivity between day and night, creating constructive interference that moderates the periods that are sustained. These harmonics are seen for periods greater than 4 hours for months around the solstice. Short period waves will be dissipated more than long period waves over a given distance.

Binning the data into months of the year increased the number of counts across the season. Dividing the number of waves found in a month by the number of nights of data used that month allowed seasonal affects to be investigated. Only waves with periods less than 4 hours were included, to prevent the weighting of the long winter nights that produce more long period waves. This showed an increase in the number of waves detected in the intensities around the equinoxes, with a minimum near the solstice. This is not clear however, especially in the winds and temperatures, due to the low number of nights in each month. Therefore, all of the seasons were combined for each site. KEOPS showed an increase in the number of waves detected in February and November for the intensities, and February and September for the winds and temperatures. This is likely to be due to a combination of the increased geomagnetic activity levels at the equinoxes, and the variations in composition of the atmosphere over the year. Svalbard showed a peak around the winter solstice and did not show an increased number of waves at the equinoxes, implying that compositional affects have more importance at this higher latitude than geomagnetic activity variations.

KEOPS data from the 2004 – 2005 season was used to compare waves found in the four cardinal look directions separately, with the proportion of waves found in any of the look directions, for each period bin. This was used to find that, on average, a wave is seen, at a confidence level above 70%, in 60% of the look directions. This shows that in over half the cases, the wave is close to the site of KEOPS, so auroral sources are likely, as the nominal auroral oval is just to the north of KEOPS. Waves that are not being seen in all of the look directions provides information on the propagation of the gravity waves, and can be due to several reasons. The wave could dissipate part way through the field of view of the FPI. This implies either a source location further away, which will only be the case on very geomagnetically quiet nights, or a weak wave. A wave may have a source within the FPI field of view and only propagating in one direction. Due to the small distances between the look directions in relation to the typical wavelengths of the gravity waves, it is more likely that gravity waves propagate away from the source in a non-uniform direction. This is likely to be in an equatorward or poleward direction, perpendicular to the auroral oval.

Comparisons between the number of gravity waves detected at KEOPS and Sodankylä over all the seasons showed a similar proportion of waves to the number of nights used for both sites, at periods outside the influence of the time resolution of the data and the length of the nights. This was as expected as the two sites are at similar latitudes and therefore locations with respect to the auroral oval, confirming this as a likely source region. Comparisons between the Svalbard data with those from KEOPS and Sodankylä again showed the effects of the time resolution of the datasets to the distribution of periods of waves detected. Svalbard, at a higher latitude, has more nights which have long hours of darkness, hence more long period gravity waves are detected. Svalbard and Sodankylä over the total range of seasons have similar average time resolutions, and both sites have a reduced number of short period waves in relation to KEOPS, which has a much shorter average time resolution. Comparisons between data sets with the same time resolution showed differences between the sites, and Svalbard showed fewer waves with short periods (less than an hour) than the KEOPS data for the same season (2005 – 2006). This gives a clear indication of the direction of flow of the gravity waves, and corroborates that the source is the auroral oval. Energy is dissipated through Joule heating in each cycle of a wave, therefore, over a given distance; short period waves lose more energy and dissipate.

6.4 FUTURE WORK

The future of the Atmospheric Physics Laboratory's ground based observations program lies with the next generation of Fabry-Perot Interferometers: SCANDI. This Scanning Doppler Imager is an All-Sky version of the FPI, so has the great advantage of being able to view large areas of the sky at once, and not just the 1° field of view that the present FPIs are limited to. This is achieved through an all-sky lens at the front end of the optics, rather than the FPI mirror system. This means that a significant fraction of the whole sky, a 130° field of view, will be viewed with each image. The sky image will be broken down into segments, of an adjustable number, but initially 25. This is also discussed by McWhirter et al., (2005).

To obtain an interference fringe profile across all areas of the CCD, the FPI fringes will be scanned. This is achieved by using a capacitance stabilised etalon (CSE). In CSEs, the optical path length of the light is changed by varying the distance between the plates of the etalon, see section 2.2.5 and McWhirter, (1993). The spacers between the plates are made of a piezoelectric material so a change in voltage across the spacer will change its length. Due to the high sensitivity of the Andor camera that will be used, very short exposures are achievable, of approximately 0.1 seconds. This is possible as although the image is broken down into 25 segments, five interference fringes are used, compared to one with the FPIs. This results in about a fifth of the photons per exposure. The fringes are then scanned after each exposure. The scans can then be co-added so that several free spectral ranges are covered in each segment. This will take approximately 60 seconds. Since an intensity map of the whole sky is obtained with each image, the data analysis and calibration are necessarily more complicated. Each of the 25 segments will have of order of 10,000 pixels over all of the exposures. Each pixel will produce an interferogram – an Airy profile of the interference fringe as it is scanned through the pixel during the exposure. As the interferograms will not be in phase for each pixel in a segment, a reference phase map will need to be created to correlate each pixel. This can then be used to calculate intensities, winds, and temperatures simultaneously in many areas across the sky.

Software will be written with help from the University of La Trobe in Australia, who have previously built and run a scanning Doppler imager in Alaska. The first SCANDI instrument is expected to be deployed in Svalbard in the 2006 – 2007 winter season. Data will be initially compared with FPI data for verification. This will be aided by a new intensity calibration system.

Photometers will be installed at each of the sites of the FPIs so that absolute flux can be determined and a scale can be put on the intensity axis. High time-resolution photometers similar to those used with SIF will be used. They will be calibrated in the laboratory from a broadband source by measuring the intensity through a number of passband filters at various wavelengths. From this, a response curve for the whole photometer system can be created. Energies of the incoming precipitation could then be determined, and the FPIs could be compared both with each other, with SCANDI and with other instrumentation.

A collaborative study with the British Antarctic Survey has recently begun at APL, which is using the fractal dimension to try to determine the physical limit of the variability of the thermosphere. Most physical systems have fractal properties, where the amount of variability of the system is similar on the different scales on which it is viewed. There will, however, be a physical limit to the scale on which the system can vary, whether this is at the atomic level or above this, for example at the ion gyro frequency. By calculating the amount of variability for different time scales, a plot of the log of the variability vs. the scale length will give a straight line for fractal data. If the gradient of this line decreases at shorter scales then this is the physical limit on variability, because no more variability is found even though the resolution is increased. This is therefore the limit where the data is no longer 'self similar'.

Studies by Lanchester et al., (1994) and Lanchester et al., (1997) use 0.2s resolution EISCAT radar data to show that the ionosphere varies down to this time scale, which corresponds to a spatial size of much less than 1km. The thermosphere would not be expected to vary at such short time scales due to its large inertia, but the limit of its variability is yet to be determined, but will be investigated by this study. The high-resolution data from the tristatic A position in KEOPS data in April 2004 is initially being used for this study. A large number of data points are however needed to provide a large statistical sample at all scale lengths, so further high resolution studies will be performed during the mid-winter when 24 hours of darkness is available.

The next step for terrestrial observations at APL will be to try to measure the thermospheres of the other planets in the solar system. The variations in the atmospheres and the magnetic fields of the other planets provide useful and interesting comparisons to the Earth's system. This would be achieved by mounting an FPI or SCANDI instrument onto large telescopes, of diameters of 3m or more.

Currently aurorae on the gas giants are measured by H_3^+ observations from the UK Infra-Red Telescope (UKIRT) and the Infra-Red Telescope Facility (IRTF) on Hawaii (e.g. Miller et al, (1990) and Stallard et al., (1999)). Slanger et al., (2001) have found evidence for 5577Å and possibly 6300Å atomic oxygen emissions on Venus. The feasibility of observing these emissions with a ground based FPI system has to be modelled. Such a system will need to have sufficient sensitivity and resolution to determine the wind speeds over a part of the planet's disk at a sensible spatial scale. Infrared ground-based observations might be more easily achievable. There is a carbon dioxide line at 3.4μm in the Martian atmosphere that could potentially be observed. An additional issue with Mars however is that there is rarely more than 10% of its nightside towards us when it is in a viewable location. A further stage of planetary observations that would solve these problems, and provide much higher resolution data, would be to fly an FPI onboard a satellite that would go into orbit around a planet.

6.5 ACKNOWLEDGEMENTS

I would like to thank my supervisor Dr Anasuya Aruliah, secondary supervisor Prof. Alan Aylward, and also Dr Ian Furniss, Dr Ian McWhirter, and Dr Eoghan Griffin at the Atmospheric Physics Laboratory at UCL. I would also like to thank Dr Vikki Howells and Dr Chris Davis at the Rutherford Appleton Laboratory for training and assistance with the EISCAT radar data, and also Dr Betty Lanchester, Dr. Anastasia Stockton-Chalk and Dr. Nickolay Ivchenko at the University of Southampton.

I am grateful to PPARC for funding this PhD studentship, and for the funding of the construction and maintenance of the FPIs. The Marie Curie Fellowship scheme funded by the European Union gave me the opportunity to work at the Sodankylä Geophysical Observatory in Finland.

The ESRANGE KEOPS facility, the Sodankylä Geophysical Observatory and LAPBIAT grant, and UNIS in Svalbard have provided generous help in logistics. EISCAT is an International Association supported by Finland (SA), France (CNRS), Germany (MPG), Japan (NIPR), Norway (NFR), Sweden (VR), and the United Kingdom (PPARC). Thanks to the Rutherford Appleton Laboratory for EISCAT data analysis. The MIRACLE network is operated as an international collaboration under the leadership of the Finnish Meteorological Institute. The IMAGE magnetometer data are collected as a Finnish-German-Norwegian-Polish-Russian-Swedish project, with thanks to Ari Viljanen (FMI) for equivalent current density data.

Many thanks also go to my family, especially my parents for their support and generosity throughout my education.

REFERENCES

Arnold, N.F., Jones, T.B., Robinson, T.R., Stocker, A.J., & Davies, J.A., *Validation of the CUTLASS HF radar gravity wave observing capability using EISCAT CP-1 data*, Annales Geophysicae, **16**, 1392-1399, (1998).

Aruliah, A.L., *The Synoptic Variability of Thermospheric and Mesospheric Winds Observed Using a Fabry-Perot Interferometer*, Ph.D. thesis, University of London, (1991).

Aruliah, A.L., & Rees, D., *The trouble with thermospheric vertical winds: geomagnetic, seasonal and solar cycle dependence at high latitudes,* Journal of Atmospheric and Terrestrial Physics, **57**, 597-609, (1995).

Aruliah, A.L., & Griffin, E.M., *Evidence of meso-scale structure in the high latitude thermosphere*, Annales Geophysicae **19**, 37-46, (2001).

Aruliah, A.L., Griffin, E.M., McWhirter, I., Aylward, A.D., Ford, E.A.K., Charalambous A., Kosch M.J., Davis, C.J., Howells, V.S.C., *First tristatic studies of meso-scale ion-neutral dynamics and energetics in the high-latitude upper atmosphere using collocated FPIs and EISCAT radar,* Geophysical Research Letters, **31**, L03802, (2004), (doi: 10.1029/2003GL018469).

Aruliah, A.L., Griffin, E.M., Aylward, A.D., Ford, E.A.K., Kosch M.J., Davis, C.J., Howells, V.S.C., Pryce, E., Middleton, H., & Jussila, J., *First direct evidence of meso-scale variability on ion-neutral dynamics co-located tristatic FPIs and EISCAT radar in Northern Scandinavia*, Annales Geophysicae, **23**, 147-162, (2005).

Babcock, H.D., *A Study of the Green Auroral Line by the Interference Method*, Astrophysical Journal, **57**, 209-224, (1923).

Balthazor, R.L., & Moffett, R.J., *A study of atmospheric gravity waves and travelling ionospheric disturbances at equatorial latitudes*, Annales Geophysicae, **15**, 1048-1056, (1997).

Balthazor, R.L., & Moffett, R.J., *Morphology of large-scale travelling atmospheric disturbances in the polar thermosphere,* Journal of Geophysical Research, **104**, 15-24, (1999).

Batchelor, G., *An Introduction to Fluid Dynamics*, Cambridge, Cambridge University Press, (1967).

Baumgardner, J., Flynn, B & Mendillo, M., *Monochromatic imaging instrumentation for applications in aeronomy of the Earth and planets*, Optical Engineering, **32**, 3028, (1993).

Bilitza, D., Rawer, D., Bossy, L., & Guliaeva, T., *International Reference Ionosphere: Past, present, and future. I - Electron density. II - Plasma temperatures, ion composition and ion drift*, Advances in Space Research, **13**, 3-13, 15-23 (1993).

Born, M., & Wolf, E., *Principles of Optics*, 6th edition, Oxford, Pergamon Press, (1987)

Bristow, W.A., Greenwald, R.A., & Samson, J.C., *Identification of high-latitude acoustic gravity wave sources using the Goose Bay HF radar*, Journal of Geophysical Research, **99**, 319-331, (1994).

Chamberlain, J.W., *Physics of the Aurora and Airglow*, Academic Press, (1961).

Codrescu, M. V., Fuller-Rowell, T.J., Foster, J.C., Holt, J.M., & Cariglia, S.J., *Electric field variability associated with the Millstone Hill electric field model*, Journal of Geophysical Research, **103**, 5265-5273, (2000).

de Deuge, M.A., Greet, P.A., & Jacka, F., *Optical observations of gravity waves in the auroral zone*, Journal of Atmospheric and Terrestrial Physics, **56**, 617-629, (1994).

England, S.L., Dobbin, A., Harris, M.J., Arnold, N.F., & Aylward, A.D., *A study into the effects of gravity wave activity on the diurnal tide and airglow emissions in the equatorial mesosphere and lower thermosphere using the Coupled Middle Atmosphere and Thermosphere (CMAT) general circulation model*, Journal of Atmospheric and Solar-Terrestrial Physics, **68**, 293-308, (2006).

Fabry, C., & Perot, A., *On a new form of interferometer,* Astrophysical Journal, **13**, 265, (1901).

Ford, E.A.K., Aruliah, A.L., Griffin, E.M., & McWhirter, I., *Thermospheric gravity waves in Fabry-Perot Interferometer measurements of the 630.0nm OI line*, Annales Geophysicae, **24**, 555-566, (2006).

Ford, E.A.K., McWhirter, I., Furniss, I., Charalambous, A. & Stockton-Chalk, A., *SIF Alignment Procedure*, Internal document, (2005).

Fordham, J.L.A., Bone, D.A., Norton, T.J., & Read, P.D., *Development of the MIC photon counting detector for astronomical applications*, SPIE Instrumentation in Astronomy VII, **1235**, 636-643, (1990).

Fordham, J.L.A., Bellis, J.G., & Bone, D.A., *The MIC photon counting detector*, SPIE Electron Image Tubes and Image Intensifiers II, **1449**, 87-98, (1991).

Fritts, D.C., & Alexander. M.J., *Gravity Wave dynamics and effects in the middle atmosphere*, Reviews of Geophysics, **41**, 1003, (2003).

Gill, A.E., *Atmosphere – Ocean Dynamics*, Academic Press, (1982).

Greenwald, R.A., Weiss, W., Neilsen, E., & Thomson, N.R., STARE: A new radar auroral backscatter experiment in northern Scandinavia, Radio Science, 13, 1021-1039, (1978).

Hargreaves, J.K., *The Upper Atmosphere and Solar – Terrestrial Relations*, Berkshire, Van Nostrand Reinhold Co. Ltd., (1979).

Hargreaves, J.K., *The Solar – Terrestrial Environment*, Cambridge, Cambridge University Press, (1992).

Hecht, E., *Optics*, 4th edition, Addison Wesley, (2001).

Hernandez, G., *Fabry-Perot Interferometers*, Cambridge, Cambridge University Press, (1986).

Hedin, A.E., *A Revised Thermospheric Model Based on Mass Spectrometer and Incoherent Scatter Data: MSIS-83*, Journal of Geophysical Research, **88,** 10170, (1983).

Hines, C.O., *Internal atmospheric gravity waves at ionospheric heights*, Canadian Journal of Physics, **38**, 1441-1481, (1960).

Hines, C.O., Paghis, I., Harts, T.R., & Fejer, J.A., *Physics of the Earth's Upper Atmosphere*, New Jersey, Prentice-Hall, (1965).

Hocke, K., *Phase estimation with the Lomb-Scargle periodogram method*, Annales Geophysicae, **16**, 356-358, (1998).

Hocke, K., & Schlegel, K., *A review of atmospheric gravity waves and travelling ionospheric disturbances: 1982-1995*, Annales Geophysicae, **14**, 917-940, (1996).

Horne, J.H., & Baliunas, S.L., *A prescription for period analysis of unevenly sampled time series,* Astrophysical Journal, **302**, 757-763, (1986).

Houghton, J., *The Physics of Atmospheres,* 3rd edition, Cambridge, Cambridge University Press, (2002).

http://nssdc.gsfc.nasa.gov/space/model/models/iri.html.

http://sgodata.sgo.fi/pub_tomo/tomo.html.

Hunsucker, R.D., *Atmospheric gravity waves generated in the high-latitude ionosphere: a review*, Reviews of Geophysics and Space Physics, **20**, 293-315, (1982).

Innis, J.L., & Conde, M., *Thermospheric vertical wind activity maps derived from Dynamics Explorer-2 WATS observations,* Geophysical Research Letters, 28, 3847, (2001).

Innis, J.L., Greet, P.A., & Dyson, P.L., *Thermospheric gravity waves in the southern polar cap from 5 years of photometric observations at Davis, Antarctica*, Journal of Geophysical Research, **106**, 15489, (2001).

Ivchenko, N., Lanchester, B.S., Galand, M., Robertson, S.C., Rees, M.H., Lummerzheim, D., Furniss, I., Peticolas, L.M., Frey, H.U., Baumgardner, J., & Mendillo, M., *High resolution measurements and modelling of auroral hydrogen emission line profiles*, Annales Geophysicae, **21**, 1629-1643, (2003).

Ivchenko, N., Rees, M.H., Lanchester, B.S., Lummerzheim, D., Galand, M., Throp, K., & Furniss, I., *Observation of O+ (4P-4D0) lines in electron aurora over Svalbard,* Annales Geophysicae, **22**, 2805 – 2817, (2004).

Johnson, S., Hanson, W.B., Hodges, R.R., Coley, W.R., Carignan, G.R., & Spencer, N.W., *Gravity waves near 300 km over the polar caps*, Journal of Geophysical Research, **100**, 23993-24002, (1995).

Kaplan, J., *A New Afterglow Spectrum in Nitrogen,* Physical Reviews, **48**, 800, (1935).

Kato, S., *Dynamics of the Upper Atmosphere*, Tokyo, Centre for Academic Publications Japan, (1980).

Killeen, T.L., Craven, J.D., Frank, L.A., Ponthieu, J.-J., Spencer, N.W., Heelis, R.A., Brace, L.H., Roble, R.G., Hays, P.B., & Carignan, G.R., *On the relationship* *between dynamics of the polar thermosphere and morphology of the aurora: global-scale observations from dynamics explorers 1 and 2,* Journal of Geophysical Research, **93**, 2675, (1988).

Kivelson, M.G., & Russell, C.T., *Introduction to Space Physics*, Cambridge, Cambridge University Press, (1989).

Kosch, M.J., Hagfors, T., Rees, D., *A new Fabry-Perot interferometer for atmospheric studies with the EISCAT incoherent radar*, Advances in Space Research, **20**, 1133, (1997).

Lanchester, B.S., Nygen, T., Jarvis M.J., & Edwards R., *Gravity wave parameters measured with EISCAT and Dynasonde*, Annales Geophysicae, **11**, 925-936, (1993).

Lanchester, B.S., Palmer, J.R., Rees, M.H., Lummerzheim, D., Kaila, K., & Turunen, T., *Energy flux and characteristic energy of an elemental auroral structure*, Geophysical Research Letters, **21**, 2789-2792, (1994).

Lanchester, B.S., Rees, M.H., Lummerzheim, D., Otto, A., Frey, H.U., & Kaila, K., *Large fluxes of auroral electrons in filaments of 100m width*, Journal of Geophysical Research, **102**, 9741-9748, (1997).

Lanchester, B.S, Rees, M.H., Robertson, S.C., Lummerzheim, D., Galand, M., Baumgardner, J., Mendillo, M., Furniss, I. & Aylward, A.D, *Proton and electron precipitation over Svalbard - first results from a new Imaging Spectrograph (HiTIES)*, Proc. of Atmospheric Studies by Optical Methods, 92, **33**, (2001).

Lilensten, J., & Amblard, P.O., *On the last movements of the thermosphere: an interpretation*, Annales Geophysicae, **18**, 1651, (2001).

Lindzen, R.S., *Dynamics in Atmospheric Physics*, Cambridge, Cambridge University Press, (1990).

Link, R. & Swaminathan, P.K., *N(2d) + O2 – A source of thermospheric 6300Å emission*, Planetary and Space Science, **40**, 699, (1992).

Lomb, N.R., *Least squares frequency analysis of unequally spaced data*, Astrophysical & Space Science, **39**, 447, (1976).

Lowry, R., *Concepts and Applications of Inferential Statistics*, Vassar, New York, USA, (1999).

Lui, A.T.Y., Venkatesan, D., Murphree, J.S., Anger, J.S., & Rostoker, G., *Dayside auroral intensifications during an auroral substorm*, Geophysical Research Letters, **14**, 415, (1987).

MacDougall, J.W., Andre, D.A., Sofko, G.J., Huang, C.S., & Koustov, A.V., *Travelling ionospheric disturbance properties deduced from Super Dual Auroral Radar measurements* Annales Geophysicae, **18**, 1550, (2001).

Makarevitch, R.A., Honary F., Koustov, A.V., & Uspensky, M.V., *Meridional motions of the afternoon radar aurora, auroral electrojets, and absorption patches under variable IMF conditions,* Annales Geophysicae **22**, 1649-1664, (2004).

Makela, J.J., Kelley, M.C., Gonzalez, S.A., Aponte, N., & McCoy, R.P., *Ionospheric topography maps using multiple wavelength all-sky images*, Journal of Geophysical Research, **106**, 29,175-29,184, (2001).

Mayr, H.G., Harris, I., Herrero, F.A., Spencer, N.W., Varosi, F., & Pesnell, W.D., *Thermospheric gravity waves: observations and interpretation using the transfer function model (TEM)*, Space Science Review, **54**, 2970375, (1990).

McIntosh, D.H., *On the Annual Variation of Magnetic Disturbance*, Philosophical Transactions of the Royal Society, **A251**, 525-552, (1959).

McWhirter, I., *Imaging Photon Detectors and their use with Single an Multiple Fabry-Perot Etalon Systems for Atmospheric Wind Measurements*, Ph.D. thesis, University of London, (1993).

McWhirter, I., Furniss, I., Aylward, A.D., Lanchester, B.S., Rees, M.H., Robertson, S.C., Baumgardner, J., & Mendillo, M., *A new spectrograph platform for auroral studies in Svalbard*, Proc. Atmospheric Studies by Optical Methods, 92, **73**, (2002).

McWhirter, I., Griffin, E.M., Aruliah, A.L., Ford, E.A.K., Charalambous, A., & Kosch, M., *Optical Instrumentation of the Atmospheric Physics Laboratory, University College London*, Proceedings of the 17th ESA/PAC Symposium, ESA SP-**590**, 315-319, (2005).

Miller, S., Joseph, R.D., & Tennyson, J., *Infrared emissions of H_3^+ in the atmosphere of Jupiter in the 2.1 and 4.0 micron region*, Astrophysical Journal, **360**, L55-L58, (1990).

Millward, G.H., Moffett, R.J., & Quegan, S., *Effects of an atmospheric gravity wave on the midlatitude ionospheric F layer*, Journal of Geophysical Research, **98**, 19,173-19,179, (1993).

Mitchell, N.J., & Howells, V.S.C., *Vertical velocities associated with gravity waves measured in the mesosphere and lower thermosphere with the EISCAT VHF radar*, Annales Geophysicae **16**, 1367, (1998).

Morgan, B, *Detecting Airglow Signatures of Gravity Waves in the Thermosphere*, MSc Project, UCL, University of London, (2003).

Nappo, C.J., *An Introduction to Atmospheric Gravity Waves,* California, Academic Press, (2002).

Nygrén, T., Markkanen, M., Lehtinen, M., Tereshchenko, E. D., Khudukon, B. Z., *Stochastic inversion in ionospheric radiotomography,* Radio Science 32, 2359, (1997).

Ogawa, T., Igarashi, K., Aikya, K., & Maeno, H., *NNSS satellite observations of medium scale travelling ionospheric disturbances at southern high latitudes*, Journal of Geomagnetism and Geoelectricity, **39**, 709-721, (1987).

Oliver, W.L., Fukao, S., Yamamoto, Y., Takami, T., Yamanaka, M.D., Yamamoto, M., Nakamura, T., & Tsuda, T., *Middle and upper atmosphere radar observations of ionospheric density gradients produced by gravity wave packets*, Journal of Geophysical Research, **99**, 6321-6329, (1994).

Oliver, W.L., Fukao, S., Sato, M., Otsuka, Y., Takami, T., & Tsuda, T., *Middle and upper atmosphere radar observations of the dispersion relation for ionospheric gravity waves*, Journal of Geophysical Research, **100**, 23763-23768, (1995).

Oyama, S., Ishii, M., Murayama Y., Shinagawa, H., Buchert, S.C., Fujii, R., Kofman, W., *Generation of atmospheric gravity waves associated with auroral activity in the polar F-region*, Journal of Geophysical Research, **106**, 18543-18554, (2001).

Press, W.H., & Rybicki, G.B., *Fast algorithm for spectral analysis of unevenly sampled data,* Astrophysical Journal, **338**, 277-280, (1989).

Rees, M.H., & Roble, R.G., *Excitation of $O(^1D)$ atoms in aurorae and emissions of the [OI] 6300Å line*, Canadian Journal of Physics, **64**, 1608, (1986).

Rees, M.H., *Physics and chemistry of the upper atmosphere*, Cambridge, Cambridge University Press, (1989).

Rishbeth, H., & Williams, P.J.S., *The EISCAT Ionospheric Radar: the System and its Early Results*, Quarterly Journal of the Royal Astronomical Society, **26,** 478-512, (1985).

Scargle, J.D., S*tudies in Astronomical time series analysis: II Statistical aspects of spectral analysis of unevenly spaced data*, Astrophysical Journal, **263**, 835-853, (1982).

Sheperd, G.G., et al., *WINDII: the wind imaging interferometer on the Upper Atmosphere Research Satellite*, Journal of Geophysical Research, **98**, 1303-1306, (1993).

Shibata, T. & Schlegel, K., *Vertical structure of AGW associated ionospheric fluctuations in the E- and lower F-region observed with EISCAT - A case study,* Journal of Atmospheric and Terrestrial Physics, **55**, 739-749, (1993).

Sica, R.J., Hernandez, G., Romick, G.J., Rees, M.H., & Roble, R.G., *Auroral zone thermospheric dynamics. II - Individual nights*, Journal of Geophysical Research, **91**, 13593-13611, (1986).

Slanger, T.G., Cosby, P.C., Huestis, D.L., & Bida, T.A., *Discovery of the atomic oxygen green line in the Venus night airglow*, Science, **291**, 463-465, (2001).

Solomon, S.C., Hays, P.B., & Abreu, V.J., *The auroral 6300A emission: observations and modeling*, Journal of Geophysical Research, **93**, 9867, (1988).

Stallard, T.S., Miller, S., Ballester, G.E., Rego, D., Joseph, R.D., & Trafton, L.M., *The H_3^+ emission profile of Saturn*, Astrophysical Journal, **521**, L149-152 (1999).

Taylor, M.J., Bishop, M.B., & Taylor, V., *All-sky measurements of short period waves imaged in the OI (557.7 nm), Na (589.2 nm) and near infrared OH and $O_2(0,1)$ nightglow emissions during the ALOHA-93 campaign*, Geophysical Research Letters, **22**, 2833-2836 (1995).

Vaughan, J.M., *The Fabry-Perot Interferometer*, Bristol, Adam Hilger, IOP Publishing Ltd., (1989).

Viljanen, A. & Häkkinen, L., *IMAGE magnetometer network.* Satellite-Ground Based Coordination Sourcebook (eds. M. Lockwood, M.N. Wild and H.J. Opgenoorth), ESA publications SP-1198, 111-117, (1997).

Williams, P.J.S., Virdi, T.S., Lewis, R.V., Lester, M., Rodger, A.S., McCrea, I.W., & Freeman, K.S.C., *Worldwide atmospheric gravity-wave study in the European sector 1985-1990*, Journal of Atmospheric and Terrestrial Physics, **55**, 683-696, (1993).

Williams, P.J.S., Mitchell, N.J., Beard, A.G., Howells, V.S.C., Muller, H.G., *The Coupling of Planetary Waves, Tides and Gravity Waves in the Mesosphere and Lower Thermosphere*, Advances in Space Research, **24**, 1571, (1999).

Wilms, J., *scargle.pro*, http://astro.uni-tuebingen.de/software/idl/aitlib/timing/scargle.html, IAAT, (2000).