MODELING ELECTRON DENSITY AT HIGH LATITUDES: DEVELOPMENT OF THE EMPIRICAL CANADIAN HIGH ARCTIC IONOSPHERIC MODEL (E-CHAIM)

by

David R. Themens

B.Sc., University of New Brunswick, 2011
M.Sc., McGill University, 2013

A Dissertation Submitted in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

in the Graduate Academic Unit of Physics

Supervisor: P. T. Jayachandran, PhD, Physics

Examiners: Ben Newling, Ph.D., Physics
Igor Mastikhin, Ph.D., Physics
Karl Butler, Ph.D., Earth Sciences

External Examiner: Jan Sojka, Ph.D., Physics, Utah State University

This dissertation is accepted by the
Dean of Graduate Studies

THE UNIVERSITY OF NEW BRUNSWICK
July, 2018

©David R. Themens, 2018
ABSTRACT

The ionosphere is an important medium for high frequency (HF) radio communications and remote sensing, as well as a hindrance to the use of the Global Positioning System (GPS); thereby, these systems require accurate ionospheric models in order to function. The highly complex nature of the high latitude ionospheric dynamics, combined with an extreme scarcity of data in the high-latitude region, has, in the past, made this area virtually impossible to model accurately. With the recent explosion of ionospheric remote sensing instruments in the polar region, it has now become possible to monitor these regions with high spatial resolution. Today there exist no accurate ionosphere models specific to the high latitude region and de facto standard ionospheric models, such as the International Reference Ionosphere (IRI), have been shown to be inaccurate at high-latitudes.

Here we present the methodology and performance of the new Empirical Canadian High Arctic Ionospheric Model (E-CHAIM), a 3D high latitude electron density model intended to replace the use of the IRI in these regions. To this end, we make use of every available high latitude radio remote sensing ionospheric data set dating back to the very first observations of the ionosphere from 1931 at the Slough ionosonde in Ditton Park, UK. Specifically, we examine lessons learned from the short comings of other empirical ionospheric models, discuss the reasoning behind the parameterizations used, and provide comparisons between the model and real observations that weren’t included in model fitting.
Overall, the E-CHAIM model is demonstrated to represent a significant improvement over current standards in the representation of the topside and F2-peak of the ionosphere, while providing comparable performance to current standards in the representation of the bottomside shape.
DEDICATION

In dedication to the memory of Joop Sanders, cherished uncle and owner/operator of
"The Voice of the Arctic". We miss you.
ACKNOWLEDGEMENTS

This work would not have been possible without the support and encouragement of several people. I would like to extend my gratitude to Dr. P.T. Jayachandran (Jay) for his years of guidance and support, without which I may never have decided to do research. Jay has supported my every endeavor over the past ten years of undergraduate and Ph.D. research, allowing me to attend numerous international and national conferences, experience informative summer schools, and publish in reputable journals. I owe Jay far more than a few words on a thesis can say.

This work also would not have been possible without the support and assistance of the CHAIN group over the years: Richard Chadwick and Todd Kelly for their technical support and patience; Anthony McCaffrey for years of insightful discussions; Lauren Heatherington and Eric Webster for their IDL assistance and codes that I still use to this day; Sajan Mushini, Chris Watson, and Hichem Mezaoui for their comradery and insight over the years; and the many other members of the group who all have contributed to this work.

I would also like to thank my family for their support and encouragement over the years and my wife, Rachael, who has entertained my non-sensical ramblings, supported me at my times of weakness, and kept me in one piece over the past ten years.

Lastly, I would like to thank Mitchell Clark, Melodie Piercey, Ilya Krouglkov, Gabriel Gan, John Harvie, Peter McLaughlin, the UNB and Escrime Mont-Royal fencing clubs, and the rest of the maritime fencing community for being so welcoming and supportive over the years. Fencing will always have a place in my life.
Infrastructure funding for CHAIN was provided by the Canadian Foundation for Innovation and the New Brunswick Innovation Foundation. Science funding is provided by the Natural Sciences and Engineering Research Council of Canada. The author would like to thank the many ionosonde operators who provided data to this project. This paper uses ionospheric data from the USAF NEXION Digisonde network, the NEXION Program Manager is Mark Leahy. This dissertation makes use of data from the Qaanaaq and Nord ionosondes, owned by the U.S. Air Force Research Laboratory Space Vehicles Directorate and supported in part by the Air Force Office of Scientific Research. The author thanks Svend Erik Ascanius of the Danish Meteorological Institute and Denmark’s Arctic Command for the operation of these ionosondes. This dissertation uses data from the Juliusruh Ionosonde, which is owned by the Leibniz Institute of Atmospheric Physics Kuehlungsborn. The responsible Operations Manager is Jens Mielich. The author is grateful to Konstantin Ratovski for the operation of the Irkutsk ionosonde. The Tromsø ionosonde is funded in part by QinetiQ. PFISR and RISR operations are supported by NSF Cooperative Agreement AGS-1133009. ISR data was acquired through the use of the Madrigal data portal at http://isr.sri.com/madrigal/. I would like to thank all Incoherent Scatter Radar (ISR) operators for their continued contribution of data to the Madrigal Database. The author acknowledges UCAR/CDAAC (Boulder, CO, USA) for the use of the CHAMP, GRACE, and COSMIC Radio Occultation data. Intercosmos-19 topside sounding data was provided by IZMIRAN. EISCAT is an international association supported by research organisations in China (CRIRP), Finland (SA), Japan (NIPR and STEL), Norway (NFR), Sweden (VR), and the United Kingdom (NERC). The
authors would also like to thank Dr. Bruno Nava for his insightful comments and suggestions regarding this manuscript.

This dissertation was supported in part by a New Brunswick Innovation Foundation graduate student grant, by a National Science and Engineering Research Council (NSERC) CGS-D Alexandre Graham Bell graduate scholarship, and by Defence Research and Development Canada (DRDC) contract #W7714-186507/001/SS.
# Table of Contents

ABSTRACT ............................................................................................................................... ii  
DEDICATION ........................................................................................................................... iv  
ACKNOWLEDGEMENTS ........................................................................................................... v  
Table of Contents .................................................................................................................... viii  
List of Tables ............................................................................................................................ xiv  
List of Figures ............................................................................................................................ xv  
List of Nomenclature and Abbreviations .................................................................................. xxxii
  
1 Introduction ............................................................................................................................ 1  
2 Physics of the Ionosphere and Thermosphere ..................................................................... 8  
2.1 Composition of the Neutral Atmosphere ........................................................................ 10  
2.2 The Thermospheric Neutral Wind .................................................................................. 15  
2.3 Continuity: Production, Loss, and Transport in the Ionosphere ..................................... 18  
  2.3.1 Photoionization ............................................................................................................ 18  
  2.3.2 Precipitation ................................................................................................................. 22  
  2.3.3 Recombination .............................................................................................................. 29  
  2.3.4 Transport ...................................................................................................................... 31  
3 The Vertical Structure of the Ionosphere ............................................................................. 40  
3.1 The D-Region .................................................................................................................... 40  
3.2 The E-Region .................................................................................................................... 42  
3.3 The F-Region .................................................................................................................... 44  
  3.3.1 F1-Layer ...................................................................................................................... 45  
  3.3.2 F2-Layer ...................................................................................................................... 50
3.4 Topside and Plasmasphere ................................................................. 52

4 Other High Latitude Features and Dynamics .......................................... 54
  4.1 Classification of High Latitude Regions ............................................. 54
  4.2 Polar Solstices .............................................................................. 55
  4.3 Main Ionospheric Trough ................................................................. 56
  4.4 Behaviour during Geomagnetic Storms .............................................. 60
    4.4.1 Thermospheric Effects ............................................................. 61
    4.4.2 Electrodynamical Effects .......................................................... 64
  4.5 Winter and Semi-Annual Anomalies ................................................ 66
  4.6 Solar Activity Variability .................................................................. 68

5 Instrumentation, Indices, and Coordinate Systems .................................. 72
  5.1 Ionospheric Effects on Radio Waves ................................................ 72
    5.1.1 The Complex Refractive Index of the Ionosphere ....................... 73
    5.1.2 The Group Refractive Index of the Ionosphere ............................. 78
  5.2 Ionosonde ................................................................................... 80
  5.3 Topside Sounder ........................................................................... 83
  5.4 Incoherent Scatter Radar (ISR) ......................................................... 84
  5.5 Global Navigation Satellite System (GNSS) ....................................... 87
    5.5.1 Principles of GNSS Ionospheric Remote Sensing ....................... 88
    5.5.2 Biases and Limitations .............................................................. 94
5.5.3 Satellite-based TEC and Radio Occultation ......................................... 101

5.6 Indices .............................................................................................................. 104

5.6.1 Sunspot Number ($R_z$) ............................................................................... 104

5.6.2 F10.7 Radio Flux .......................................................................................... 105

5.6.3 IG Index ........................................................................................................ 106

5.6.4 Planetary K Index (Kp) and Planetary a Index (ap) ........................................ 107

5.6.5 Auroral Electrojet Index (AE) ........................................................................ 107

5.6.6 Dst Index ..................................................................................................... 108

5.7 Altitude Adjusted Corrected Geomagnetic (AACGM) Coordinates ............ 109

6 Existing Empirical Ionospheric Models .......................................................... 112

6.1 The International Reference Ionosphere (IRI) ............................................... 112

6.1.1 The F2-Peak .................................................................................................. 114

6.1.2 F2-Region Bottomside ................................................................................. 137

6.1.3 Topside ......................................................................................................... 142

6.1.4 TEC .............................................................................................................. 152

6.2 The NeQuick .................................................................................................. 177

6.2.1 Formulation ................................................................................................ 178

6.2.2 Validation .................................................................................................... 179

7 Introduction to the E-CHAIM Models ............................................................ 195

8 Modeling the Ionospheric Peak Density: $NmF2$ .......................................... 198

8.1 Data ................................................................................................................ 198
10.2.3 Topside Shape Errors................................................................. 251
10.3 Refitting and re-Parameterizing the NeQuick ................................. 255
10.3.1 New r and g constants .................................................................. 256
10.3.2 Re-fitting $H_0$............................................................................. 259
10.3.3 A Topside for E-CHAIM............................................................... 262
10.4 Behaviour of the E-CHAIM Topside ............................................... 266
11 The Bottomside Model ....................................................................... 272
11.1 The IRI Bottomside ......................................................................... 272
11.2 Revisiting the NeQuick Bottomside .................................................. 276
11.3 E-CHAIM Bottomside ..................................................................... 277
    11.3.1 Data.......................................................................................... 277
    11.3.2 Bottomside Function................................................................. 279
    11.3.3 $h_mE$ and $h_mF1$.................................................................... 281
    11.3.4 Fitting the Bottomside Function ............................................... 286
    11.3.5 Parameterization of the Bottomside Scale Height Amplitudes.... 290
    11.3.6 Validation.................................................................................. 290
    11.3.7 Challenges due to Precipitation ............................................... 296
12 Conclusions and Future Work ............................................................. 300
12.1 Inclusion of an Auroral E-Region Model .......................................... 300
12.2 $h_mF2$ Storm Model ...................................................................... 301
12.3 Merging E-CHAIM with IRTAM ................................................................. 302
12.4 New Physics ............................................................................................ 303
12.5 Further Validation .................................................................................. 304

Bibliography .................................................................................................. 305
Curriculum Vitae ......................................................................................... 323
List of Tables

Table 1 Sources for solar, ionospheric, and geomagnetic indices. .................................. 104

Table 2 CHAIN Station Geographic Locations and Status ............................................. 118

Table 3 Summary of the characteristics of the topside sounder missions used in this study. ................................................................................................................................. 243

Table 4 RMS errors from each hmF1 modeling method tested................................. 286
List of Figures

Figure 2.1 Example neutral, ion, and electron number density profiles at ionospheric altitudes above White Sands, New Mexico (32°N, 106°W). Taken from Rishbeth and Garriott [1969]. ................................................................. 9

Figure 2.2 Example electron density profiles generated using the IRI at (60°N, 270°E) for local midnight (red) and local noon (black) at low (dashed lines) and high (solid lines) solar activity corresponding to June 2009 and June 2013, respectively. ......................... 10

Figure 2.3 Neutral atmospheric composition generated by the US Standard Atmosphere, taken from Hargreaves (1992). ........................................................................................................... 13

Figure 2.4 An illustration of the atmospheric neutral temperature profile (left) and ionospheric electron density (right) with some traditional nomenclature annotations. This figure is taken from Hunsucker and Hargreaves [2003]. ......................................................... 14

Figure 2.5 Neutral wind vectors and temperature contours at 300km altitude generated using the cTIPe full-physics model for December 19th, 2003, at 00UT. Taken from Huba et al. (2014). ......................................................................................................................... 17

Figure 2.6 Altitude at which the atmospheric optical depth equals unity at zenith. Principle ionizing species are marked with ranges above the height curve. “Air” refers to all atmospheric constituents and O$_2$(SR) refers to Schumann-Runge continuum.

“Ionization thresholds” refer to the maximum solar radiation wavelength at which the listed atmospheric constituents could be ionized. Taken from Rees [1989]......................... 19

Figure 2.7 Normalized Chapman production function plotted vs. reduced height ($z = h - hm0H$) for various solar zenith angles. Taken from Hunsucker and Hargreaves [2003]. 21
Figure 2.8 An illustration of the main plasma regions of the magnetosphere and their relation to the magnetic field. From Hunsucker and Hargreaves [2003].

Figure 2.9 A statistical representation of the auroral oval (shaded) based on all-sky imager data for quiet, moderate, and active geomagnetic conditions. Taken from Kivelson and Russell [1995].

Figure 2.10 Vertical profiles of ionization rates for energetic electron (top) and proton (bottom) precipitation for various initial energies. From Watson [2016].

Figure 2.11 Particle precipitation as a function of geomagnetic latitude. Taken from Kelley [2009].

Figure 2.12 Southward Bz convection patterns plotted in geomagnetic latitude and magnetic local time with (b) and without (a) accounting for the effects of corotation. Taken from Hunsucker and Hargreaves [2003].

Figure 2.13 Statistical convection patterns sorted by IMF orientation. Colour and equipotential lines correspond to the electric potential, with red (blue) shades corresponding to positive (negative). Equipotential contours are plotted at 6 kV intervals. Taken from Pettigrew et al. [2010]. Note, the plasma convection is assumed to follow these equipotentials.

Figure 3.1 Electron density contours from the North face of the Resolute Incoherent Scatter Radar (RISR-N) on May 6th, 2013, from all available beams. Note the daytime enhancement of the E-region (10UT - 02UT) and the sporadic-E structures that form near sunset (22UT - 04UT). For reference, local noon is at 17:45UT.

Figure 3.2 Illustrations of the principles behind the formation of the F1-Layer/Ledge. The case on the left is for a situation where the alpha-to-beta transition is above the location.
of maximum photoionization and the case of the right is for when this altitude is below the altitude of maximum photoionization. In terms of nomenclature, references to “Attachment-type” refer to beta-type processes, while references to “recombination-type” refer to alpha-type processes. Taken from Yonezawa [1965].

Figure 3.3 F1-Layer/Ledge occurrence rate versus solar zenith angle derived using global ionosonde observations at various geomagnetic latitudes (λ). 1968 corresponds to a period of high solar activity and 1964 corresponds to one of low solar activity. Taken from Scotto et al. [1998].

Figure 3.4 Ion concentration profiles measured by the Arecibo Incoherent Scatter Radar in Puerto Rico for daytime (top) and nighttime (bottom). Taken from Schunk and Nagi [2009].

Figure 4.1 Plot of the solar terminator altitude vs. local time along the 270°E meridian at 50°N (black), 60°N (blue), 70°N (green), and 80°N (red) during the winter and summer solstices. There are no lines for 70°N and 80°N during the summer, as these regions are sunlit 24 hours of the day.

Figure 4.2 Illustration of the extent and morphology of the MIT for the Northern (right) and Southern (left) hemispheres. Taken from Sojka and Schunk [1989].

Figure 4.3 Illustration of the relative location of the MIT, Auroral Oval, and Polar Hole with respect to the high latitude convection pattern. Taken from Kelley [2009].

Figure 4.4 Meridional cross sections of thermospheric winds, demonstrating the effect of high-latitude heating during geomagnetic storms on neutral atmospheric circulation. Situation a) corresponds to quiet conditions, b) corresponds to modest heating (10^{11} J/s),
and c) corresponds to storm periods with substantial auroral heating (10^{12} \text{ J/s}). Taken from Hargreaves [1992].

Figure 4.5 IDA4D estimates of vertical TEC over the region of EMPIRE analysis for 20 November 2003 at (a) 17:10, (b) 17:30, (c) 17:50, (d) 18:10, (e) 18:30, and (f) 18:50 UT. Taken from Huba et al. [2014].

Figure 4.6 Maps of the distribution of Winter and Semi-Annual Anomaly behaviour in foF2 at low (bottom), moderate (top right), and high (top left) solar activity. Note that (A) and (B) distinctions regard the intensity of the winter anomaly, where (A) corresponds to winter anomaly behaviour with foF2 less than 2 MHz greater than that during the equinoxes and (B) corresponds to winter anomaly behaviour with foF2 more than 2 MHz greater than during the equinoxes. Black dots represent the geomagnetic poles. Taken from Rishbeth [1998].

Figure 4.7 Top: NmF2 from ionosondes between 60° and 70° magnetic latitude from 1970 to 2016. Bottom: F10.7 solar flux corresponding to the periods with ionosonde data. See Section 8.1 for a description of the dataset used in generating this figure. (Generated by the author using the dataset discussed in Section 8.1)

Figure 4.8 Plot of O/N_2 ratio vs. F107 flux (black) and 3-month smoothed F10.7 flux (red) at 300km altitude in January (left) and June (right) between 1961 and 2016 for (65°N, 260°E). Neutral densities were generated here using the MSIS-E-90 climatological neutral atmosphere model.

Figure 5.1 The collision-less complex refractive index (a) and its real (b) and imaginary components (c) plotted against X with values of Y < I for arbitrary angles of propagation. The upper (ordinary) and lower (extraordinary) signs of the Appleton-Hartree Equation...
are given by the vertically and horizontally shaded regions, respectively. T here refers to
the transverse limit and L refers to the longitudinal limit. Taken from Ratcliffe [1959].

Figure 5.2 Ionogram at Moscow ionosonde MO155 on June 21, 2013, at 14:01 UTC.
Pink corresponds to the ordinary mode and green corresponds to the extraordinary model.
Superimposed in black is the virtual height trace of the O-mode and the true height
profile after inversion. The y-axis is altitude in km and the x-axis is frequency in MHz.

Figure 5.3 ISR spectra for O\textsuperscript{+}, He\textsuperscript{+}, and H\textsuperscript{+} ions for three values of T\textsubscript{e}/T\textsubscript{i}. From Evans [1969].

Figure 5.4 Schematic illustration of the geometric projection of slant GPS measurements
to vertical. From Themens et al. [2013].

Figure 5.5 Example of the summed standard deviations versus test bias at Resolute for
24-hour integration on September 19, 2009. The minimum summed standard deviations
is marked by a red asterisk at the receiver bias value of 34.8 TECU. From Themens et al.
[2015].

Figure 5.6 An illustration of the GNSS Radio Occultation geometry.

Figure 5.7 Hourly AU (blue), AL (red), and Dst (black) indices for May 2010.

Figure 5.8 An example of AACGM (red) and Magnetic Apex (blue) coordinate
grids at the surface. Shaded areas denote locations where AACGM coordinates are undefined.
Taken from Laundal and Richmond [2017].

Figure 6.1 A map of the CHAIN GPS stations used in this and following studies. Red
markers correspond to locations where ionosondes are collocated with GPS observations.
Approximate lower boundaries of the auroral oval (MLAT 65N) and polar cap (MLAT
75N) are marked with dashed and dotted lines, respectively.
Figure 6.2 CADI-measured and IRI modelled NmF2 using both CCIR and URSI coefficient maps at Resolute between 2008 and 2011 ................................................................. 122

Figure 6.3 Percent differences between CADI and IRI NmF2 using both CCIR and URSI coefficient maps at Resolute. ........................................................................................................ 123

Figure 6.4 Monthly RMS errors between CADI and IRI NmF2 using CCIR (dashed) and URSI (solid) coefficient maps at Resolute (black), Eureka (red), Pond Inlet (blue), and Cambridge Bay (green). ........................................................................................................ 124

Figure 6.5 CADI-derived NmF2 from the Cambridge Bay, Pond Inlet, and Eureka CHAIN stations .................................................................................................................. 126

Figure 6.6 Digisonde and IRI modelled NmF2 using URSI and CCIR coefficient maps at the Qaanaaq GIRO station between 2004 and 2011 ......................................................... 127

Figure 6.7 Percent differences between Digisonde and IRI NmF2 using URSI and CCIR coefficient maps at Qaanaaq. .................................................................................................. 128

Figure 6.8 Monthly RMS errors between Digisonde and IRI NmF2 using URSI (solid) and CCIR (dashed) coefficient maps at Qaanaaq between 2004 and 2011. ................. 129

Figure 6.9 CADI-derived and IRI-modelled hmF2 using URSI, CCIR, and CADI foF2 at Resolute between 2008 and 2011. ......................................................................................... 130

Figure 6.10 Percent differences between CADI and IRI hmF2 at Resolute for each of the options of Figure 10. ................................................................................................. 131

Figure 6.11 Monthly RMS errors between CADI and IRI hmF2 using URSI (solid black), CCIR (dashed black), and CADI (solid red) foF2 at Resolute. ................................. 133

Figure 6.12 CADI-derived and IRI-modelled M(3000)F2 at Resolute between 2008 and 2011 ......................................................................................................................... 135
Figure 6.13 Percent differences between CADI and IRI M(3000)F2 at Resolute between 2008 and 2011................................................................. 136

Figure 6.14 CADI-derived and IRI-modelled B0 at Resolute using both the Gulyaeva and Table options................................................................. 139

Figure 6.15 Percent differences between CADI and IRI B0 at Resolute using the Gulyaeva and Table options................................................................. 141

Figure 6.16 Example of ionosonde – RISR-N NmF2 (top) and hmF2 (bottom) comparison after calibration. RISR-N mean (black) and median (blue) curves for elevation angles greater than 60° are compared to CADI (red) NmF2 and hmF2 measurements. Right panels show histograms of the ratio of RISR-N to CADI NmF2 (top) and hmF2 deviation (bottom). Mean and median ratios for the electron density are ~0.99 +/- 0.02 (standard deviation of ~20%)................................................................. 145

Figure 6.17 ISR-derived, monthly median, peak normalized, topside electron density (as percentage of NmF2) at Resolute between September 2009 and August 2010........... 147

Figure 6.18 IRI-modelled, monthly median, peak normalized, topside electron density (as percentage of NmF2) at Resolute between September 2009 and August 2010........... 148

Figure 6.19 Percent differences between IRI and ISR monthly median, peak normalized, topside electron density at Resolute between September 2009 and August 2010........... 149

Figure 6.20 Topside scale parameter (Ho) derived from ISR (solid black) and IRI topside profiles using URSI foF2 (solid red) and CADI-ingested foF2 and hmF2 (dashed red) at Resolute between September 2009 and August 2010................................................. 150
Figure 6.21 Contours of monthly median diurnal vTEC behaviour between 2009 and 2015 at Edmonton (a), Sanikiluaq (b), Iqaluit (c), and Pond Inlet (d) using GPS (left), IRI with IG12 (center), and IRI with monthly IG (right). ................................................................. 155

Figure 6.22 Monthly IG index (solid) and IG12 (dashed) between 2009 and 2015. ....... 156

Figure 6.23 Overall RMS errors at all stations between 2009 and 2015 using IG12 (dashed) and monthly IG index (solid) with the IRI CCIR (red) and URSI (black) options. Standard errors of the differences are provided as error bars. ......................... 159

Figure 6.24 Differences in monthly RMS errors between the use of IRI with IG12 and monthly IG index for the URSI option. Positive values imply an improvement through the use of the monthly IG index ................................................................. 161

Figure 6.25 Monthly median GPS (left), IRI with IG12 (middle), and IRI with monthly IG index (right) vTEC versus latitude. .................................................................................................. 163

Figure 6.26 Monthly IRI with IG12 (red) and IRI with monthly IG index (blue) RMS errors for all stations (arranged by latitude from top left to bottom right). GPS RMS standard deviations (black) are also included for each station. ......................... 164

Figure 6.27 Slopes calculated by fitting monthly diurnal variations between GPS and IRI vTEC for IG12 (solid) and monthly IG index (dashed) ......................................................... 167

Figure 6.28 Pearson correlation coefficients of monthly diurnal variations between GPS and IRI vTEC for IG12 (solid) and monthly IG index (dashed) ................................. 169

Figure 6.29 Contours of monthly median diurnal variations of bottomside TEC calculated using CADI (left), the IRI with the URSI option (center-left), the IRI with the CCIR option (center-right) and the IRI with CADI-ingested NmF2 (right) at the Resolute CHAIN station. ................................................................. 171
Figure 6.30 Contours of monthly median diurnal variations of topside TEC calculated using CADI (left), the IRI with the URSI option (center-left), the IRI with the CCIR option (center-right) and the IRI with CADI-ingested NmF2 (right) at the Resolute CHAIN station. ................................................................. 172

Figure 6.31 Monthly median bottomside TEC (top) and monthly RMS errors in bottomside TEC (bottom) for the IRI URSI (blue), IRI CCIR (red), IRI with CADI-NmF2 (dashed black) and GPS (black) at the Resolute CHAIN location. ...................... 174

Figure 6.32 Monthly median topside TEC (top) and monthly RMS errors in topside TEC (bottom) for the IRI URSI (blue), IRI CCIR (red), IRI with CADI-NmF2 (dashed black) and GPS (black) at the Resolute CHAIN location................................................................. 175

Figure 6.33 Example peak-relative bottomside electron density profiles for a) January 2, 2009, at 11:00UTC, b) January 14, 2009, at 00:30UTC, c) June 14, 2009, at 14:00UTC, and d) June 1, 2009, at 2:30UTC. The black solid line corresponds to the measured profile, the orange squares correspond to the best fitted NeQuick profile (i.e. method B4), the green triangles correspond to the profile calculated using method B3, the red stars correspond to the profile calculated using method B1, and the blue diamonds correspond to the profile calculated using method B2. Corresponding B2Bot values for each of the methods is provided in the legend of each image in km. ..................................................... 182

Figure 6.34 Monthly median B2Bot calculated by methods B4, B3, B1, and B2, respectively from left to right. ................................................................. 184

Figure 6.35 Monthly RMS errors in B2Bot for situations B1 (dashed black line), B2 (dashed red line), and B3 (solid black line) at Resolute. ................................................. 185
Figure 6.36 Plot of fitted B2Bot vs. the NmF2 to (dN/dh)max ratio (black points). The blue line represents the expected relationship given Equation 6.9. 186

Figure 6.37 Monthly median topside thickness at RISR as determined by methods T1, T2, T3, and T4, respectively from left to right. 189

Figure 6.38 Same as Figure 6.36 but for PFISR. 190

Figure 6.39 Percent errors in monthly median H0 at PFISR for methods T2, T3, and T4, respectively from left to right. 192

Figure 6.40 Monthly RMS errors in topside thicknesses derived using IRI bottomside parameters (solid red line), using measured bottomside parameters (solid black line), and fitted B2Bot (dashed black line) at RISR (a) and PFISR (b). 194

Figure 8.1 Plot of the global distribution of ionosondes used in the creation of the NmF2 phase of E-CHAIM (red dots). The dashed line corresponds to the lower boundary of the model at 50°N geomagnetic latitude. The dotted line corresponds to 65°N geomagnetic latitude. Black X’s represent stations that were not included in the model fitting and were instead used in the model validation. 199

Figure 8.2 Ionosonde-measured (left column), E-CHAIM-modeled (middle column), and IRI-modeled (right column) NmF2 for the Resolute (A), Alert (B), Eielson (C), and Cambridge Bay (D) validation stations. 205

Figure 8.3 Monthly RMS errors in E-CHAIM (solid line) and IRI (dashed line) foF2 at the Resolute (top left), Alert (top right), Cambridge Bay (bottom left), and Eielson (bottom right) validation stations. 207
Figure 8.4 Contours of foF2 from E-CHAIM (A) compared to that generated by the IRI (B), for representative months in 2013 at 00 (top row), 08 (middle row), and 16 (bottom row) UTC.

Figure 8.5 Correlations (top) and slopes (bottom) for linear fits of IRI (dashed) and E-CHAIM (solid) NmF2 to that from the Resolute ionosonde for each local time hour.

Figure 8.6 Contours of median NmF2 from ionosonde observations (top) and E-CHAIM (bottom) between 50°N and 55°N AACGM latitude plotted against local time and F10.7 solar flux for a) December, b) September, and c) June. Please note that white space in the above plots corresponds to an absence of data.

Figure 8.7 Contours of median NmF2 from ionosonde observations (top) and E-CHAIM (bottom) between 50°N and 55°N AACGM latitude plotted against month and F10.7 solar flux for a) local noon, and b) local midnight.

Figure 8.8 Contours of median NmF2 from ionosonde observations (top) and E-CHAIM (bottom) between 70°N and 75°N AACGM latitude plotted against local time and F10.7 solar flux for a) December, b) September, and c) June.

Figure 8.9 Contours of median NmF2 from ionosonde observations (top) and E-CHAIM (bottom) between 70°N and 75°N AACGM latitude plotted against month and F10.7 solar flux for a) local noon, and b) local midnight.

Figure 8.10 Contours of median NmF2 from ionosonde observations (top) and E-CHAIM (bottom) between 85°N and 90°N AACGM latitude plotted against local time and F10.7 solar flux for a) December, b) September, and c) June.
Figure 8.11 Contours of median NmF2 from ionosonde observations (top) and E-CHAIM (bottom) between 85°N and 90°N AACGM latitude plotted against month and F10.7 solar flux for a) local noon, and b) local midnight. ................................................................. 220

Figure 8.12 (top row of each quadrant) Ionosonde-measured (black), E-CHAIM modeled (blue), and IRI-modeled (red) NmF2 around the May 29th, 2010 geomagnetic storm at the Resolute (top left - RESC), Eureka (top right - EURC), Cambridge Bay (bottom left - CBBC), and Pond Inlet (bottom right - PONC) validation stations. (middle row of each quadrant) Differences between observations and the E-CHAIM (blue) and IRI (red) modeled observations for the corresponding stations. (bottom row of each quadrant) Kp index for the periods presented. ........................................................................................................... 224

Figure 8.13 Same as Figure 8.12 but for Eielson between June 15th and July 5th, 2013 (left) and October 1st and October 25th, 2013 (right). ......................................................................................... 226

Figure 9.1 Plot of the global distribution of ionosondes used in the creation of the hmF2 phase of E-CHAIM. The dashed line corresponds to the lower boundary of the model at 50°N geomagnetic latitude. The dotted line corresponds to 65°N geomagnetic latitude. .................................................................................................................. 229

Figure 9.2 Contours of hmF2 from E-CHAIM (A) compared to that generated by the IRI (B), for representative months in 2013 at 00 (top row), 08 (middle row), and 16 (bottom row) UTC. ............................................................................................................................... 232

Figure 9.3 Ionosonde-measured (left column), E-CHAIM-modeled (middle column), and IRI-modeled (right column) hmF2 for the Eielson (A), Cambridge Bay (B), and Resolute (C) validation stations. ........................................................................................................................................ 234
Figure 9.4 Monthly RMS errors in E-CHAIM (solid line) and IRI (dashed line) hmF2 at the Resolute (top), Cambridge Bay (middle), and Eielson (bottom) validation stations.

Figure 10.1 a) Plot of NeQuick scale height vs. altitude for the following r and g combinations (r,g): red (20,0.2024), black (100,0.125), blue (100,0.5), and green (100,1.0). Dotted lines represent the asymptotes for the corresponding curve. b) Plot of NeQuick scale height vs. altitude subsection of the domain presented in a), where dotted lines now represent the corresponding H = Ho + g(h-hmax) curve.

Figure 10.2 A map of the locations of the ISRs used in this study with measurement density from all data sources superimposed (density is presented in 2° x 2° latitude x longitude bins).

Figure 10.3 Plots of monthly median Ho derived from measurements (black), from the traditional IRI/NeQuick (red), and from refitting the NeQuick k parameterization (green) for 00LT, 06LT, 12LT, and 18LT local time hours.

Figure 10.4 A) Plots of monthly RMS Ho errors for the traditional IRI/NeQuick (red) and for the refitted NeQuick k parameterization (green) for 00LT, 06LT, 12LT, and 18LT local time hours between 2007 and 2016. B) Same as Figure 10.3, but only for 2007-2016.

Figure 10.5 a) Distributions of Ho differences from measurement-derived values for the traditional IRI/NeQuick (red) and the refitted k parameterization (black). b) Mean (solid) and standard deviation ranges (dashed) of Ho vs. geomagnetic latitude. c) Mean (solid) and standard deviation ranges (dashed) of Ho vs. magnetic local time. d) Mean (solid) and standard deviation ranges (dashed) of Ho vs. integrated AE index.
Figure 10.6 Left) Normalized distributions of absolute electron density errors, as a percentage of NmF2, vs. altitude for the traditional IRI/NeQuick. Right) Normalized distributions of the corresponding relative errors in electron density vs altitude. .......... 252

Figure 10.7 Same as Figure 10.6 but for the best-fitted NeQuick parameterization (r,g) = (100,0.125). ........................................................................................................ 254

Figure 10.8 Contour plots a) of the average NmF2-relative RMS error and b) the average correlation coefficient of best-fitted NeQuick profile functions using varying values of r and g. The + marks the corresponding minimum RMS or maximum correlation for each plot. The X marks the values used by the traditional NeQuick model. ...................... 257

Figure 10.9 Same as Figure 10.6 but for the best-fitted new NeQuick model function using (r,g) = (20,0.2024). ........................................................................................................ 258

Figure 10.10 Same as Figure 10.5 but the black curves now represent the new H₀ parameterization of Equation 10.2 and the red curves represent the refitted traditional NeQuick k parameterization using the new (r,g) = (20,0.2024). ........................................ 260

Figure 10.11 Same as Figure 10.6 but for the new NeQuick parameterization of Equation 10.2 with (r,g) = (20,0.2024).................................................................................. 262

Figure 10.12 Same as Figure 10.10 but for H₀ derived from the E-CHAIM parameterization (black) and that of Equation 10.2 (red). .......................................................... 265

Figure 10.13 Same as Figure 8.23 but for the E-CHAIM topside model. .................. 266

Figure 10.14 Contours of quiet time topside thickness (H₀) from E-CHAIM for representative months in 2009 (A) and 2013 (B) at 00 (top row), 08 (middle row), and 16 (bottom row) UTC. Note the change in contour colour bar........................................ 267
Figure 10.15 Contours of the topside thickness progression of the June 1\textsuperscript{st}, 2013, geomagnetic storm. The number at the top of each plot indicates the number of hours past 00UT May 30\textsuperscript{th}, 2013. Plots are provided every eight hours at 00UTC, 08UTC, and 16UTC on each day, such that each row of the above plot can be taken as a day between May 31\textsuperscript{st} (top) to June 3\textsuperscript{rd} (bottom). The above plots are provided in AACGM latitude (radius), and magnetic local time (angle counter clockwise from the bottom of each plot).

Figure 11.1 Illustration of the six components of the IRI bottomside profile. 1) The Topside, 2) the F2 bottomside, 3) the F1-layer/ledge, 4) An intermediate region between the F1-layer/ledge and the E-region valley, 5) the E-region valley, and 6) the bottom of the E-region and the D-region.

Figure 11.2 Bottom) Examples of the IRI electron density profile function fitted to ionosonde observations at Ramey, Puerto Rico. Top) A zoom in to the F1-F2 transition for various F1-layer/ledge thicknesses. Taken from Reinisch and Huang [2000].

Figure 11.3 Same as Figure 8.1 but for the stations used in the E-CHAIM bottomside shape model.

Figure 11.4 Plot of mean hmF1 behaviour vs. AACGM latitude (top left), AACGM local time (top right), integrated AE index (bottom left), and solar zenith angle (bottom right). Ionosonde measured values are represented by black squares, E-CHAIM is represented by connect black stars, NeQuick is represented by connected blue stars, the “traditional” parameterization is represented by connected red stars, and a best fitted quadratic function in solar zenith angle is represented by connected green stars.
Figure 11.5 Example ionosonde-derived electron density profiles at Cambridge Bay (a-c) and corresponding scale height profiles (d-e) for three situations: a,d) no E-region trace, b,e) no F1-layer/ledge trace, and c,f) a profile with all three layers present. d-f) also demonstrate the various components of the scale height function, where the dotted line is the scale height function for the F1-layer/ledge, the dashed line is the constant term, the dash-dotted is the scale height function for the E-region, and the solid line is the final bottomside scale height function.

Figure 11.6 Contours of monthly median fitted $H_{F2}$ (top), $H_{F1}$ (middle), and $H_{E}$ (bottom) for the Dourbes ionosonde between 2006 and 2013.

Figure 11.7 F2-peak normalized electron density profiles at the Millstone Hill ISR for various UT times between 2007 and 2015. Note that local time at Millstone Hill is UT – 4.8 hours.

Figure 11.8 Model-to-measurement bottomside electron density errors for E-CHAIM (left) and the IRI (right) at Millstone Hill between 2007 and 2015.

Figure 11.9 Differences between IRI and E-CHAIM bottomside electron density profile RMS errors. Negative values correspond to periods and locations where E-CHAIM outperforms the IRI.

Figure 11.10 Same as Figure 11.9 but for RISR (left) and the Tromso ISR (right).

Figure 11.11 Bottomside electron content measured by the Millstone Hill (left) and Tromso (right) ISRs and that modeled by the IRI (red) and E-CHAIM (blue) between 2007 and 2015 at various UT times. Note that the Millstone local time is UT – 4.8 and that for Tromso is UT + 1.3.
Figure 11.12 Example electron density profiles from PFISR without (top) and with (bottom) auroral precipitation structures. Solid lines correspond to measured profiles, dotted lines correspond to E-CHAIM, and dashed lines correspond to the IRI. Note that ISR hmF2 and NmF2 have been used here in place of the model values to facilitate comparison of just the profile shape.

Figure 11.13 Model-to-measurement bottomside electron density errors for E-CHAIM (left) and the IRI (right) at PFISR between 2010 and 2013. Note local time at PFISR is UT – 9.8 hours.
List of Nomenclature and Abbreviations

*B0*: IRI Bottomside thickness parameter
*B1*: IRI Bottomside shape parameter
*B2Bot*: NeQuick bottomside thickness parameter
*CHAIN*: Canadian High Arctic Ionospheric Network
*CHAMP*: Challenging Minisatellite Payload Satellite
*COSMIC/FORMOSAT3*: Constellation Observing System for Meteorology, Ionosphere, and Climate
*DIDBase*: Digital Ionogram Database
*E-CHAIM*: Empirical Canadian High Arctic Ionospheric Model
*foE*: Peak critical frequency of the E Layer
*foF1*: Peak critical frequency of the F1 Layer
*foF2*: Peak critical frequency of the F2 Layer
*GIRO*: Global Ionospheric Radio Observatory
*GLONASS*: Russian Global Navigation Satellite System
*GNSS*: Global Navigation Satellite System
*GPS*: Global Positioning System
*GRACE*: Gravity Recovery and Climate Experiment satellite
*H0*: NeQuick topside thickness parameter
*HF*: High Frequency
*hmE*: Peak height of the E Layer
*hmF1*: Peak height of the F1 Layer
*hmF2*: Peak height of the Ionosphere
*ICEPAC*: Ionospheric Communications Enhanced Profile Analysis and Circuit Prediction Program
*IMF*: Interplanetary Magnetic Field
*IRI*: International Reference Ionosphere
*ISIS*: International Satellites for Ionospheric Studies
*ISR*: Incoherent Scatter Radar
*M3000F2*: F2-region transmission factor at 3000km
**MIDAS**: Multi-Instrument Data Analysis System

**MIT**: Main Ionospheric Trough

**MUF**: Maximum Usable Frequency

**NmE**: Peak density of the E Layer

**NmF1**: Peak density of the F1 Layer

**NmF2**: Peak density of the ionosphere

**OTH**: Over the Horizon Radar

**RO**: Radio Occultation

**SPIDR**: Space Physics Interactive Data Resource

**SW-M-I**: Solar wind-Magnetosphere-Ionosphere

**TEC**: Total Electron Content

**TECU**: TEC Units

**UHF**: Ultra High Frequency

**VOACAP**: Voice of America Coverage Analysis Program for HF Propagation Prediction and Ionospheric Communications Analysis
1 Introduction

The ionosphere, a layer of ionized plasma located between roughly 60km and 2000km altitude, is produced by photoionization of the upper atmosphere by solar radiation and, at high latitudes, by precipitating energetic particles. At mid latitudes, the ionosphere’s dynamics are largely governed by photoionization and relatively simple electrodynamics. The high-latitude region, however, is one of the most dynamic and variable regions of the ionosphere, where the Earth’s magnetic field connects directly with the Interplanetary Magnetic Field (IMF) or maps to the outer magnetosphere, resulting in complex Solar Wind-Magnetosphere-Ionosphere (SW-M-I) interactions [Hunsucker and Hargreaves, 2003]. The dynamics of the high-latitude region are further complicated due to heating in the auroral oval, which can drive neutral winds, produce neutral composition changes, and enhance recombination rates that significantly alter the state of the ionosphere; furthermore, high latitude electric fields drive plasma circulation at various small and large scales. All of these factors, combined with an extreme scarcity of data at high latitudes, have made accurate operational modelling the ionospheric electron density of this area virtually impossible to accomplish in the past. Despite these high latitude challenges, the ionosphere has important implications for ground-to-ground and satellite-to/from-ground communications.

In the case of high frequency (HF) propagation and communications, a standard in military and aviation communications, the ionosphere serves as a reflector for these
signals allowing them to be received at distances of over 3000km. Not all HF signals are reflected by the ionosphere. The maximum frequency that can be used (MUF), and thus the maximum bandwidth that is available, is proportional to the square of the ionosphere’s maximum electron/ion density; thus, the electron/ion density of the ionosphere requires regular forecasting to facilitate the most efficient choice of frequency [Jodalen et al., 2001]. Additionally, HF over-the-horizon Radar (OTHR), which uses the ionosphere as a virtual mirror for target detection at several thousand kilometers range, requires accurate propagation models in order to properly range targets; thus, the accuracy of such ranging techniques is severely limited by the user’s ionospheric model [Davies, 1990].

In the case of trans-ionospheric communications (satellite-to/from-ground communications), signals that travel through the ionosphere are bent and slowed down. This leads to significant delays in the travel of these signals, where the magnitude of this delay is proportional to the total amount of ionized material between the satellite and the receiver, typically referred to as Total Electron Content (TEC). In the case of the Global Positioning System (GPS), this leads to significant errors in the system’s capability to determine the position of a receiver of roughly 10m to 100m [Hernández-Pajares et al., 2002]. While errors of this magnitude may not typically be significant to many traditional GPS users, industrial use of GPS data for automated mining and resource extraction, as well as aviation altimetry, requires accurate ionospheric GPS correction. Also, the ionosphere can induce GPS system failures by means of scintillation, which can cause a receiver to lose lock in certain circumstances. This system failure problem is often
exacerbated during periods of enhanced geomagnetic activity [Kintner et al., 2007]. For these reasons, it is important that researchers and systems engineers have access to accurate ionospheric electron density information. This, however, has proven to be difficult to provide at high latitudes.

Themens et al. [2014] and Themens and Jayachandran [2016] demonstrate that the International Reference Ionosphere (IRI), which is considered the climatological standard for ionospheric specification, is extremely inaccurate at high latitudes and fails to reproduce even some of the most basic of ionospheric variabilities. In Themens and Jayachandran [2016], the 2007 version of the IRI is found to produce errors in excess of 50% when modeling the TEC above various Arctic sites, while demonstrating errors within 10% at mid latitudes. The majority of these errors were shown to result from issues in the IRI’s topside representation and to occur predominantly during low solar activity and equinox periods. Themens et al. [2014] goes one step further, diagnosing specific components of the IRI’s vertical electron density representation. They demonstrate that IRI 2007 exhibits errors at times in excess of 70% in the peak electron density (NmF2) of the ionosphere, while mismodeling the peak height (hmF2) of the ionosphere by up to 80km. In terms of propagation, the IRI is found to demonstrate critical failures in modeling the diurnal variability in M(3000)F2 propagation factor (a measure of the MUF), where diurnal variability is found to be the opposite of that modeled by the IRI during summer periods. Meanwhile, Themens et al. [2017b] diagnosed errors in the IRI and NeQuick empirical model’s bottomside and topside representations, finding that the shape function used to represent the bottomside of the
ionosphere in the NeQuick model is incapable of representing the profile shape and that both models struggle in representing the thickness of the topside. Themens et al. [2017b] goes on to demonstrate that accepted methods of mitigating errors in the topside through adjustments of the bottomside parameterization are incapable of resolving these issues and in fact make errors far worse.

Issues in modeling and observing the high latitude ionosphere are not, however, solely constrained to issues in modeling electron density. Themens et al. [2013] demonstrated that the sparsity of data in high latitude regions contributes to significant errors in the representation of TEC in these regions by global TEC maps and Themens et al. [2015] showed that the strong gradients in TEC at high latitudes significantly degrade the performance of standard Global Positioning System (GPS)-based TEC calibration techniques. In fact, the technique proposed in Themens et al. [2015] represents the sole GPS receiver bias estimation technique that has been specifically designed and validated for use at high latitudes, demonstrating improvements over other techniques of a few TEC units (1 TEC unit = $10^{16}$e/m$^2$).

Errors in ionospheric models can have drastic consequences for model users. With respect to OTHR, accurate electron density information is necessary operationally for coordinate registration and frequency selection. Not only is accurate electron density needed during operation, it is also crucial in system design, where electron density information can be used to simulate system performance and capacity. For example, empirical electron density models are typically used to assess future radar placement and
orientation, system frequencies, and the expected surveillance area observed by these instruments [Thayaparan et al., 2016; Saverino et al., 2013]. In this way, an accurate electron density model is necessary in order to properly assess the capabilities, performance, and design of future OTHR systems. Currently, one of the main challenges in making such assessments at high latitudes lies in the inaccuracy of current empirical models in these regions and limited local ionospheric observations. Typically, the IRI is used for such purposes [Cacciamano et al., 2009]; however, the previously mentioned errors in the IRI could result in catastrophic system design failings.

For HF communications and OTHR, prediction of the MUF of HF propagation is dependent on the electron density of the ionosphere. Current prediction models including the Voice of America Coverage Analysis Program (VOACAP), Ionospheric Communications Enhanced Profile Analysis and Circuit Prediction Program (ICEPAC), and other IRI/NeQuick ionospheric map-based prediction systems have all been shown to demonstrate significant shortcomings in their application to high latitude regions, largely due to shortcomings in their electron density specifications [Athieno et al. 2015, Athieno and Jayachandran, 2016].

For radiolocation, which uses triangulation with respect to three or more reference receivers, previous studies have shown that significant, large-scale ionospheric gradients in electron density, such as those seen in the Main Ionospheric Trough (MIT) region, can produce significant errors in positioning in their vicinity [Stocker et al., 2007; Warrington et al., 2012]. These gradients have been shown to be significantly mis-modeled by the IRI.
wherein MIT gradients are almost completely smoothed out [Themens et al., 2017a; Karpachev et al., 2016]. These results all suggest that a new model must be specifically developed for the high latitude region.

Since the creation of the IRI, and similarly the NeQuick [Nava et al., 2008], a plethora of data have become available for use in empirical modeling, namely that from new ionosonde deployments in the arctic regions and from radio occultation (RO)-based electron density inversion. These new data sources allow for the modeling of spatial scales that were not available to previous models. Satellite data, in particular, promise to improve the representation of the ionosphere in regions of sparse ground instrument coverage, such as in the arctic regions and over the oceans. It is our belief that, with the deployment of these new satellite missions and the recent development and expansion of the Canadian High Arctic Ionospheric Network (CHAIN), we now have the resources and data sufficient to model the high-latitude ionosphere [Jayachandran et al., 2009].

With the increased industrial and military interest in the Arctic due to diminishing sea ice in this region and limitations on the use of geostationary satellite communications at high latitudes, it becomes crucial that the ionosphere above this region be accurately modeled. Recent deglaciation in Arctic regions is also unveiling new targets for resource exploration, while the melting of Arctic sea ice is opening up transportation corridors and has triggered the need for new seaports and the development of Arctic communities. As a result, development of communications and navigation technologies for Arctic travel, exploration, mining, and infrastructure development will be essential in the coming years.
The following thesis will present a more in-depth look at the impact of the ionosphere on radio signals and on the performance of other ionospheric models at high latitudes, while identifying the main priorities for new model development, before fully detailing the Empirical Canadian High Arctic Ionospheric Model (E-CHAIM), a new empirical ionospheric electron density model that we believe to be a substantial improvement over current standards.

In the following, I will review some basic principles of the high latitude ionosphere and thermosphere in Chapters 2 – 4. In Chapter 5, I discuss the implications of the ionosphere for radio communications, as well as introduce the many instruments used in the construction and validation of the E-CHAIM model. In Chapter 6, I discuss existing standards in ionospheric empirical electron density modeling. Finally, I present and discuss the E-CHAIM model parameterization, features, and performance in Chapters 7 through 11 before discussing future work in Chapter 12. The author was responsible for the work presented in Chapters 7 – 11, as well as the GNSS calibration work, which is presented in Chapter 6.
2 Physics of the Ionosphere and Thermosphere

The ionosphere is a region of free electrons and ions embedded in the upper atmosphere at altitudes between roughly 60km and 2000km. This layer is produced through the ionization of neutral atmospheric species by solar Extreme Ultraviolet (EUV) and X-ray radiation, as well as by energetic particles. It decays due to recombination, often facilitated by interactions with neutral atmospheric species. The interplay between the ionized and neutral atmosphere, combined with further interactions with the Earth’s magnetic field, makes the dynamics of this region of the upper atmosphere incredibly complex and highly nuanced.

A representation of the neutral atmospheric composition in this region is presented in Figure 2.1. The ionized portion of the ionosphere makes up only a minority of the total number density of the region, which is dominated by neutral species even above the altitude of the peak density of the ionosphere (hmF2). Also, as is illustrated in Figure 2.1 and later in Figure 2.2, the lower and upper boundaries of the ionosphere are not well defined, where gradual exponential behavior is seen at both the top and bottom of the ionosphere.
Figure 2.1 Example neutral, ion, and electron number density profiles at ionospheric altitudes above White Sands, New Mexico (32°N, 106°W). Taken from Rishbeth and Garriott [1969].

The vertical structure of the ionosphere is usually classified into a series of “regions” ranging from the D-Region at the lowest altitudes of the ionosphere to the F-Region, which consists of the peak electron density of the ionosphere. These regions are distinguished based on either substantial differences in chemical processes (e.g. the D-Region) and/or by a characteristic layer-like structuring in the vertical electron density profile (e.g. the E-Region and F2-Layer). These regions are illustrated in Figure 2.2, where example electron density profiles are presented for night and daytime conditions during solar minimum and solar maximum at a sub-auroral location. The mechanisms behind the formation of these regions will be discussed in detail in Section 3. Before that, the following sections will discuss some fundamental processes within the ionosphere and thermosphere.
Figure 2.2 Example electron density profiles generated using the IRI at (60°N, 270°E) for local midnight (red) and local noon (black) at low (dashed lines) and high (solid lines) solar activity corresponding to June 2009 and June 2013, respectively.

2.1 Composition of the Neutral Atmosphere

A discussion of the ionosphere cannot proceed without at least a brief treatment of the neutral atmosphere, where the composition of the neutral atmosphere has a significant impact on the vertical structure of the ionosphere, mainly through its impact on the production, loss, and transport of plasma.

The vertical structure of the density of any given neutral atmospheric constituent can be represented by the barometric relationship derived under the assumption of hydrostatic equilibrium in conjunction with the ideal gas law. This relationship is given below

\[ n = n_0 \exp z' \]  \hspace{1cm} (2.1)
\[ z' = \int_{h_0}^{h} \frac{dh}{H} \]  

\[ H = \frac{kT}{mg} \]  

where \( n \) is the neutral density, \( n_0 \) is the neutral density at a reference altitude \( h_0 \), \( h \) is the altitude, \( k \) is Boltzman’s constant, \( T \) is the neutral temperature, \( m \) is the molecular mass of the neutral constituent, \( g \) is the acceleration due to gravity, and \( H \) is the atmospheric scale height. In the derivation of the above relationship production and loss of neutral constituents is neglected. Of course, this assumption is incorrect for some neutral species, such as atomic oxygen and ozone, but may still be used in regions away from the regions of production and loss. For example, at altitudes near and above 100km, molecular oxygen (\( O_2 \)) is dissociated into atomic oxygen (\( O \)) by solar ultraviolet (UV) radiation in the range of 102.7nm to 175.9nm. This process serves as the main source of neutral atomic oxygen at ionospheric altitudes and as a sink for molecular oxygen.

If we assume \( T \) and \( g \) to be constant in altitude, then \( z' \) may instead be written

\[ z' = \left( \frac{h - h_0}{H} \right) \]  

In the lower atmosphere, the neutral gas is well mixed, such that the composition of the neutral atmosphere is largely unchanged with altitude up to the turbopause at \( \sim 100 \)km and such that most neutral gases have the same characteristic scale height. Above the
turbopause, diffusion plays a much more important role in controlling atmospheric composition, such that the scale height of the various neutral species is proportional to the inverse of their individual molecular masses (see Equation 2.3).

The above features are illustrated in Figure 2.3, where we present a more detailed example of the neutral atmospheric composition. One may notice from this figure that atomic oxygen only becomes an appreciable neutral constituent at ~100km altitude and becomes the dominant neutral species throughout much of the region between 150km and 600km. One may also note that above ~100km, the neutral constituents no longer maintain a constant relative composition and the atmosphere broadens significantly. This sudden increase in scale height can be attributed to a sudden increase in the temperature of the neutral atmosphere at this altitude.
Figure 2.3 Neutral atmospheric composition generated by the US Standard Atmosphere, taken from Hargreaves (1992).

In Figure 2.4, we show a standard illustration of the temperature profile of the Earth’s atmosphere. Within the troposphere, the atmosphere cools with increasing altitude with a lapse rate of ~7K/km until ~10km altitude (i.e. the tropopause). Above the tropopause, the stratosphere is characterized by an increase in temperature resulting from the absorption of solar UV radiation by ozone. Above the stratosphere one finds the
mesosphere, where radiative cooling results in a temperature depression. Above the mesosphere we find the thermosphere, within which the majority of the ionosphere is embedded. The thermosphere is associated with a very large, sudden increase in temperature caused by the absorption of solar radiation within the ionosphere. It is through this mechanism that the ionosphere has a profound impact on the structure of the neutral atmosphere. As we will show in the following sections, the neutral atmosphere will, vice versa, have a profound impact on the structure of the ionosphere as well.

**Figure 2.4** An illustration of the atmospheric neutral temperature profile (left) and ionospheric electron density (right) with some traditional nomenclature annotations. This figure is taken from Hunsucker and Hargreaves [2003].
2.2 The Thermospheric Neutral Wind

As the ionosphere is embedded in the neutral atmosphere, it is fair to presume that the neutral atmospheric circulation may have a significant impact on the drift of ionospheric plasma. The nature of this interaction will be discussed in the following section, but we will here introduce the basic circulations and winds within the neutral atmosphere to facilitate our later discussion regarding their impact on the ionosphere.

The momentum equations governing the motion of the thermosphere are largely identical to what one would encounter for the troposphere, where dynamics are governed by advection, pressure gradient forces, Coriolis effects, and viscous drag. There is, however, one major difference for which one must account: ion drag. Ions are highly restricted by the magnetic field, where ion motions are free to move along magnetic field lines but restricted across field lines. The ion drag, induced by the thermospheric wind, can have the effect of generating dynamo electric fields in the ionosphere and thus plays an important role in the dynamics of the ionosphere. Accounting for this interaction, the momentum equations for the neutral atmosphere can be written as

\[
\frac{d\mathbf{U}}{dt} + 2\boldsymbol{\Omega} \times \mathbf{U} = \mathbf{g} - \frac{1}{\rho} \nabla p - \nabla \phi + \frac{\mu}{\rho} \nabla^2 \mathbf{U} - \nu_{ni}(\mathbf{U} - \mathbf{V}_i) \tag{2.5}
\]

where \(\mathbf{U}\) is the neutral wind vector, \(\boldsymbol{\Omega}\) is the earth’s angular velocity, \(\mathbf{g}\) is the gravity vector, \(\rho\) is the neutral density, \(p\) is the pressure, \(\phi\) is a scalar potential associated with
tidal forcing, $\mu$ is the coefficient of molecular viscosity, $\nu_{nl}$ is the ion-neutral collision rate, and $\mathbf{V}_i$ is the ion drift velocity [Rishbeth and Garriott, 1969].

Within the thermosphere, solar heating of the neutral atmosphere causes an increase in thermospheric temperature resulting in a high pressure “bulge” in the illuminated thermosphere. In the absence of ion drag, the dominant neutral circulation within the thermosphere would be via thermal winds and adhere well to the geostrophic approximation, where winds follow pressure isobars. In the presence of ion drag, however, winds are driven across isobars such that there is wind divergence in the solar illuminated thermosphere and convergence in the nightside. This essentially results in warm thermospheric gas moving almost directly toward regions of cold thermospheric gas. This also results in the lifting of the gas in the region of heating. This process is illustrated in Figure 2.5, where thermospheric temperature contours at 300km altitude are plotted with superimposed neutral wind vectors.
Figure 2.5 Neutral wind vectors and temperature contours at 300km altitude generated using the cTIPe full-physics model for December 19th, 2003, at 00UT. Taken from Huba et al. (2014).

By the above mechanism, at high latitudes, the neutral wind is directed in the anti-sunward direction, across the polar cap during equinox periods. During these periods, a Hadley-type circulation pattern exists, where winds drive the neutral gas upward at the equator and poleward. During the solstices, the winter hemisphere experiences an anti-sunward wind while the summer hemisphere experiences an equatorward wind. During these periods, a single global circulation cell exists, where neutral gas is lifted at the summer pole and moves toward the winter pole, uninterrupted at the equator, with a return flow at the lower boundary of the thermosphere.
2.3 Continuity: Production, Loss, and Transport in the Ionosphere

Now that we have introduced the structure and dynamics of the neutral atmosphere we can begin discussing the physical framework that governs the ionosphere. To begin we will first briefly introduce the ionospheric continuity relationship. Conservation of mass and charge dictates that changes in electron density must result either from production, loss, or transport processes. Formally, for a given ion species, this relationship can be written

\[
\frac{\partial N}{\partial t} = q - L - \nabla \cdot (Nv) \tag{2.6}
\]

where \( q \) is the rate of production, \( L \) is the rate of loss through recombination, and \( \nabla \cdot (Nv) \) is the transport of plasma by velocity \( v \). The following sections will describe each term in the above relationship in detail.

2.3.1 Photoionization

For a single species that adheres to the hydrostatic (barometric) neutral atmosphere relationship (Equation 2.1), the photoionization component of the production term of Equation 2.6 may be approximated by the Chapman function
\[ q = \eta \sigma n I = \eta \sigma n_0 I_0 \exp(-z' - \tau_z) \]  
\[ \tau_z = \sigma H \sec \chi n_0 \exp(z') \]  

where \( I \) is the solar radiation flux at height \( h \), \( \sigma \) is the absorption cross section, \( n \) is the neutral number density, \( \eta \) is the ionization efficiency, \( n_0 \) is the neutral number density at a reference altitude \( h_0 \), \( I_0 \) is the incident radiation flux prior to entering the atmosphere, \( H \) is the neutral atmospheric scale height given as \( H = kT/mg \), \( T \) is the neutral temperature, \( g \) is the acceleration due to gravity, \( k \) is Boltzmann’s constant, \( m \) is the neutral molecular mass, \( \tau_z \) is the optical depth, and \( \chi \) is the solar zenith angle. In the above relationship, it has been assumed that \( g \) and \( T \) are constant with altitude. The altitude at which the optical depth equals unity, and thereby the altitude where the Chapman production function reaches a maximum, is plotted for the entire spectrum between 1 nm and 300 nm in Figure 2.6.

**Figure 2.6** Altitude at which the atmospheric optical depth equals unity at zenith. Principle ionizing species are marked with ranges above the height curve. “Air” refers to
all atmospheric constituents and \( \text{O}_2 \) (SR) refers to Schumann-Runge continuum. “Ionization thresholds” refer to the maximum solar radiation wavelength at which the listed atmospheric constituents could be ionized. Taken from Rees [1989].

To better illustrate the behaviour of the Chapman Production Function, it is often convenient to re-express the above Chapman relationships with respect to the maximum production rate at zenith. In this case, the Chapman production function can be written as

\[
q = q_m \exp(1 - z - \sec \chi e^{-z})
\]  

(2.9)

\[
z = \int_{h_{m0}}^{h} \frac{dh}{H} = \left( \frac{h - h_{m0}}{H} \right)
\]  

(2.10)

where \( q_m \) is the maximum production rate at zenith and \( h_{m0} \) corresponds to the altitude of that maximum production rate at zenith. Each term of the above production equation represents a regime of behaviour: when \( z \) is large and negative, the exponential term within the exponent dominates the production function, illustrating a strong solar zenith angle control of production in the region below the peak; when \( z \) is large but positive, the linear term within the exponent dominates, illustrating the exponential decrease in production due to the corresponding exponential decrease in neutral density with altitude in the region above the peak. One may also note that the maximum production rate and the height of maximum production are highly dependent on the solar zenith angle through this relationship. As one tends to higher solar zenith angles, the maximum production rate decreases and its altitude increases. This behaviour is illustrated in Figure 2.7, where we
present a plot of this production function for various solar zenith angles holding the incident radiation intensity constant.

Figure 2.7 Normalized Chapman production function plotted vs. reduced height \( z = (h - h_m)/H \) for various solar zenith angles. Taken from Hunsucker and Hargreaves [2003].

The Chapman function relationships presented here were developed under the assumption of a flat earth. To modify them for a spherical earth, one must simply include a grazing angle term in place of the secant term, where this grazing angle term is given by
\[
    f(\chi) = R' \sin \chi \int_0^{\chi} \exp \left[ R' \left( \frac{1 - \sin \chi}{\sin \varphi} \right) \right] \csc^2 \varphi d\varphi
\]

(2.11)

\[
    R' = \frac{R_E + h}{H}
\]

(2.12)

where \(R_E\) is the radius of the earth [Al’Pert, 1973].

While these relationships explain the photoionization component of production, at high latitudes, one must also take energetic particle precipitation into account.

### 2.3.2 Precipitation

At high latitudes, production can also occur as a result of collisional ionization of the neutral atmosphere by energetic particles (electrons and protons) precipitating from the magnetosphere or from the solar wind. While the exact mechanism governing particle precipitation is beyond the scope of this study, particle precipitation acts as a significant contributor to ionospheric production within the auroral oval and polar cap. Precipitation produces climatologically significant structures in the high latitude ionosphere that we hope to reproduce with our electron density model, E-CHAIM, and thus we cannot neglect their presence.

Particle precipitation in the ionosphere is mostly constrained to the auroral oval region, which maps along magnetic field lines to a region of energetic particles within the
magnetosphere. An illustration of the structure of the magnetosphere is presented in Figure 2.8.

**Figure 2.8** An illustration of the main plasma regions of the magnetosphere and their relation to the magnetic field. From Hunsucker and Hargreaves [2003].

In Figure 2.9 we present an illustration of the typical location and extent of the auroral oval region. Basic characteristics of this region include the tendency for the auroral oval
to be extended toward the nightside, for the oval to be narrow in the dayside, and for the oval to extend and broaden during periods of increased geomagnetic activity.
Figure 2.9 A statistical representation of the auroral oval (shaded) based on all-sky imager data for quiet, moderate, and active geomagnetic conditions. Taken from Kivelson and Russell [1995].
When energetic particles precipitate into the upper atmosphere in the auroral zones they collide with neutral atmospheric particles along their trajectory, ionizing these neutrals if they have sufficient energy. For the domain of our interest (i.e. between 60km and 2000km), precipitating electron energies correspond to the range between a few tenths of keV, which produces ionization within the F-region, and a few dozen keV, which produces ionization within the E and D regions. The relationship governing the ionization rate due to energetic particle precipitation is given by Hargreaves [1992] and Rees [1989] as

\[
q = \frac{F E_p}{\Delta E} \Lambda \frac{\rho}{r_o \rho_o}
\]  

(2.13)

where \( F \) is the incident particle flux, \( E_p \) is the incident particle energy, \( \Delta E \) is the ionization energy required to produce an ion-electron pair, \( \rho \) is the neutral mass density, \( r_o \) is the depth at which the particle stops, \( \rho_o \) is the neutral mass density at \( r_o \), and \( \Lambda \) is a normalized energy dissipation function that is dependent on the fractional depth of the particle with respect to \( r_o \). For ions, an additional term should be added to Equation 2.13 to accommodate the effect of momentum transfer and excitation.
Figure 2.10 Vertical profiles of ionization rates for energetic electron (top) and proton (bottom) precipitation for various initial energies. From Watson [2016].

The vertical distribution of the ionization rate is illustrated in Figure 2.10 for electrons and protons with various initial energies. As the particles descend into the upper atmosphere, collision rates increase by virtue of the exponentially increasing neutral density with decreasing altitude. This increase in collision rate results in a corresponding
exponential increase in ionization rate with decreasing altitude; however, each collision acts to remove energy from precipitating electrons until such an altitude whereat the particles no longer have sufficient energy to ionize the neutral atmosphere. Higher energy particles thus penetrate deeper into the atmosphere. While the vertical distribution of ionization is governed by the energy of the precipitating particles, the rate of ionization is ultimately controlled by the particle flux; thus, low energy particles, which produce less ionization individually, may still produce significant ionospheric structures at F-region altitudes if particle fluxes are sufficiently high.

Energetic protons behave in a similar manner to electrons; however, the increased mass and collision rate of protons results in slightly different vertical distributions from that of electrons (seen in Figure 2.10) and a tendency for proton precipitation to occur equatorward of the corresponding electron precipitation regions. An illustration of the latitudinal distribution of precipitation regions is provided in Figure 2.11.
Figure 2.11 Particle precipitation as a function of geomagnetic latitude. Taken from Kelley [2009].

2.3.3 Recombination

The main loss of ionization in the ionosphere is the result of recombination, whereby electrons and ions directly or indirectly recombine into neutral species. While direct recombination is rather straightforward in that the ions and electrons recombine through collisions, indirect recombination is the process whereby recombination is first facilitated by some other prior chemical process.
For direct recombination of an ion with an electron, such as through dissociative recombination, the loss term of Equation 2.6 can be written as

\[ L = \alpha N_{\text{ion}} N_e \]  

(2.14)

where \( \alpha \) is the recombination coefficient, \( N_{\text{ion}} \) is the ion number density, and \( N_e \) is the electron number density. We shall term this type of process as an alpha-process. In regions of the ionosphere where molecular ions are the dominant species (lower ionosphere) or where molecular neutral species are sparse (topside), an alpha-type process is the dominant recombination reaction.

In regions where neutral densities are comparably low and atomic ions are the dominant species (such as in the F-region), the fastest recombination reaction generally follows a two-step process, whereby a chemical reaction occurs prior to an alpha-type process. In these regions, this chemical reaction is often the limiting reaction and thus controls the recombination rate. For these chemical reactions, such as charge exchange with neutral species, the rate of reaction is given as

\[ L = \beta N_{\text{ion}} \]  

(2.15)

where \( \beta \) is the reaction coefficient. We shall term this type of process as a beta-process. It is through the interplay between alpha and beta recombination processes and recombination rates that neutral atmospheric dynamics contribute to the vertical
structuring of the ionosphere. This is discussed in detail in Sections 3.1 - 3.4, where we discuss these recombination processes in the context of each vertical regime within the ionosphere.

2.3.4 Transport

Ionospheric plasma transport is perhaps the most complicated component of ionospheric behaviour and dynamics. Diffusion, ExB drift, and neutral wind drag all play an important role in the vertical and horizontal structure of ionospheric electron density. The equation of motion for an ionospheric plasma may be written as the following for both electrons and ions

\[
Nm \left( \frac{\partial}{\partial t} + \mathbf{v} \cdot \nabla \right) \mathbf{v} = Nm g \pm eN (\mathbf{E} + \mathbf{v} \times \mathbf{B}) - \nabla p - Nm \vartheta (\mathbf{v} - \mathbf{u})
\]  

(2.16)

where \( e \) is the charge of an electron, \( m \) is the ion or electron mass, \( \mathbf{v} \) is the electron or ion velocity, \( \mathbf{g} \) is the acceleration due to gravity, \( \mathbf{E} \) is the electric field vector, \( \mathbf{B} \) is the magnetic field vector, \( p \) is the electron or ion gas pressure, \( \vartheta \) is the electron/ion-to-neutral collision frequency, and \( \mathbf{u} \) is the neutral wind velocity. The + in the above relationship is used in the case of positive ions and the – is used in the case of electrons.
2.3.4.1 Diffusion in the Absence of a Magnetic Field

Diffusion in an ionospheric plasma is slightly different from that of a neutral gas in that ionospheric constituents are composed of electrons and ions, which pull on one-another by virtue of coulomb forces and follow the orientation of the magnetic field. As these two constituent gasses experience different gravitational forces by virtue of their mass, gravity will act to separate the gasses by mass. Unlike a neutral gas, the coulomb force between the electrons and ions, however, acts against this gravitational sorting, further complicating the diffusion of the gas as a whole. In this situation an electric field is generated that forces ions and electrons to the same diffusion rate. In the absence of a magnetic field, the diffusion velocity can be written as

\[
\mathbf{v} = \frac{1}{N m_i \nu_i} \{ \nabla [N k (T_i + T_e)] \}
\]

(2.17)

where \( m_i \) is the ion mass, \( \nu_i \) is the ion collision rate, \( T_i \) is the ion temperature, and \( T_e \) is the electron temperature. This relationship is derived under the assumptions that the mass and collision rate of electrons is negligibly small with respect to that of the ions, that the electrons and ions are moving at the same velocity, and that the electron and ion number densities are the same.

In the presence of a magnetic field, the diffusion of the gas becomes anisotropic, such that the diffusion no longer follows the direction of the density gradient and is instead
deflected such that, for a vertical density gradient, the diffusion may have a horizontal component. Likewise, in the presence of horizontal gradients, diffusion is restricted across magnetic field lines and may thus have a vertical component. The reduced rate of diffusion across field lines has important implications for the lifetime of anomalous density structures, such as patches, at high latitudes, where recombination may play the dominant role in the decay of such structures, in place of diffusion, because of this restriction.

### 2.3.4.2 Field-Aligned Transport

In the presence of a magnetic field, under similar assumptions as those leading to Equation 2.17, one may find that the component of the ion velocity in the field-aligned direction can be written as

\[
V_i^\parallel = U^\parallel + V_d^\parallel = U^\parallel + D_a \left( H_p^{-1} - H_{de}^{-1} \right)
\]  (2.18)

where

\[
D_a = k \left( T_i + T_e \right)/m_i v_i
\]  (2.19)

\[
H_p = \left\{ \frac{1}{N} \nabla^\parallel N + \left( \frac{N}{T_i + T_e} \right) \nabla^\parallel (T_i + T_e) \right\}^{-1}
\]  (2.20)

\[
H_{de} = \left\{ \frac{m g^\parallel}{k(T_i + T_e)} \right\}^{-1}
\]  (2.21)

\[
U^\parallel = U_x \cos I + U_z \sin I
\]  (2.22)
Here, $U^\parallel$ is the component of the neutral wind along the magnetic field, $g^\parallel$ is the component of gravity along the magnetic field, $V_d^\parallel$ is the field aligned diffusion velocity, $H_p$ is the scale height of the gas, while $H_{de}$ is the scale height of the gas at diffusive equilibrium, and $D_a$ is the ambipolar diffusion coefficient [Rodger et al., 1992]. In this sense, the departure of the gas from diffusive equilibrium is mapped to a diffusion velocity through the diffusion coefficient. For the F-region, where O+ dominates, the above relationship is sufficient; however, at lower altitudes or in the topside, where multiple ion species may make up sizable contributions to the total ion number density, separate equations of motion must be considered for each ion species. A thorough discussion of this type of correction for the H+ and O+ gas of the topside ionosphere is provided in St. Maurice and Schunk [1977]. Regardless, the relationships presented in Equations 2.18 – 2.22 represent a powerful tool for interpreting the impacts of neutral winds and bulk or localized heating on the field-aligned motion of the ionospheric plasma and have been used to explain various depletion features observed within the polar cap [Rodger et al., 1992].

2.3.4.3 Transport Perpendicular to the Magnetic Field

To examine the motion perpendicular to the magnetic field orientation, let us consider a simplification of the equations of motion such that
\[ e(E + v_i \times B) - m_i \vartheta_i (v_i - u) = 0 \] (2.23)

For the component perpendicular to the direction of the magnetic field, the following tensor relationship may be used to solve for the velocity of the plasma

\[ v_{i\perp} = k_{\perp} \cdot F_{\perp} \] (2.24)

\[ k_{\perp} = \begin{bmatrix} k_1 & \pm k_2 \\ \mp k_2 & k_1 \end{bmatrix} \] (2.25)

\[ F_{\perp} = \begin{bmatrix} eE_x + m_i \vartheta_i u_x \\ eE_y + m_i \vartheta_i u_y \end{bmatrix} \] (2.26)

where \( x \) and \( y \) are coordinates perpendicular to \( B \) and \( k_{\perp} \) is a subset of the mobility tensor with \( k_1 \) corresponding to the component parallel to \( E \) and \( k_2 \) corresponds to the component perpendicular to \( E \) [Rishbeth and Garriott, 1969]. The upper sign corresponds to that for positive ions and the lower sign corresponds to that for electrons. The components of the mobility tensor may be written as

\[ k_1 = \frac{1}{eB} \frac{\vartheta \omega_B}{\vartheta^2 + \omega_B^2} \] (2.27)

\[ k_2 = \frac{1}{eB} \frac{\omega_B^2}{\vartheta^2 + \omega_B^2} \] (2.28)

where \( \omega_B \) is the gyrofrequency. Interestingly, the electron and ion mobility tensors may be related to the ionospheric conductivity tensor such that
\[ \sigma = Ne^2(k_l + k_e) \]  \hspace{1cm} (2.29)

where \( \sigma_{(1,1)} \) is the Pederson Conductivity and \( \sigma_{(2,1)} \) is the Hall conductivity.

The ratios of the gyrofrequencies and the collision frequencies in the above relationships allow us to decompose the cross-field transport into regimes. For example, for the case of electrons, the collision frequency is much smaller than the gyrofrequency for virtually all altitudes above the D-region and thus, to a good approximation, the electrons move according to \( \mathbf{v}_\perp = \frac{E \times B}{B^2} \), the so-called \( \mathbf{E} \times \mathbf{B} \) drift relationship, at most altitudes within the ionosphere. For the case of ions, the collision frequency is only much smaller than the gyrofrequency at F-region altitudes and thus this \( \mathbf{E} \times \mathbf{B} \) drift approximation is only truly valid for ions within the F-Region. Below the F-Region, neutral drag becomes a significant consideration. The differences between the ion and electron behaviour here allows for the production of dynamo electric fields.

The above \( \mathbf{E} \times \mathbf{B} \) relationship has interesting implications for the circulation of the F-Region at high latitudes, induced through interaction with the solar wind. In the Earth’s reference frame, the solar wind electric field can be written

\[ E_{sw} = -u_{sw} \times B_{sw} \]  \hspace{1cm} (2.30)
where $\mathbf{u}_{SW}$ is the solar wind velocity and $\mathbf{B}_{SW}$ is the solar wind magnetic field. For a southward Interplanetary Magnetic Field (IMF), this electric field is directed from dawn to dusk and maps along open field lines into the polar cap. At high latitudes, the magnetic field is near vertical, so the resulting $\mathbf{E} \times \mathbf{B}$ drift is directed in the anti-sunward direction, across the polar cap. The speed of this cross-polar cap convection can range from a few hundred to over a thousand meters per second, where the speed is largely governed by the $z$- (north-south) component of the IMF [e.g. MacDougall and Jayachandran, 2001]. The resulting magnetospheric circulation produces a return flow toward the sunlit ionosphere in both the dawn and dusk sectors near the equatorward boundary of the auroral oval. This circulation is illustrated in Figure 2.12 in a geomagnetic local time and magnetic latitude coordinate system both with (a) and without (b) accounting for the rotation of the earth.
The orientation of the IMF has significant implications on the structure and orientation of the high latitude convection pattern, for example: the $B_y$ component of the IMF can affect the orientation of the cross-polar cap convection, such that positive $B_y$ results in a dawnward shift in the location of the cross-polar cap convection pattern, and positive $B_z$ can result in weaker convection or the production of multiple convection cells and thus multiple cross-polar cap flow channels [Förster et al., 2008]. An illustration of the high latitude convection pattern for various IMF orientations is presented in Figure 2.13. For a full discussion of the effect of IMF orientation on the convection pattern we invite the reader to consult Hunsucker and Hargreaves [2003], Hargreaves [1992], or Kelley [2009] for a thorough summary of these effects.
In the following section we will examine the relative roles of production, loss, and transport on the general vertical structure of the ionosphere.
3 The Vertical Structure of the Ionosphere

As illustrated in Figure 2.2 the vertical structure of the ionosphere is composed of various regions and layers, with each of these regions/layers being governed by different physical processes and dynamics. We will here use the principles developed in the previous section to interpret the behaviour of these various regions/layers.

3.1 The D-Region

The D-Region is the lowest region of the ionosphere, located between approximately 60 km and 90 km altitude, and features a much denser accompanying neutral atmosphere than other regions of the ionosphere with higher proportional abundances of molecular species. This region is produced largely by photoionization of NO by solar Lyman α radiation; however, it is also significantly enhanced by hard X-ray ionization of a broad range of neutral species (particularly N₂ and O₂) during periods of high solar activity. At its lower boundary, ionization may also be supplemented by Cosmic Rays. At its upper boundary these sources of ionization may also be supplemented by photoionization of N₂, O₂, and NO by solar EUV radiation. In the case of N₂⁺ ions, a charge exchange reaction occurs,
\[ N_2^+ + O_2 \rightarrow N_2 + O_2^+ \]  \hspace{1cm} (3.1)

which rapidly depletes \( N_2^+ \) ions in favour of \( O_2^+ \) ions. This produces a situation whereby the D region is largely composed of \( \text{NO}^+ \) and \( \text{O}_2^+ \) ions above \( \sim 75 \) km. Recombination at these altitudes is thus mainly the result of dissociative recombination, whereby diatomic ion species dissociate into neutral atomic species upon combining with an electron. An example of this process for \( \text{NO}^+ \) ions is given as

\[ \text{NO}^+ + e \rightarrow N + O \]  \hspace{1cm} (3.2)

Below \( \sim 75 \) km, the situation is far more complicated, where various hydrates make up the dominant ion species and negative ions serve an important role in the recombination process. Chemical models of this region involve a complex interplay of neutrals with many positive and negative ions, where some of the more intricate models include as many as 50+ different positive and negative ions [Hunsucker and Hargreaves, 2003].

Beyond photoionization and Cosmic ray ionization, at high latitudes the D-Region may also be enhanced by the ionization of neutral species by energetic electrons precipitating from the magnetosphere with energies greater than \( \sim 50 \) keV. This is illustrated in Figure 2.10, where precipitating high energy electrons and protons lead to substantial ionization rates at D-Region altitudes.
As the E-CHAIm model does not provide a D-Region representation, we shall truncate our discussion of the chemical processes within the D-Region here but invite the reader to consult the works of Hunsucker and Hargreaves [2003] or Kelley [2009] for a more in-depth discussion of D-Region aeronomy and dynamics.

3.2 The E-Region

The E-Region is found directly above the D-Region between altitudes of 90km and 150km. This region of the ionosphere can be separated into two types: the primary E-region and sporadic-E.

While sporadic-E makes up a significant contribution to E-region dynamics at high latitudes, its mechanisms for production are complex and multifaceted. It can be produced through precipitation of high energy electrons and protons in the auroral oval and polar cap or by shearing instabilities of metallic ions [Hunsucker and Hargreaves, 2003; MacDougall et al., 2000a/b]. Due to its complex nature and the fact that ionosondes -standard ionospheric sounding instruments- are not capable of producing full electron density profiles in the presence of sporadic-E, we will limit our discussion of the E-region to the primary E-region, which is the only component currently being represented in our model.
The primary E-region is produced through the photo-ionization of virtually all main neutral species by X-rays of wavelengths between 1nm and 10nm, as well as by the photo-ionization of O_2 by EUV radiation at wavelengths between ~80nm and 100nm. Subsequent chemical reactions result in the dominant ion species being O_2^+ and NO^+, like the upper D-region. Loss in this region is dominated by dissociative recombination (ex: Equation 3.2) of O_2^+ and NO^+. In the absence of sporadic-E, this region is well represented by chapman theory and photo-chemical equilibrium.

In Figure 3.1 we present example electron density contours from May 6th, 2013, at the Resolute Incoherence Scatter Radar, located within the polar cap. In this figure, one may note the presence of both the primary E-region during daytime periods (e.g. 10UTC to 22UTC), as well as sporadic-E layers following sunset.
Figure 3.1 Electron density contours from the North face of the Resolute Incoherent Scatter Radar (RISR-N) on May 6th, 2013, from all available beams. Note the daytime enhancement of the E-region (10UT - 02UT) and the sporadic-E structures that form near sunset (22UT - 04UT). For reference, local noon is at 17:45UT.

3.3 The F-Region

The F-Region is found above the E-region, contains the ionospheric electron density maximum, and makes up the majority of the ionospheric electron content. This region is mainly formed through the photoionization of atomic oxygen and is thus mostly made up
of O+ ions. The F-Region can often be separated into two separate features, the F1-Layer or -Ledge and the F2-Layer.

3.3.1 F1-Layer

The F1-Layer is a feature of the lower F-Region electron density profile characterized by an inflection in the profile shape, illustrated in Figure 2.2. This layer is located near the height of maximum photoionization and, interestingly, is not always present. Like the E-Region, however, the F1-Layer is well approximated by a Chapman Function, implying that it can be well represented without taking transport into consideration.

In the bottomside of the F-Region, there exists a transition between alpha- and beta-dominant recombination processes. In the lower ionosphere, molecular ions are abundant; thus, recombination is largely regulated by the rate of an alpha-type, dissociative recombination process. As one tends to higher altitudes, atomic ions, such as O+, serve as the dominant ion species; thus, because a direct recombination reaction of O+ and electrons is slow, a beta-type reaction (ex: Equation 3.1), where charge exchange is undertaken between atomic ions (O+) and molecular neutrals (O2, N2, NO), takes place prior to the main dissociative recombination process (ex: Equation 3.2). Because this beta-process serves as the limiting reaction, the rate of recombination in this region can be taken as the rate of this beta-reaction (see Equation 2.15). This change between alpha-
and beta-controlled recombination processes is of crucial importance to the formation of the F1-Region.

In an alpha-type process, the overall recombination rate is proportional to the electron and ion number density; thus, when photoionization increases the electron and ion number density, it correspondingly increases the recombination rate (see Equation 2.14) in a sort of feedback manner. However, for higher altitudes, where a beta-type process is the limiting reaction, the recombination rate (Equation 2.15) is dependent on the ion and neutral number density; thus, the recombination rate in this case is far less dependent on the amount of ionization and tends to decrease exponentially with increasing altitude as a result of decreasing neutral density. This is particularly limiting as the abundance of molecular neutral species decreases at a faster rate with altitude than that of atomic oxygen, the main photoionized species, simply by virtue of their mass.

This means that there is a reduction in the recombination rate in the transition region between alpha- and beta-dominated recombination processes, where the beta process region has a slower recombination rate than that of the alpha process region. If the altitude of the maximum photoionization is below the altitude of this transition, electron density will peak at the location of the photoionization maximum only to decrease above this altitude, as one would expect of a typical Chapman Function. Electron density thus continues to decrease with altitude until it reaches the beta process region and recombination rates decrease, allowing net electron density to increase. If the altitude of maximum photo-ionization is above the beta-to-alpha process transition region, then no
such decrease in electron density will appear and thus there will be no inflection in the profile above the F1-peak. These two cases are illustrated in Figure 3.2.

**Figure 3.2** Illustrations of the principles behind the formation of the F1-Layer/Ledge. The case on the left is for a situation where the alpha-to-beta transition is above the location of maximum photoionization and the case of the right is for when this altitude is below the altitude of maximum photoionization. In terms of nomenclature, references to “Attachment-type” refer to beta-type processes, while references to “recombination-type” refer to alpha-type processes. Taken from Yonezawa [1965].

Since the altitude of maximum production is dependent on the solar zenith angle, such that high solar zenith angles result in higher maximum production altitudes (see Section 2.3.1), the F1-Layer/Ledge does not occur at high solar zenith angles, where the altitude of maximum production is above the beta-to-alpha transition height. As solar zenith angles decrease approaching local noon, the height of maximum production decreases.
and an F1-Layer/Ledge may form, appearing to migrate to lower altitudes as one approaches local noon. Because of this solar zenith angle dependence, the F1-Layer/Ledge is often not observed during the winter, where the solar zenith angle remains high throughout the day. A statistical representation of the occurrence rate of the F1-Layer/Ledge versus solar zenith angle is presented in Figure 3.3. One may notice that the occurrence rate of an F1-Layer/Ledge decreases at high solar activity in the midlatitude case of this figure. This occurs because the altitude of the alpha-to-beta transition decreases at high solar activity due to competing processes of thermal expansion and transport of thermospheric neutral populations, and increased photoionization at high solar activity.
Interestingly, while one would expect the maximum of the ionospheric electron density profile to be located coincident with the location of maximum production, that is not the case; in fact, photo-chemical equilibrium alone would suggest that the electron density would increase with altitude and not have an explicit maximum at all, as the
recombination rate decreases with altitude more quickly than the rate of photoionization. This runs contrary to observation, where we see a maximum in electron density above the F1-Layer; thus, it is obviously insufficient to consider only photo-chemical processes in interpreting the vertical structure of the F-Region and we must now take transport into consideration when discussing altitudes above the F1-Layer.

3.3.2 F2-Layer

The F2-Layer contains the electron density maximum of the ionosphere. In this region of the ionosphere, the dominant photoionized species is atomic oxygen, forming O+ ions. As mentioned in Section 3.3.1, recombination above the F1-Layer (and thus the main recombination reaction in the F2-Layer) is achieved through a two-step process, where first O+ ions undertake a charge exchange reaction with molecular neutral species (ex: Equation 3.1), such as O₂ and N₂, producing molecular ions that then undertake dissociative recombination (ex: Equation 3.2). In this way, the beta-type charge exchange reaction acts as the limiting step in F2-Layer recombination.

Unlike the E-Region and F1-Layer, the F2-Layer does not feature a photoionization maximum and rather owes its existence to a rapid reduction in recombination rate with increasing altitude with respect to the correspondingly slower decrease in photo-ionization rate with altitude. In this way, electron density increases exponentially with altitude above the F1-Layer until such altitude where diffusion dominates, producing a
subsequent exponential decrease in electron density with altitude. Thus, one may consider the F2-Layer as being photo-chemically controlled below the electron density peak and being diffusion-dominant above the electron density peak. The peak itself can be approximated to be at the altitude where the recombination rate is approximately equal to $D/H^2$, where $D$ is the diffusion coefficient (see Section 2.3.4) and $H$ is the scale height. Equivalently, the peak altitude is associated with the location where the photochemical and diffusion time scales are equal. The delicate equilibrium struck in the F2-layer makes this layer extremely susceptible to changes in thermospheric composition, where the ratio of atomic oxygen to molecular nitrogen is often used as a measure of the role of the thermosphere in driving F2-layer dynamics.

The above mechanism suffices to explain the existence of the F2-layer during periods of photo-ionization; however, unlike the mid-latitude E-Region and the F1-Layer, the F2-Layer is known to persist throughout the night. To partially explain this, one simply has to consider that the recombination rates at F2-Layer altitudes are much lower than at lower altitudes, where molecular neutral densities are greater. Because of this difference in recombination rate with altitude, the F-region will appear to rise in altitude as the night progresses due to the bottom of the profile rapidly recombining.

While this explanation can explain why there is an F-Region at night, it cannot explain why the densities remain relatively high. For a more complete picture, one must again consider diffusion. During the daytime, where ions are being produced within the ionosphere, diffusion serves to transport these ions upward along magnetic field lines.
where they undergo charge exchange with atomic hydrogen and helium, which are then transported into the plasmasphere and may even be transported to the conjugate hemisphere. During the nighttime, when the F-Region is recombining and cooling, the reverse happens and the plasmasphere and conjugate hemisphere serves as a source of ions [Anderson et al., 1998].

3.4 Topside and Plasmasphere

Above the F-Region peak in electron density, one will find the topside ionosphere and plasmasphere. Because the electron density model in this study does not explicitly model the plasmasphere, we will restrict our discussion here to just the topside ionosphere.

In the lower topside, near the F-Region peak, O\(^+\) ions form the dominant ion species; however, as one tends to higher altitudes, charge exchange occurs between O\(^+\) and atomic hydrogen and atomic helium is directly ionized, such that He\(^+\) and H\(^+\) ions form the dominant species. The altitude where this transition between dominant species occurs is often termed the topside transition height. Below this transition height the electron density can be thought to adhere to the O\(^+\) scale height; however, above this transition height, the electron density is governed by the H\(^+\) scale height instead. As hydrogen has a much lower atomic mass than oxygen, its scale height is much larger and thus a “kink” can appear in the topside electron density profile at this transition. This process is illustrated in Figure 3.4. This “kink” is particularly evident during nighttime conditions,
when O+ recombines in the F-region and lower topside, and during low solar activity periods.

**Figure 3.4** Ion concentration profiles measured by the Arecibo Incoherent Scatter Radar in Puerto Rico for daytime (top) and nighttime (bottom). Taken from Schunk and Nagi [2009].
4 Other High Latitude Features and Dynamics

In Section 2.3, we introduced the various principles underlying the behaviour of the ionosphere, such as production, transport, and chemistry. We will here begin to put these various principles together in the context of the high latitude ionosphere in order to introduce and explain various important features that we hope to later reproduce using our model. The models presented later in this study will be judged largely on their ability to reproduce characteristic behaviours and features of the high latitude ionosphere. Some of the more crucial of these behaviours and features are described in this chapter.

4.1 Classification of High Latitude Regions

Before proceeding, we shall here first describe the classification of the various regions of the high latitude ionosphere. The high latitude region is often loosely characterized as the region embedded within the magnetospheric convection (see Section 2.3.4.3). Within this region, there exist several sub-domains that are characterized by different dominant dynamics and interactions with the magnetosphere and solar wind. The lower boundary of the high latitude region is delineated by the Main Ionospheric Trough (MIT), which is described in detail in Section 4.3. At latitudes above the MIT, one will find the Auroral Oval, a region that marks the transition from closed to open field lines and is characterized by enhanced particle precipitation (see Section 2.3.2). Finally, enclosed by the Auroral Oval is the Polar Cap. The Polar Cap is a region of “open” field lines,
wherein the Earth’s magnetic field is directly connected to the Interplanetary Magnetic Field.

4.2 Polar Solstices

Aside from interesting dynamic interactions with the earth’s magnetic field, the high latitude region is also characterized by unique features of solar illumination, such that the ionosphere at high latitudes may be sunlit or in darkness for several months at a time. During the summer, the high latitude ionosphere is sunlit all day. During the winter the reverse occurs, such that the high latitude ionosphere may be in darkness all day. These periods are referred to as Polar Daytime and Polar Nighttime, respectively. To illustrate this, we have plotted the solar terminator versus altitude for various latitudes during summer and winter periods in Figure 4.1. During the summer, the ionosphere (above 100km) remains sunlit 24 hours per day, even at 60°N. This state of persistent photo-ionization acts to smooth out ionospheric gradients, resulting in fewer small scale structures and a weakening of the sharp latitudinal gradient in electron density within the Main Ionospheric Trough Region (Section 4.3). During the winter at the very highest latitudes (> 80°N), the ionosphere remains in darkness all day, contributing to the region of depleted plasma, known as the Polar Hole Region (Section 4.3). Interest in the effects of the altitude-dependence of the solar terminator on the behaviour of the ionosphere have recently been renewed [Verhulst and Stankov, 2017].
Figure 4.1 Plot of the solar terminator altitude vs. local time along the 270°E meridian at 50°N (black), 60°N (blue), 70°N (green), and 80°N (red) during the winter and summer solstices. There are no lines for 70°N and 80°N during the summer, as these regions are sunlit 24 hours of the day.

4.3 Main Ionospheric Trough

An important feature of the high latitude horizontal F-region structure is the Main Ionospheric Trough (MIT), as well as the Polar Hole. The MIT is a longitudinally elongated region of depleted plasma most often found in the nightside ionosphere at sub-auroral latitudes and often rotated about the geomagnetic pole toward the morning sector. This feature is most well defined during the winter but can also be found during the equinoxes. The equatorward edge of the MIT features a gradual electron density gradient, while the polarward edge forms a sharp boundary. The features of the MIT are well
illustrated in Figure 4.2, where we present an example of the structure of the MIT as generated through model simulations by Sojka and Schunk [1989].
Figure 4.2 Illustration of the extent and morphology of the MIT for the Northern (right) and Southern (left) hemispheres. Taken from Sojka and Schunk [1989]
The MIT is largely thought to result from plasma stagnation in the dusk sector, where westward convection acts against eastward co-rotation, resulting in plasma remaining in a region absent of photoionization for several hours [Rodger et al., 1992]. This plasma stagnation is assisted by the fact that the equatorward boundary of the MIT tends to be located at, or near, the plasmapause boundary and thus the MIT region is not supplemented with plasma from the plasmasphere. An illustration of the high latitude convection pattern accounting for corotation and with the region of the MIT superimposed is presented in Figure 4.3.

**Figure 4.3** Illustration of the relative location of the MIT, Auroral Oval, and Polar Hole with respect to the high latitude convection pattern. Taken from Kelley [2009].
A dayside trough may also be observed during some UT periods; however, the mechanisms for its production differ from those of the nightside trough [Sojka et al., 1985]. Poleward of the trough is an auroral region enhancement in plasma density thought to be produced by particle precipitation [Rodger et al., 1992; Sojka and Schunk, 1989] and convection of dayside plasma into polar regions [Middleton et al., 2008].

In Figure 4.2 and Figure 4.3 you will also note the appearance of the Polar Hole. This region of depleted plasma is thought to be produced as a result of localized ionospheric heating due to large differences between the ion and neutral velocities, subsequent ion outflow, and the significant increase in O+ recombination rates with increased ion temperature [Rodger et al., 1992]. During the winter, the auroral region enhancement, due to low-energy precipitation, is substantially larger than the background density; thus, a polar hole may also exist purely as a consequence of auroral enhancement. The location of the polar hole about local midnight has been shown to depend on IMF $B_y$, where the polar hole will occur in the midnight-dawn sector for negative $B_y$ and in the midnight-dusk sector for positive $B_y$ [Sojka et al., 1991].

### 4.4 Behaviour during Geomagnetic Storms

The ionospheric response to geomagnetic activity is composed of a complicated combination of competing processes. We will here introduce and discuss only a handful of these interactions so as to provide context regarding behaviour we shall see later in this
study. The mechanisms that lead to the main ionospheric response to geomagnetic activity, discussed herein, are via: thermospheric wind and composition changes, and changes in electrodynamic convection. The following discussion will provide a very brief overview of the impact of these phenomena to facilitate the interpretation of our modeling results; however, readers are directed to specific studies and reviews of these subjects for more detailed discussions. We will here neglect the impact of increased high energy precipitation, as the impacts of this are not represented in the present E-CHAIM model.

4.4.1 Thermospheric Effects

During periods of enhanced geomagnetic activity, strong electric fields, generated in the E-region, act to joule heat the thermospheric neutral gas [Buonsanto, 1999]. This heating, along with heating by precipitation, results in the generation of traveling atmospheric disturbances that propagate upward and toward the equator [Kintner et al., 2007]. If heating is sufficiently prolonged, a hemispheric circulation, similar to a Hadley cell, whereby there is lifting at the poles with equatorward transport aloft, may develop. The circulation change due to increased geomagnetic activity is illustrated in Figure 4.4.
Figure 4.4 Meridional cross sections of thermospheric winds, demonstrating the effect of high-latitude heating during geomagnetic storms on neutral atmospheric circulation. Situation a) corresponds to quiet conditions, b) corresponds to modest heating ($10^{11}$ J/s), and c) corresponds to storm periods with substantial auroral heating ($10^{12}$ J/s). Taken from Hargreaves [1992].

The upward transport at the poles results in a substantial change in the composition of the neutral atmosphere at F-region altitudes in high latitude regions. By virtue of a gas’ scale height dependence on the molecular mass, the vertical gradient in molecular gas species, such as O$_2$ and N$_2$, is much steeper than that of atomic species, such as atomic oxygen; thus, vertical winds will disproportionately result in an increase in the relative abundance of molecular species compared to atomic species. Since atomic oxygen is the main ionized species at F-region altitudes and molecular species are the main facilitators of plasma loss through recombination, this transport results in a large increase in recombination and a relatively smaller increase in photoionization-based production. The net impact of this heating is thus a significant increase in the recombination rate relative
to production and consequently a substantial decrease in the F-region electron density at high latitudes during the main phase of a geomagnetic storm that can persist for several days [Buonsanto, 1999]. This is often referred to as a negative ionospheric storm [Buonsanto, 1999]. As the lower ionosphere is produced through photoionization of molecular species, this region is less impacted by vertical thermospheric winds, such that these regions may have higher density than the F2-region during storms. In the topside, thermospheric heating results in increased diffusion and thus a broadening of the topside ionosphere in the heating region [Kintner et al., 2007].

At latitudes outside the region of upward thermospheric winds, the equatorward thermospheric wind will also have a significant impact on the state of the ionosphere. Due to the upward-slanted magnetic field lines at mid-latitudes, an equatorward neutral wind has the effect of dragging ions upward along the magnetic field [Werner et al., 1999; Jones and Rishbeth, 1971]. This upward displacement of the F-region ionosphere moves the ionosphere into a region of lower neutral density and thus lower recombination rate [Jones and Rishbeth, 1971]. Ultimately, this equatorward neutral wind creates a so-called positive storm effect on the mid latitude ionosphere, where densities increase by virtue of decreased recombination.

Localized heating by precipitation or joule heating and localized regions of fast plasma convection may also produce high latitude trough structures; however, the scale of these structures will often be too small to be resolved by empirical models and the complicated and localized nature of their production necessitates very careful treatment of the physics.
if one wishes to model these structures. Other storm-time structures not discussed here include polar cap and auroral arcs produced via high energy precipitation.

### 4.4.2 Electrodynamical Effects

During disturbed periods with high geomagnetic activity and southward IMF $B_z$, the southward expansion of the auroral oval, and thereby the high latitude convection pattern, has a substantial impact on the structure of the ionosphere. These impacts can be lumped into two main categories: the southward propagation of the main ionospheric trough; and the transport of plasma from mid and low latitudes toward high latitudes. In the first case, the location of convection-induced plasma stagnation, associated with the MIT, migrates southward, drawing the trough with it [Werner and Prolss, 1997]. In the second case the expansion of the auroral oval sees the high latitude convection pattern reach regions of abundant dayside plasma at mid latitudes. In the convergence region in the cusp, dense plasma from this region is drawn into the convection pattern toward high latitudes resulting in a region of Storm Enhanced Density (SED) at mid latitudes and a Tongue of Ionization or plasma patches at high latitudes. An example of these features is presented in Figure 4.5. Complete treatments of these phenomena are respectively available from the following sources: Werner and Prolss [1997] (Trough Migration); Anderson et al. [1988], Lockwood and Carlson [1992], Sojka et al. [1994] (Patches); Kelley et al. [2004], Foster et al. [2005] (SEDs and Tongue of Ionization).
Figure 4.5 IDA4D estimates of vertical TEC over the region of EMPIRE analysis for 20 November 2003 at (a) 17:10, (b) 17:30, (c) 17:50, (d) 18:10, (e) 18:30, and (f) 18:50 UT. Taken from Huba et al. [2014].
4.5 Winter and Semi-Annual Anomalies

Anomalies in the ionosphere generally refer to departures from solar-controlled behaviour in the F2-layer. The Winter Anomaly is the observed pattern where electron densities in the F2-region are greater than those during the summer. The Semi-Annual Anomaly is the situation whereby the electron density in the F2-layer is greatest in the equinoxes. The typical locations associated with this behaviour are illustrated in Figure 4.6.

![Figure 4.6](image)

**Figure 4.6** Maps of the distribution of Winter and Semi-Annual Anomaly behaviour in foF2 at low (bottom), moderate (top right), and high (top left) solar activity. Note that (A) and (B) distinctions regard the intensity of the winter anomaly, where (A) corresponds to
winter anomaly behaviour with $f_0F_2$ less than 2 MHz greater than that during the equinoxes and (B) corresponds to winter anomaly behaviour with $f_0F_2$ more than 2 MHz greater than during the equinoxes. Black dots represent the geomagnetic poles. Taken from Rishbeth [1998].

These anomalies originate from seasonal variations in neutral composition due to seasonal variations in thermospheric circulation. Examining the quiet time behaviour of the thermospheric meridional circulation, illustrated in Figure 4.4, one will note that there is a downward thermospheric wind at high latitudes during the equinoxes and in the winter hemisphere. This downward thermospheric wind acts to increase the O/N$_2$ ratio in the F-region in the reverse manner discussed in Section 4.4.1, where the steeper vertical gradient in N$_2$ results in more N$_2$ loss than atomic oxygen loss due to downward transport [Rishbeth et al., 2000; Millward et al., 1996]. In the summer hemisphere, we see the same behaviour as we noted for geomagnetic storms, where there is upward neutral atmospheric transport at high latitudes and thus a decrease in the O/N$_2$ ratio. As discussed in Section 4.4.1 the increased/decreased O/N$_2$ ratio results in increased/decreased F-region electron density [Fuller-Rowell, 1998]. This combined with the seasonal variation in solar illumination creates the following net impact:

**Winter:** Increased electron density due to increased O/N$_2$ but weak or no solar illumination.

**Summer:** Decreased electron density due to decreased O/N$_2$ but high solar illumination.

**Equinox:** Moderately increased electron density from increased O/N$_2$ and moderate solar illumination.
Thus, the strength and orientation of the thermospheric winds can modulate the seasonal behaviour of the F-region electron density. Depending on the local behaviour of the thermosphere, some regions experience no anomaly while others experience winter and semi-annual anomalies. Generally, the weaker thermospheric circulation at low solar activity results in weaker anomaly behaviour that is often located closer to the equator than during high solar activity periods. As solar activity increases, thermospheric winds strengthen and anomaly behaviour dominates globally [Yonezawa, 1971, 1972].

4.6 Solar Activity Variability

Variations in solar flux have a significant impact on the ionosphere, resulting in ionospheric variations on solar cycle time scales. This is illustrated in Figure 4.7, where we have presented NmF2 from global ionosonde observations at upper-mid latitudes and corresponding solar F10.7 flux, a proxy for photo-ionizing EUV flux. Note the increase in NmF2 associated with corresponding increases in solar flux.
Figure 4.7 Top: NmF2 from ionosondes between 60° and 70° magnetic latitude from 1970 to 2016. Bottom: F10.7 solar flux corresponding to the periods with ionosonde data. See Section 8.1 for a description of the dataset used in generating this figure. (Generated by the author using the dataset discussed in Section 8.1)

Despite solar flux having a significant impact on photoionization, its relationship to the plasma density within the ionosphere is not simply a direct linear relationship within the F-region. Increased photoionization results in thermospheric changes, such as increased thermospheric temperature, and thus an expansion of the neutral atmosphere, as well as strengthening of meridional winds with increased solar activity [Ma et al., 2009]. It is expected that these changes in the neutral atmosphere act to counteract the increased photoionization from increased solar flux through a substantial decrease in F-region O/N2
with increasing solar activity [Liu et al., 2006]. Figure 4.8 demonstrates this trend in O/N₂ by plotting its variation against solar radio flux.

**Figure 4.8** Plot of O/N₂ ratio vs. F107 flux (black) and 3-month smoothed F10.7 flux (red) at 300km altitude in January (left) and June (right) between 1961 and 2016 for (65°N, 260°E). Neutral densities were generated here using the MSIS-E-90 climatological neutral atmosphere model.

This ultimately results in a saturation effect at high solar fluxes, whereby increasing solar flux does not produce a significant increase in F-region plasma density. This tendency is complicated by the suspected nonlinear behaviour of solar flux proxy indices, such as F10.7, with true EUV flux. Balan et al. [1996] and Kane [1991] attempted to demonstrate that this saturation effect in fact does not exist when EUV flux is used instead of a proxy index at North American midlatitudes. Nonetheless, recent studies have demonstrated that this saturation effect persists, even when measured EUV fluxes are used, when looking at NmF₂ measurements in multiple geographic regimes and varying local times.
Ma et al., 2009; Liu et al., 2006; Lakshmi et al., 1988]. The saturation behaviour is thus suspected to vary wildly with geographic location, season, and local time, due to its expected dependence on thermospheric composition changes.
5 Instrumentation, Indices, and Coordinate Systems

This thesis will make use of virtually every ionospheric radio remote sensing instrument available at high latitudes including: ionosondes, topside sounders, Incoherent Scatter Radars (ISRs), and GNSS instruments. Each of these instruments has strengths, weaknesses, and points of caution that one must have a comprehensive understanding of prior to using their data. In this chapter, we present the theory and data processing methods associated with each of these instruments. I here begin in Section 5.1 with a summary of the impacts of the ionosphere on electromagnetic signals, which forms the basis for many of the remote sensing techniques used in this study. I then briefly describe the techniques used to infer ionospheric electron density information from ionosonde, topside sounder, ISR, and GNSS in Sections 5.2, 5.3, 5.4, and 5.5, respectively. Section 5.5.2 contains a summary of the author’s work on GNSS receiver calibration, originally presented in Themens et al. [2013] and [2015]. Following this, the chapter concludes with a discussion of the various geomagnetic indices and coordinate systems used by the E-CHAIM model in Sections 5.6 and 5.7, respectively.

5.1 Ionospheric Effects on Radio Waves

This study proposes a set of empirical electron density models aimed at improving radio communications and navigation forecasting capabilities in Arctic Regions. These models themselves are built on a large archive of data derived from UHF navigation systems, HF
radars, and other radio remote sensing instruments. It is thus important that we here introduce several fundamental concepts of ionospheric radio propagation relevant to these systems.

5.1.1 The Complex Refractive Index of the Ionosphere

The complex refractive index of a magneto-ionic medium can be derived through careful consideration of the equation of motion of the medium in combination with Maxwell’s Equations. The derivation of this complex refractive index is straightforward and can be found in several different forms in many texts, such as Ratcliffe [1959], Kelso [1964], Al’Pert [1973; 1990], and Budden [1966]. For this reason, we do not present the derivation of the relationship here, sufficing to present the equation and discuss the assumptions made in arriving at this relationship.

The complex refractive index of the ionosphere is given by the Appleton-Hartree Equation (sometimes also referred to as the Appleton-Lassen Equation), which is presented below

\[ n^2 = (\mu - i\chi)^2 = 1 - \frac{X}{1 - iZ - \frac{1}{2}Y_T^2/(1 - X - iZ) \pm \left[ \frac{1}{4} Y_T^4/(1 - X - iZ)^2 + Y_L^2 \right]^{1/2}} \]  

\[ Z = \frac{\nu}{\omega} \quad Y_L = \frac{\omega H}{\omega} \cos \theta \quad Y_T = \frac{\omega H}{\omega} \sin \theta \quad X = \frac{\omega N}{\omega^2} \]

73
where \( \nu \) is the electron collision frequency with heavy particles (ions and neutrals), \( \omega \) is the angular wave frequency, \( \omega_H \) is the angular electron gyrofrequency, \( \theta \) is the angle between the wave vector and the magnetic field vector, \( \omega_N \) is the angular electron plasma frequency, \( \mu \) is the real component of the refractive index, \( \chi \) is the imaginary component of the refractive index (also referred to as the absorption index), and \( n \) is the complex refractive index.

In the derivation of the Appleton-Hartree Equation, several approximations/simplifications are undertaken; for example, the effects of the magnetic field of the wave itself are neglected and this relationship only holds for cold plasmas, where the pressure gradient force may be neglected and one may essentially treat particles independently. This relationship is also derived under the assumption that the effects of ions on electromagnetic waves within the medium is negligible by virtue of the ionic plasma- and gyro-frequencies being much smaller than that of electrons. The validity of this assumption, however, depends largely on the frequencies of interest. If one is concerned with frequencies in the vicinity of the ion plasma frequency, they would simply need to adjust the above relationship to accommodate one or more ion species. As this is beyond the scope of interest for this study, we shall merely refer the reader to the discussion of the effect of a single ion species presented in Ratcliffe [1959] and the tensor treatment of multiple ion species discussed in Al’Pert [1990]. Both of these texts provide complimentary discussions of the impact of ions on the refractive index of the ionosphere.
In this work, we will largely be concerned with the real component of the refractive index, which relates to the phase and propagation of electromagnetic signals through the ionosphere. The imaginary component of the refractive index relates to the absorption of these signals as they propagate. While this absorption is important for radio communications, particularly at frequencies at the lower end of the HF band and for propagation in the D-Region, its impact is not considered in the remote sensing techniques used in this study; thus, we will only consider relationships governing signals within collisionless plasma. Those curious about the very interesting use of ionospheric radio wave absorption for ionospheric reconstruction may consult the works of Kero et al. [2014] for a detailed discussion.

One may arrive at a collisionless version of the refractive index expression by simply removing the $Z$ terms of Equation 5.1, resulting in the following relationship

$$n^2 = (\mu - i\chi)^2 = 1 - \frac{X}{1 - \frac{1}{2} \frac{Y_T^2}{(1 - X)} \pm \left[ \frac{1}{4} \frac{Y_T^2}{(1 - X)^2} + \frac{Y_L^2}{X} \right]^{1/2}}$$

In this relationship, the refractive index is real everywhere $n^2$ is positive and imaginary everywhere $n^2$ is negative. It is this form of the Appleton-Hartree Equation, and simplifications thereof, that we will be working with in the remainder of this study.
It is often convenient to discuss the Appleton-Hartree relationship in terms of limiting cases in order to conceptualize the behaviour of the refractive index. Here we will first consider a signal that is propagated along the magnetic field (longitudinally) in a collisionless plasma. The refractive index for propagation of this type is given by

\[ n^2 = 1 - \frac{X}{1 \pm Y} \]  

(5.3)

where \( Y = \frac{\omega_H}{\omega} \). In this case, \( n^2 \) is given by a pair of linear functions of \( X \) with roots at \( X = 1 - Y \) and \( X = 1 + Y \). Similarly, in the case of propagation perpendicular to the magnetic field (transverse), we arrive at the following pair of relationships

\[ n_{(u)}^2 = 1 - X \]  

(5.4)

\[ n_{(l)}^2 = 1 - \frac{X(1 - X)}{1 - X - Y^2} \]  

(5.5)

In this case, the second relationship, corresponding to the lower sign solution, retains the same roots as those found for the longitudinal case, but the first relationship is identical to that for propagation in the absence of a magnetic field and has a root at \( X = 1 \).

Through these limiting cases we may set bounds to the behavior of the refractive index. These limiting cases are illustrated in Figure 5.1, where we present the refractive index versus \( X \) for the case of \( Y < 1 \) (i.e. for the case of frequencies at HF and above). The L in this figure refers to the limiting case where propagation is parallel to the magnetic field.
(i.e. longitudinal propagation) and the T refers to the limiting case of propagation perpendicular to the magnetic field (i.e. transverse propagation). All possible values of the refractive index are bounded by these transverse and longitudinal cases. As the Appleton-Hartree Equation has two solutions, the upper sign, typically referred to as the ordinary mode, is given by the vertically shaded region and the lower sign, typically referred to as the extraordinary mode, is given by the horizontally shaded region.
Figure 5.1 The collision-less complex refractive index (a) and its real (b) and imaginary components (c) plotted against $X$ with values of $Y < 1$ for arbitrary angles of propagation. The upper (ordinary) and lower (extraordinary) signs of the Appleton-Hartree Equation are given by the vertically and horizontally shaded regions, respectively. $T$ here refers to the transverse limit and $L$ refers to the longitudinal limit. Taken from Ratcliffe [1959].

5.1.2 The Group Refractive Index of the Ionosphere
In many applications, and particularly in radar remote sensing, it will often become important to determine the group refractive index of the ionosphere. The group refractive index ($\mu'$) is related to the real component of the phase refractive index through the following relationship

$$\mu' = \mu + \omega \left( \frac{d\mu}{d\omega} \right)$$  \hspace{1cm} (5.6)

For propagation in a collisionless plasma and in the absence of a magnetic field, $\mu^2 = 1 - X$; thus, in this case, the phase and group refractive indices are related by

$$\mu'\mu = 1$$

This is of course a highly simplified situation that is only valid to the first order for propagation at UHF and higher frequencies. For HF propagation in the F-region, we cannot neglect the effect of the magnetic field but may still assume a mostly collisionless plasma. In this situation, for positive $n^2$, the group refractive index is given by Shin and Whale [1952] as

$$\mu' = \mu + \frac{1}{\mu D} \left[ 1 - \mu^2 - X^2 + \frac{(1 - X^2)(1 - \mu^2)Y_L^2}{2S} \right]$$  \hspace{1cm} (5.8)

$$S = \pm \sqrt[4]{\frac{1}{4} \frac{Y_T^4}{(1 - X)^2} + Y_L^2}$$  \hspace{1cm} (5.9)
\[ D = (1 - X) - \frac{Y_T^2}{2} + S \] 

The upper sign in the above corresponds to the ordinary mode and the lower sign 
corresponds to the extraordinary mode. The phase refractive index, \( \mu \), used in the above, 
is that corresponding to Equation 5.2, when \( n^2 \) is positive. A detailed discussion of the 
above relationship is provided in Shin and Whale [1952]. A discussion of the impacts of 
collisions on the group refractive index of HF propagation is available in Scotto and 
Settimi [2013], wherein the group refractive index is calculated by numerically solving 
Equation 5.6 using the full Appleton-Hartree Equation for the phase refractive index.

### 5.2 Ionosonde

Ionosondes are perhaps the most important instruments used in the development of E-
CHAIM and are one of the oldest and most reliable methods of sounding ionospheric 
electron density. Ionosondes are HF radars that sound the bottomside portion of the 
ionosphere, providing electron density profiles up to the F2-peak. They accomplish this 
by sweeping through HF frequencies from 1MHz to 20 - 30MHz and measuring the time 
of flight of echoes from the ionosphere. These signals are reflected by the ionosphere at 
the point where \( \mu = 0 \); thus, the ordinary mode is reflected when the signal frequency is 
equal to the plasma frequency of the medium. In this manner, in the absence of collisions, 
the frequency of the reflected ordinary mode signals can be related to the plasma density 
at the point of reflection by
\[ N_e = 1.24 \times 10^{10} \cdot f^2 \]  

(5.11)

where the frequency is in MHz. Through using the speed of light, the echo time of the reflected signal can be related to the “virtual height” as \( h = \frac{ct}{2} \), where \( c \) is the speed of light in a vacuum and \( t \) is the time of flight of the signal. An example of the virtual height profile of ionosonde echoes, called an ionogram, in Moscow, is presented in Figure 5.2.

In this figure, we see two distinct sets of echoes: one related to the ordinary mode (pink/red) and the other related to the extraordinary mode (green).

![Figure 5.2](image)

**Figure 5.2** Ionogram at Moscow ionosonde MO155 on June 21, 2013, at 14:01 UTC. Pink corresponds to the ordinary mode and green corresponds to the extraordinary mode. Superimposed in black is the virtual height trace of the O-mode and the true height profile after inversion. The y-axis is altitude in km and the x-axis is frequency in MHz.
In order to process these ionograms into electron density profiles, one must convert the observed virtual ranges into “real heights”. This is done through inversion algorithms such as that of Titheridge [1985, 1988]. In these methods, the group refractive index is calculated for each point of the profile. The true height is then calculated sequentially, from the bottom up, by integrating the true height group refractive index profile up to that point. The integration is started by assuming that the group refractive index is unity below the lowest recorded virtual height. These methods are generally accurate to within 3km, where errors are largely related to the inability of ionosondes to sound the valley between the E- and F-Regions, thus requiring that one make assumptions about the shape and depth of these valleys [Liu et al., 1992]. Also, for the ordinary mode, Scotto and Settimi [2013] have shown that the assumption of a collisionless ionosphere has little effect on the inversion of ionograms.

The main source of error in ionosonde sounding of the bottomside is related to the tracing of the virtual height profile of the ordinary mode from ionograms. The following challenges make the automated processing of ionograms difficult, often requiring manual interpretation, which can sometimes be a bit of an artform:

1) Small-scale irregularities in the ionosphere can cause broadening of the frequency and height of the ionogram trace, called spread-F [Al’Pert, 1973].
2) Absorption can result in E-region echoes being absent or even result in the loss of F-region echoes at high latitudes during solar X-ray flare events, periods of high energy precipitation, or during periods of very low densities.

3) So-called “Z-mode” traces can contaminate ionograms at high latitudes [Ratcliffe, 1959].

4) Patches, Travelling Ionospheric Disturbances (TIDs), and other mesoscale ionospheric structures can produce multiple O- and X-mode traces in ionograms [Moskaleva and Zaalov, 2013].

In the absence of scaling errors, ionosondes are accurate and reliable instruments for the sounding of the bottomside ionosphere.

5.3 Topside Sounder

Topside sounders are satellite-mounted, downward-facing ionosondes that provide topside electron density profiles down to the F2-peak altitude (hmF2) [Bilitza et al., 2003]. These instruments function in a similar manner as ground-based ionosondes but suffer additional challenges in their processing as a result of being immersed within a plasma environment. Due to these challenges, most topside sounder data has had to be manually interpreted, creating a significant inhibition to processing these datasets. To facilitate further use of these data, a specialized software called TOPside Ionogram Scaler with True height algorithm (TOPIST) was developed [Huang et al., 2002].
5.4 Incoherent Scatter Radar (ISR)

ISRs are capable of determining the electron density profile of the ionosphere from altitudes as low as the lower E-Region and D-Region to well into the F-Region topside or even the lower plasmasphere. They are also capable of determining plasma composition, temperatures, and drift velocities. In this way, these instruments are some of the most powerful tools for studying the ionosphere. Unfortunately, their extremely high construction and operations costs make these instruments prohibitively expensive; thus, only a few ISRs have been deployed since Gordon [1959]’s proposal of the ISR technique. An in-depth review of the ISR method is available from Evans [1969]; however, we shall here nonetheless touch on the basics of the technique.

The ISR technique was originally conceived based on the theory of Thompson Scatter, whereby electromagnetic signals directly cause electrons to oscillate and re-radiate energy at the same frequency. This re-radiated energy, which is proportional to the density of electrons within the scattering volume, could then be detected at the ground. In the case of Thompson Scatter, the scattering cross section is simply that of an electron given by

\[ \sigma_e = 4\pi (r_e \sin \varphi)^2 \]  \hspace{1cm} (5.12)
where $r_e$ is the radius of an electron and $\varphi$ is the scattering angle ($\frac{\pi}{2}$ for backscatter). In this case, the scattering cross section is very small (on the order of $10^{-28}$ m$^2$); thus, high power radars with large apertures are necessary in order to employ this technique.

The above relationship is purely for the case of Thompson Scatter; however, for transmission wavelengths near or larger than the Debye length, shielding by ions results in Landau Damping, significantly altering this relationship. In such a situation, the signal can no longer be thought of as scattering from individual electrons and is rather scattering from very small irregularities in the plasma, principally ion-acoustic waves and electron waves at the plasma frequency. The scattering cross section of the ion-acoustic irregularities is given in Buneman [1962] as

$$\sigma = \frac{\sigma_e}{(1 + \alpha^2)(1 + T_e/T_i + \alpha^2)}$$ (5.13)

where $\alpha = 4\pi D/\lambda$, $D = 69(T_e/N_e)^{1/2}$ is the Debye length, and $\lambda$ is the signal wavelength. For the sake of simplicity, the effects of ion mass and the magnetic field have been neglected here; however, to illustrate these effects on ISR scattering spectra, Figure 5.3 provides example spectra for three different ion species common to the topside ionosphere.
Figure 5.3 ISR spectra for O\(^+\), He\(^+\), and H\(^+\) ions for three values of $T_e/T_i$. From Evans [1969].

In effect, the total area of the ion-acoustic scattering spectrum depends on the electron density, the width of the spectrum is dependent on the ion and electron temperatures, as well as the ion mass, the mean doppler depends on the ion drift, and the depth of the valley at spectrum center depends on the ratio of the ion and electron temperatures. Through fitting these spectra, one can derive the electron temperature, ion temperature, ion mass, and electron density.

Alternatively, one may examine the spectrum of electron plasma waves, often referred to as plasma lines or plasma resonance lines, which are centered at a doppler equal to the
ionospheric plasma frequency within the sampling volume. As with ionosondes, this frequency can be directly related to the electron density. In this way, the plasma line observations of an ISR can be thought of as an ionogram but without the necessity for a virtual height to real height correction and able to sample to topside ionosphere. The electron density derived through plasma line observations can be very precise, as the observed spectrum is solely dependent on the electron density [Vierinen et al., 2017]. This and the ability to sample plasma lines at very high rates has seen plasma lines become a popular tool in examining ionospheric irregularities and mesoscale structures.

5.5 Global Navigation Satellite System (GNSS)

GNSS is a very valuable tool in remote sensing the state of the ionosphere. In the case of ground-based GNSS, global networks of advanced dual-frequency receivers provide the majority of our information on the horizontal structure of the ionosphere. In the case of satellite-borne receivers, Radio Occultation (RO) techniques allow GNSS measurements to provide information on the vertical structure of the ionosphere with unparalleled spatial coverage. We will here describe the techniques used with these instruments to infer information about the ionosphere, which are later applied to form a significant portion of E-CHAIM’s fitting dataset and have been refined as part of this study in preparation for the use of this data in an E-CHAIM assimilation scheme.
5.5.1 Principles of GNSS Ionospheric Remote Sensing

The primary ionospheric parameter derived from GNSS observations is the ionospheric total electron content (TEC). TEC is defined as the total number of electrons within a 1 m$^2$ column along a path through the ionosphere and is measured in TEC Units (TECU), where 1 TECU = $10^{16}$ electrons/m$^2$.

GPS, primarily broadcasting two L-band frequencies (L1 = 1575.42 MHz and L2 = 1227.60 MHz), provides a host of observables pertinent to ionospheric studies; those particularly important to TEC calculation include pseudorange and carrier-phase measurements. Using these observables, one may calculate signal delays, which are related to ionospheric TEC along the ray path of the GPS signal through the refractive index’s dependence on electron density. The time delay of GPS signals, including the effects of the ionosphere and instrumental biases measured by GPS receivers, is converted to pseudorange (code) values for both L-band frequencies via a conversion using the vacuum speed of light. The receiver also measures carrier phase for both frequencies. These pseudorange and carrier-phase measurements can be described by the following:

\[
P_1 = \rho_k^p + c(d_{1,p} - d_k) + I_{k,1}^p + T_k^p + c(d_{k,1} + d_{1}^p) \quad (5.14)
\]

\[
P_2 = \rho_k^p + c(d_{2,p} - d_k) + I_{k,2}^p + T_k^p + c(d_{k,2} + d_{2}^p) \quad (5.15)
\]
\[ L_1 = \rho_k^p + c(dt^p - dt_k) + I_{k,1,P}^p + T_k^p - \phi_{k,1}^p - \phi_k^p + \lambda_k N_{k,1}^p \]  
\[ L_2 = \rho_k^p + c(dt^p - dt_k) + I_{k,2,P}^p + T_k^p - \phi_{k,2}^p - \phi_k^p + \lambda_k N_{k,2}^p \]  

(5.16)  
(5.17)

as in Leick [2004], where \( P_f \) is the pseudorange measured on frequency \( f \), \( L_f \) is the carrier phase measured on frequency \( f \), \( \rho_k^p \) is the geometric range between receiver \( k \) and satellite \( p \), \( dt^p \) and \( dt_k \) are respectively satellite and receiver clock errors, \( I_{k,f,P}^p \) and \( I_{k,f,L}^p \) are the ionospheric code delay and phase advance on signal \( f \), respectively (in meters), \( T_k^p \) is the neutral atmosphere (tropospheric) delay (in meters), \( d_{k,f} \) and \( d_f^p \) are respectively receiver and satellite hardware delays on signal \( f \) code measurements, \( \phi_{k,f} \) and \( \phi_f^p \) are respectively receiver and satellite hardware delays on signal \( f \) phase measurements, \( \lambda_f \) is the wavelength of signal \( f \), and \( N_{k,f}^p \) is the integer phase ambiguity in signal \( f \). It should be noted at this point that \( P_1, P_2, L_1, \) and \( L_2 \) have all been converted to units of meters in the above representative equations for the sake of simplicity; also, multipath and measurement noise are not explicitly noted in the equations.

Taking the difference of \( P_2 \) and \( P_1 \) as well as \( L_1 \) and \( L_2 \), we are left with

\[ P_{k,GF}^p = P_2 - P_1 = I_{k,2,P}^p - I_{k,1,P}^p - c(DCB^p + DCB_k) \]  
\[ L_{k,GF}^p = L_1 - L_2 = I_{k,1,L}^p - I_{k,2,L}^p - c(DPB^p + DPB_k) + n_{k}^p \]  

(5.18)  
(5.19)

89
where \( n_k^p = \lambda_1N_{k,1}^p - \lambda_2N_{k,2}^p \) and \( DCB^p = d_1^p - d_2^p \), \( DCB_k = d_{k,1} - d_{k,2} \), \( DPB^p = \phi_1^p - \phi_2^p \), and \( DPB_k = \phi_{k,1} - \phi_{k,2} \) are, respectively, satellite and receiver differential code and differential phase biases (in units of seconds) [Arikan et al., 2008]. We have thus removed all frequency independent terms, including the tropospheric and geometric unknowns. These difference equations are known as the geometry-free linear combinations of pseudorange and carrier phase, respectively. At UHF frequencies, the magnetic field and collisional terms of the refractive index equations can be dropped such that the phase and group refractive indices are, to the first order, given by the following

\[
\begin{align*}
\mu &= \sqrt{1 - X} \\
\mu' &= \frac{1}{\sqrt{1-X}}
\end{align*}
\] (5.20)

If we then do a Taylor expansion, we arrive at first order linear relationships in terms of the electron density

\[
\begin{align*}
\mu &= 1 - \frac{X}{2} - \ldots \\
\mu' &= 1 + \frac{X}{2} + \ldots
\end{align*}
\] (5.21)

From the integration of these relationships along the ray path, the ionospheric term of Equations 5.18 and 5.19 can be written as

\[
I_{k,f,P}^p = -I_{k,f,L}^p \approx A_s^{TEC} \frac{C_i^p}{f^2}
\] (5.22)
where \( A = 40.3 \) and \( f \) is the signal frequency in MHz. Combining this with Equations 5.18 and 5.19 yields two relations for determining sTEC via GPS observables

\[
P_{k,GF}^p = A \left( \frac{f_1^2 - f_2^2}{f_1^2 f_2^2} \right) sTEC_k^p - c \left( DCB_k^p + DCB_k \right) \quad (5.23)
\]

\[
L_{k,GF}^p = A \left( \frac{f_1^2 - f_2^2}{f_1^2 f_2^2} \right) sTEC_k^p - c \left( DPB_k^p + DPB_k \right) + n_k^p \quad (5.24)
\]

The sTEC determined from pseudorange measurements via the above relation contains no non-instrumental ambiguity but can be extremely noisy, while the sTEC determined from phase measurements via the above relation retains a phase ambiguity but is extremely precise. This phase ambiguity term must be removed if we are to capitalize on the high-precision of phase measurements in our analysis. This is done by a phase-leveling procedure that effectively levels the phase-derived TEC to the pseudorange-derived TEC. The method used in our analysis is a similar procedure to those that can be found in Horvath and Crozier [2007], Arikan et al. [2008], and Makela et al. [2001]. In this phase leveling procedure, we first determine a leveling constant \( W \)

\[
W = \frac{1}{N_{arc}} \sum_{N=1}^{N_{arc}} \left( P_{k,GF}^p - L_{k,GF}^p \right) \quad (5.25)
\]

where \( N_{arc} \) is the total number of measurements over one arc of lock, and \( N \) is an epoch index. In order to minimize the effects of multipath and low elevation-angle noise in the estimation of the leveling constant, we limit the range considered in Equation 6.15 to data
acquired within $10^\circ$ of the peak elevation angle of each arc. We can then substitute

Equations 5.23 and 5.24 into Equation 5.25, resulting in

$$W = \langle cDPB_k^p \rangle - \langle cDCB_k^p \rangle - \langle n_k^p \rangle \quad (5.26)$$

where $DPB_k^p$ and $DCB_k^p$ simply represent the sum of the receiver and satellite biases in
carrier phase and pseudorange measurements, respectively. Since these DCBs and DPBs
are considered constant over periods far longer than an arc [Sardon and Zarraoa, 1997]
and the integer ambiguity is constant over an arc, the terms in Equation 5.26 can be taken
such that,

$$W = cDPB_k^p - cDCB_k^p - n_k^p \quad (5.27)$$

If we then rearrange Equation 5.24 for sTEC and substitute in Equation 5.27, we are left
with

$$sTEC_k^p = \frac{1}{A} \left( \frac{f_1^2 f_2^2}{f_1^2 - f_2^2} \right) \left( L_{k,gr}^p + W + cDCB_k^p \right) \quad (5.28)$$

where the sTEC term can be taken as the phase-leveled sTEC. It should be noted that, in
this manner, the phase-leveled sTEC is independent of the DPB and $n_k^p$ terms previously
found in phase Equation 5.24.
Often, for the purpose of comparing TEC measured from different satellites or at different sites, converting GPS-derived sTEC to equivalent vertical TEC (vTEC) is desirable. We achieve this by first removing satellite and receiver DCBs from GPS-derived sTEC values, where P1-C1 and P1-P2 satellite biases can be retrieved from the University of Bern ftp database at ftp://ftp.unibe.ch/aiub/CODE/ and ftp://ftp.unibe.ch/aiub/BSWUSER50/ORB/, respectively, and receiver biases must be calculated independently through either single-station or network approaches (see Section 5.5.2). GPS sTEC measurements can then be mapped to vTEC via a simple geometric mapping function derived via the single layer ionosphere model (SLIM), pictured in Figure 5.4.

\[ vTEC = sTEC \cdot M(e) \quad (5.29) \]

where \( M(e) \) is represented by the following relationship:

\[ M(e) = \cos(\chi) = \sqrt{1 - \left(\frac{R \cos(e)}{R + h}\right)^2} \quad (5.30) \]

where \( R \) is the mean radius of the Earth, \( e \) is the elevation angle measured from the horizon at the receiver to the ray from the satellite to the receiver, \( \chi \) is the ray’s zenith angle at the intersect of the ray and the ionospheric shell, and \( h \) is the ionosphere thin shell height, often assumed to be 400km. This mapping results in the determination of
vTEC through the ionospheric pierce point (IPP), the point at which the ray from the satellite to the receiver intersects the thin shell ionosphere.

Figure 5.4 Schematic illustration of the geometric projection of slant GPS measurements to vertical. From Themens et al. [2013].

5.5.2 Biases and Limitations

One of the more severe limitations in the use of GNSS-based TEC is the problem of differential receiver biases. In order to provide accurate ionospheric information, GPS measurements must be calibrated to account for both receiver and satellite biases [Warnant, 1997; Rideout and Coster, 2006]. While several techniques exist for determining these biases, one must take care in their application in regions outside their initial design [Lanyi and Roth, 1988; Ma and Maruyama, 2003; Ma et al., 2005; Rideout and Coster, 2006; Arikan et al., 2008]. Recent studies have attempted to characterize variabilities in these biases, estimated through single-station approaches, using real data and simulations [Ciraolo et al., 2007; Mazzella, 2009; Zhang et al., 2009; Brunini and
Azpilicueta, 2010; Zhang et al., 2010; Conte et al., 2011; Coster et al., 2013; Themens et al., 2015]. These studies highlight the need to understand not only the nature of true bias variability but also the impact of the fundamental assumptions made in standard bias-estimation techniques on bias estimation.

There are several different means of determining GPS receiver biases, all of which are expected to demonstrate appreciable errors in their application to the high latitude region. Themens et al. [2013] showed that the least squares (LSQ) receiver-bias-estimation method of Lanyi and Roth [1988] is highly unstable in its application to the high latitude region while used in the same manner as used at mid-latitudes. The minimization of standard deviations (MSD) method of Ma and Maruyama [2003] is also shown to demonstrate appreciable errors while applied in the polar cap region; however, these errors were found to be far less significant than those of the LSQ method [Themens et al., 2013]. While the LSQ and MSD methods can be considered bias-projection-based estimation methods, other methods generally use an ionospheric reference model to estimate biases [Arikan et al., 2008; Keshin, 2012]. In mid-latitude regions, where accurate ionospheric models, such as the International Reference Ionosphere (IRI) and IONospheric EXchange (IONEX) ionospheric maps, are available, reference-based methods demonstrate good performance [Arikan et al., 2008]. In high latitude regions, however, Themens et al. [2013] demonstrates that the IONEX maps suffer significant errors due to a lack of contributing data from these regions and Themens et al. [2014] shows that there are appreciable errors in the IRI-2007’s representation of the polar cap ionosphere. The lack of adequate reference data in this region makes it necessary for
operators to explore alternative receiver-bias-estimation approaches for application in the high latitude region. It should be noted that receiver biases provided by sources, such as the University of Bern, or calculated using any of the above receiver bias estimation techniques generally refer not only to the bias of the receiver itself but also to biases potentially produced by the antenna hardware or cabling.

There are predominantly two popular single-station receiver bias estimation methods: namely, MSD [Ma and Maruyama, 2003] and various permutations of the LSQ method [Lanyi and Roth, 1988]. These methods, like most single station bias estimation approaches, rely on three basic assumptions:

1) The ionosphere is locally horizontally homogeneous.
2) The ionosphere can be roughly approximated as a thin spherical shell above the Earth at a specified altitude, or at the very least, the ionosphere can be represented by some standard vertical profile function.
3) The GPS receiver bias is the only geometry/elevation-independent parameter affecting GPS TEC measurements.

While both methods share these assumptions, they do not necessarily behave in the same manner, as each method has some ability to accommodate violations of their fundamental assumptions. In Themens et al. [2015], we used a series of simplified simulations and measured data to assess the impact that the method assumptions have on the MSD and LSQ bias approaches.
These approaches to bias estimation are built on the fact that receiver biases are not geometrically-dependent properties; thus, when one projects bias-contaminated sTEC to vTEC, one is also erroneously projecting the biases. This leads to elevation-dependent errors in the projected vTEC. Biases can be considered correctly removed when the estimated vTEC no longer contains these erroneous projected biases. This principle is treated slightly differently in each approach, pertaining largely to how it is decided that the resulting vTEC no longer contains projected biases.

The errors resulting from erroneously projecting receiver biases should be different for each satellite measurement; thus, there will be a bias-dependent spread in the derived vTEC values if there are receiver biases in the sTEC measurements. In the MSD approach to bias estimation, it is assumed that, if the bias is correctly removed, this spread in vTEC should be minimized. By iterating through a series of test biases, removed before vTEC projection, and summing the standard deviation of vTEC measurements at each time step, we can identify the correct bias as that which minimizes the summed standard deviations (SSD). An example of this process is presented in Figure 5.5 where we have iterated through a series of biases between 0.0 and 50.0 TECU in 0.1 TECU steps over an integration period of 24 hours for September 19, 2009, at Resolute. This example demonstrates a clear minimum in summed standard deviations at a bias of 34.8 TECU (marked by a red star).
Figure 5.5 Example of the summed standard deviations versus test bias at Resolute for 24-hour integration on September 19, 2009. The minimum summed standard deviations is marked by a red asterisk at the receiver bias value of 34.8 TECU. From Themens et al. [2015].

In the LSQ receiver bias estimation method, one models the ionospheric vTEC by a simple polynomial, such as the following:

\[
vTEC = a + b\theta + c\varphi + d\varphi\theta + e\theta^2 + f\varphi^2
\]

(5.31)

where \(a, b, c, d, e\) and \(f\) are fitting coefficients, \(\theta\) is latitude, and \(\varphi\) is local time. The sTEC measured by the receiver can then be expressed as

\[
sTEC = \frac{vTEC}{M(e)} + DCB
\]

(5.32)
where $DCB$ refers to the receiver bias. By fitting GPS sTEC measurements to the above relationship by a linear regression, one can extract the expected receiver bias.

While typical applications of this technique use short integration periods (2-4 hours), this application has been shown to result in highly erroneous and unstable biases for stations in the polar cap region [Themens et al. 2013]. The high degree of horizontal inhomogeneity and variability of the polar cap ionosphere forces one to use longer integrations in this region to overcome some of these issues.

In Themens et al. [2015], we demonstrated the first explicit validation of GPS receiver bias estimation methods and their assumptions using independent observations from an external instrument: the Resolute ISR. We first examined the effect of shell height on bias estimation, finding that bias sensitivity to shell height is locally linear and varies seasonally and with solar cycle, ranging from sensitivities of 40km per TECU in solar maximum summer to over 1000km per TECU in solar minimum winter. Despite this linear relationship, it was found that, under ideal circumstances, it is possible to determine shell height and bias simultaneously; however, these ideal circumstances are not experienced in real ionospheric systems. Future research may yet be able to decouple bias and shell height in real applications.

We next examined the behaviour of the LSQ bias estimation method using local time and magnetic coordinate systems. While using a $10^\circ$ elevation cutoff, LSQ biases determined using a local time coordinate system during winter periods are consistently lower than
those found using the magnetic coordinate system. Using a 30° elevation cutoff results in lower biases by the magnetic LSQ method, particularly during periods of high geomagnetic activity.

As we are interested in the high latitude region, where satellite elevations rarely exceed ~60°, the choice of elevation cutoff can have significant implications for the bias estimation. 10° cutoff biases tend to be lower than those from their 30° counterpart during winter periods for MSD, and spring for local time LSQ. All methods produce lower biases using 10° cutoffs during solar minimum but transition to producing larger biases as solar and geomagnetic activity increases [Themens et al., 2015].

To evaluate the performance of the MSD and LSQ bias estimation methods, and their permutations, we undertook a direct comparison to ISR observations. At mid-latitudes this would not be possible, as plasmaspheric content cannot be observed by the ISR system. In the Polar Cap, however, plasmaspheric content is largely negligible. Using this comparison, we find that 10° cutoff MSD biases outperform the other bias methods/permutations tested, with overall RMS errors of 2.68 TECU. Using MSD and local time LSQ, 10° cutoffs outperform 30° cutoffs by ~0.4 TECU. Over the period tested, local time LSQ is found to outperform magnetic local time LSQ.

Also interesting, by comparing GPS- and ISR-derived sTEC, we find that GPS biases do in fact demonstrate real seasonal, but not solar cycle, variability. These ISR-derived biases are found to correlate well with outdoor temperature at the site tested, likely due to
temperature-dependent dispersion in the cabling and antenna hardware. This study was the first such study to determine the existence of these true bias variations in absolute TEC, corroborating previous studies that inferred such variations from relative TEC from collocated GPS receivers [Coster et al., 2013, Brunini and Azpilicueta, 2010].

We also found that the erroneous solar cycle variability in estimated biases cannot be explained solely by shell height variability, but rather, likely results from large scale ionospheric gradients correlated with the vertical projection function [Themens et al., 2015]. To account for these errors we found that simply correlating estimated biases to 28-day smoothed sunspot number and removing the trend results in a significant improvement in method performance. Assuming then that true bias variability is only driven by temperature changes, we fit the sunspot de-trended biases to temperature to again improve the performance and stability of bias estimation, achieving overall RMS errors of 1.66 TECU [Themens et al., 2015]. While these measures constitute a significant improvement in bias estimation performance, they require long (several year) periods over which to undertake the sunspot and temperature correlations, making these adjustments impossible to achieve in short-term deployment situations.

5.5.3 Satellite-based TEC and Radio Occultation

Satellite-based GNSS observations provide a unique method of determining and constraining the vertical structure of the ionosphere, which cannot be achieved through
ground-based methods. One of the more popular applications of satellite-based GNSS observation of the ionosphere has been in its use for Radio Occultation (RO) remote sensing of the vertical electron density profile.

Satellite-based line of sight GNSS observations of the ionosphere follow largely the same principles as those used for ground-based observations but are subject to some additional challenges, particularly in the solution of the receiver bias problem. A full assessment of satellite-based GPS TEC estimation errors can be found in Watson et al. [2018]. For Low Earth Orbit (LEO) satellites, once one has calibrated line of sight TEC, they may then apply the RO method.

In the RO method, one records the sTEC between the GNSS satellite and a LEO receiver as the LEO satellite occults behind the earth. If one removes the contribution from above the LEO satellite orbit altitude, the remaining sTEC may be written as

\[
sTEC(a) = 2 \int_a^{R_{LEO}} \frac{N_e(r)}{\sqrt{r^2 - \alpha^2}} \, dr
\]

where \( \alpha \) is the altitude of the point where the signal ray path is tangent to a circle centered about the Earth center, the so called tangent point, and \( R_{LEO} \) is the altitude of the LEO satellite [Schreiner et al., 1999]. An illustration of the RO geometry is presented in Figure 5.6.
Figure 5.6 An illustration of the GNSS Radio Occultation geometry.

Applying an Abel Transform, one may invert Equation 5.33 for the electron density, such that

\[
N_e(r) = \int_r^{R_{LEO}} dsTEC(a) \frac{da}{\sqrt{a^2 - r^2}}
\]  

(5.34)

where \(\frac{dsTEC(a)}{da}\) is the derivative of sTEC with respect to the tangent point altitude. The above inversion is only valid under the assumption of a spherically symmetric ionosphere, which makes its application at high latitudes and near the equatorial anomaly challenging [Schreiner et al., 1999; Yue et al., 2010]. Also, bottomside electron density from this inversion can exhibit significant errors due to the technique’s reliance accurate inversion of electron density above each altitude, such that errors accumulate as one tends to lower altitudes [Kelley et al., 2009]. An improved inversion method has now been developed by Pedatella et al. [2015] that incorporates an accommodation for horizontal gradients.
5.6 Indices

The empirical electron density model presented in this study uses several solar and geomagnetic indices to represent ionospheric variations driven by solar photo-ionization and geomagnetic storms, respectively. In this section, we will discuss these various indices, as well as those commonly used by other empirical models. The indices listed in this section are available from the databases listed in Table 1.

Table 1 Sources for solar, ionospheric, and geomagnetic indices.

<table>
<thead>
<tr>
<th>Index</th>
<th>Database Link</th>
<th>Source Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>IG</td>
<td><a href="https://www.ukssdc.ac.uk/cgi-bin/wdec1/secure/geophysical_parameters.pl">https://www.ukssdc.ac.uk/cgi-bin/wdec1/secure/geophysical_parameters.pl</a></td>
<td>United Kingdom World Data Center</td>
</tr>
<tr>
<td>PC</td>
<td><a href="http://pcindex.org/">http://pcindex.org/</a></td>
<td>Technical University of Denmark</td>
</tr>
<tr>
<td>AE, Dst</td>
<td><a href="http://wdc.kugi.kyoto-u.ac.jp/dstae/index.html">http://wdc.kugi.kyoto-u.ac.jp/dstae/index.html</a></td>
<td>Kyoto World Data Center</td>
</tr>
</tbody>
</table>

5.6.1 Sunspot Number (Rz)
Solar irradiance at X-ray and EUV wavelengths follows an oscillation with average period of ~11 years (generally between 9 and 14 years). This cycle can also be thought of as being a ~22 year cycle, where the Sun’s magnetic field reverses after each ~11 year period. Closely correlated with these irradiances is the abundance of sunspots visible in the Sun’s photosphere. At the beginning of each ~11 year period, sunspots form around ~30° heliocentric latitude and form closer to the heliocentric equator as the cycle progresses. As this occurs, increasingly more sunspots are formed until solar maximum, after which sunspot abundance begins to decrease rapidly. To represent this cycle numerically, Johann Rudolph Wolf developed the Wolf Sunspot Number in 1848, later referred to as the Zurich sunspot number (Rz). Due to reliable records of sunspot number existing as far back as 1848 (and less reliable records going back several centuries), Rz is very commonly used as a proxy index for solar irradiance on climatological time scales. More recently, however, an abnormally weak solar minimum and a drift in the sunspot number over the previous couple cycles highlighted significant shortcomings in the traditional sunspot number, leading to several re-imaginations of the sunspot number [Clette and Lefèvre, 2016]. Due to the relative uncertainty in the stability of this index into the future, we have chosen to avoid the use of this index for our model.

### 5.6.2 F10.7 Radio Flux

F10.7 radio flux is the solar radio noise measured at 10.7cm (2800MHz). It is often referred to as the Covington Index after Arthur Covington, the Canadian physicist who
first related this radio emission to the solar sunspot cycle. The strong correlation between F10.7 flux and solar EUV flux makes this measurement an excellent proxy for solar photoionizing radiation.

Reliable measurements of F10.7 flux date as far back as Covington’s original measurements near Ottawa in 1947 and continue to this day. The observatory used for estimating this flux has changed over time. In 1960, observations moved to the Algonquin Radio Observatory (ARO) outside Ottawa in an effort to reduce the impact of local radio noise contamination. A second instrument was later developed in 1964 at the Dominion Radio Astrophysical Observatory (DRAO) in British Columbia. In 1991, due to the loss of funding for ARO, DRAO became the sole measurement site for F10.7 flux.

5.6.3 IG Index

IG index is a monthly ionospheric index derived from a selection of global ionosonde measurements of foF2. The index is derived by using a linear regression of the difference between observed foF2 and that modeled using the CCIR reference maps (see Section 6.1 for details) to the difference between the IG index and the reference sunspot number. The full methodology is laid out in Minnis [1955]. Due to its derivation from observations of the ionosphere, this index is generally found to be a better representation of the solar cycle variability of the ionosphere than solar proxy indices, such as F10.7 flux and Rz [Liu et al., 1983]. IG index is available starting in 1943.
5.6.4 Planetary K Index (Kp) and Planetary a Index (ap)

Kp is a geomagnetic disturbance index derived by averaging the K index recorded at several globally distributed geomagnetic observatories. The K index, first proposed by Bartels et al. [1939], is derived from the amplitude range of the most disturbed of two magnetic components (horizontal and declination components) in a three-hour interval after the removal of their smooth daily variation [Perrone and De Franceschi, 1998]. Both the K and Kp indices range from values of 0 to 9 on a quasi-logarithmic scale. ap is a linearized version of the Kp index that ranges from 0 to 400, with each unit approximately equal to a 2 nT change in magnetic flux density. Both Kp and ap are available from 1932 onward.

5.6.5 Auroral Electrojet Index (AE)

The AE index, originally proposed by Davis and Sugiura [1966], is intended as a measure of the auroral electrojet and is derived from the AU and AL indices, where $AE = AU – AL$. The AU and AL indices are calculated from the departure of the horizontal (H) component of the magnetic field from quiet time values at several stations located within the auroral zone. For each UT hour, the largest value of the H component among all stations is taken as the AU index (eastward electrojet) and the lowest value is taken as the AL index (westward electrojet) [Perrone and De Franceschi, 1998]. In this way, the AE
index can be conceptualized as the envelop width of all H-component observations from auroral magnetometers. The AE index is available from 1957 to 1975 and from 1978 to the present date. An example of the AU and AL indices, and thus the AE index, for May 2010 is presented in Figure 5.7.

![Graph showing Dst and AE indices for May 2010](image)

**Figure 5.7** Hourly AU (blue), AL (red), and Dst (black) indices for May 2010.

### 5.6.6 Dst Index

The Dst index is very similar to the AE index but is instead a measure of the equatorial electrojet. Like the AE indices, Dst is calculated from the horizontal component of
magnetic field observations. Unlike the AE indices, Dst is calculated from magnetic field observations near the geomagnetic equator. After the removal of the instrument baseline and solar quiet variation at each site, a Dst index is calculated from the remaining disturbance magnetic perturbation on an hourly basis. Dst index is available from 1957 onward. An example of the Dst variation for the same period as that presented for AE index is presented in Figure 5.7.

5.7 Altitude Adjusted Corrected Geomagnetic (AACGM) Coordinates

In this study, extensive use is made of Altitude Adjusted Correction Geomagnetic (AACGM) coordinates, which are the coordinate system chosen as a basis for E-CHAIM. AACGM coordinates were first proposed by Baker and Wang [1989] as a way to better compare coherent backscatter radar observations from conjugate hemisphere locations. In this coordinate system, a ray is traced from a user-supplied geographic longitude, latitude, and altitude along an International Reference Geomagnetic Field (IGRF) field line to the centered magnetic dipole (CD) equator. This point is then mapped back along the magnetic dipole field to the Earth’s surface, where the geographic dipole latitude and longitude of this point are taken as the AACGM coordinates of the user-supplied point. In this way, all points along a magnetic field line map to the same AACGM coordinate. Like Magnetic Apex Coordinates [Richmond, 1995], the AACGM coordinate system is not orthogonal due to its use of higher order magnetic field spherical harmonics. The AACGM coordinate grid is illustrated in Figure 5.8. Also, because this method relies on
tracing IGRF field lines to the CD equator, these coordinates are undefined in regions where these field lines do not intersect the CD equator. This undefined region is illustrated as the shaded region of Figure 5.8. This, of course, is not a concern for this study, as we are only applying these coordinates to high latitude regions.

We will often also use Magnetic Local Time (MLT) calculated in AACGM coordinates in this study. Because there are multiple different definitions of MLT and each coordinate system will come to a different MLT value, differences in MLT definitions can range as high as 30% [Laundal and Richmond, 2017]. It is thus important to define how exactly we have derived MLT for this study. In our case, we have simply defined MLT as the hour difference between the AACGM longitude of the desired point and the AACGM longitude of the reference sub-solar point. This approach minimizes the time rate difference from solar local time that one incurs due to the geographically uneven spacing of the AACGM grid [Laundal and Richmond, 2017], which is often exacerbated by the traditional approach of using the geographic longitude of the CD northern dipole as the midnight meridian reference in MLT calculation [Baker and Wang, 1989]. This latter approach is that used by the International Reference Ionosphere. The specific version of the AACGM used in this study is that of Shepherd [2014].
Figure 5.8 An example of AACGM (red) and Magnetic Apex (blue) coordinate grids at the surface. Shaded areas denote locations where AACGM coordinates are undefined. Taken from Laundal and Richmond [2017].
6 Existing Empirical Ionospheric Models

Empirical modeling is perhaps the most popular form of operational ionospheric modeling. Typically, the approach of these models is to fit historical ionospheric measurements to a simplified set of basis functions in an attempt to provide a large-scale representation of the climatological state of the ionosphere. It is important here to discuss these models, as we have used the shortcomings and successes of these models to inform our choice of parameterization and to make a case for the necessity of E-CHAIM.

There exist two primary models used to represent the 4-dimensional electron density state of the ionosphere: the International Reference Ionosphere (IRI) and the NeQuick. In this chapter, we will first undertake a comprehensive validation of the IRI, based on the work of Themens et al. [2014] and [2016]. We will then undertake a similar validation of the NeQuick model, based on the work of Themens et al. [2017b]. These validations will cover everything from the F2-peak density and height to the bottomside and topside parameterizations.

6.1 The International Reference Ionosphere (IRI)

The IRI, initially developed by a collaboration lead by Dr. Karl Rawer, is the defacto standard for the representation of the ionosphere and is recognized as such by the International Standards Organization [Bilitza and Reinisch, 2008]. It is an empirical,
climatological model of the ionosphere based on a host of datasets from around the world, including the global network of ionosondes, incoherent scatter radars, the ISIS and Alouette topside sounders, and various rocket observations. It is developed and maintained by a Committee on Space Research (COSPAR) and International Union of radio Science (URSI) joint task group, which regularly updates the model’s coefficients and proposes improvements for future versions of the IRI [Bilitza and Reinisch, 2008]. The available IRI code can output various ionospheric parameters and allows for the application of a selection of topside, bottomside, and foF2 coefficient models. For this study, we have developed an Interactive Data Language (IDL) command line code in order to interface with the IRI2007 code available from the National Space Science Data Center FTP site at http://spdf.gsfc.nasa.gov/pub/models/iri/. This code retains all of the functionality of the original IRI2007 scripts, including the capability for user specification of measured hmF2 and NmF2 values.

The IRI is widely used in applications such as the evaluation of the performance of HF modems [Jodalen et al., 2001], as a reference ionosphere in Over The Horizon Radar (OTHR) system planning [Thayaparan et al., 2016; Saverino et al., 2013; Cacciamano et al., 2009], and as a base-line ionosphere in data assimilation models [Komjathy, 1998; Hernandez-Pajares et al., 2002; Bust et al., 2004; Schmidt et al., 2008; Pezzopane et al., 2011; Galkin et al., 2012]. At mid-latitudes, the IRI offers accurate modeled ionospheric parameters, such as the heights and peak electron densities of the ionosphere’s various layers, as well as Total Electron Content [Coïsson et al., 2006; Bilitza et al., 2012]. The same cannot necessarily be said for its application to high-latitude regions, like the Polar
Cap, where there is a significant lack of available data and characteristically different dynamics.

In the following sections, we will present the formulation of the IRI’s various components, as well as a discussion of the performance of these components in their application at high latitudes. The work presented in this section is based on the author’s paper, Themens et al. [2014].

6.1.1 The F2-Peak

In order to model the vertical structure of ionospheric electron density, the IRI uses the F2-region peak (hmF2, NmF2) as an anchor point to which shape functions are normalized. This practice places a particular importance on accurately modeling hmF2 and NmF2.

6.1.1.1 Formulation

The peak electron density in the IRI is provided via two options: the Consultative Committee on International Radio (CCIR) foF2 maps or the URSI foF2 maps. The CCIR foF2 maps were developed, based on data from a global network of ionosondes operated between 1954 and 1958 [CCIR, 1982]. The model is a set of two maps: one for low solar activity and one for high solar activity. Data is first fitted to an 8th order Fourier
expansion in Universal Time for each month and each solar activity period to represent
the diurnal variability within each month. For each of these months, the coefficients of
this expansion are subsequently fit to a basis of Jones and Gallet [1965] special functions,
which are expanded in modified dip latitude, geographic latitude, and geographic
longitude. These Jones and Gallet [1965] special functions, are a set of Legendre
Functions that are characteristically similar to spherical harmonics but are orthogonal in
the desired coordinate system. Ultimately, this CCIR foF2 model is composed of 988
coefficients for each month and both solar activity conditions. To determine the electron
density at other solar activity periods, a linear interpolation is used; however, in light of
the saturation effect in the relationship between solar activity and foF2, R12, the 12-
month-smoothed sunspot number, which is used for the solar activity interpolation, is
forced constant to a value of 150 when solar activity exceeds Rz = 150. A later form of
this model incorporates an adjustment to these maps using the IG12 index.

The URSI foF2 maps are built in the same manner as the CCIR maps; however, they
feature more data from ionosondes operated in 1975-76 and 1978-79 and synthetic
theoretical data generated to fill gaps in coverage over the oceans and at high latitudes.
The dataset of this model included 111 ionosondes, 47 of which were located above
45°N, and 24 of which did not operate in the above-mentioned periods but were instead
interpolated to the solar activity conditions of these periods using a polynomial fit. The
theoretical data used to fill data gaps were generated using a time-dependent ion
continuity model from Anderson [1973] with
1) Assumed reaction rates from Rush [1983].

2) An $O_2$, $N_2$, and $O$ neutral atmosphere generated by the MSIS empirical neutral atmospheric model of Hedin et al. [1977].

3) The simplified diffusion scheme of Anderson [1973].

4) Winds that were inferred by fitting the model to measured foF2 at available ionosonde locations, neglecting electro-dynamic drift.

This real and synthetic data was then fit in the same manner used for the CCIR model.

Similar to foF2, the IRI also includes maps of the M(3000)F2 propagation factor. This propagation factor is defined as the ratio of the Maximum Usable Frequency at 3000km range (MUF(3000)F2) to foF2. The IRI features only a single M(3000)F2 option: the CCIR M(3000)F2 maps, which were developed at the same time as the corresponding foF2 maps. The M(3000)F2 maps are used as an integral part of the IRI’s hmF2 parameterization.

For hmF2, the IRI uses a modified form of the Bradley-Dudeney formulation, where CCIR M(3000)F2 maps are used in cooperation with foF2 and foE in order to determine hmF2 [Bilitza et al., 1979]. The parameterization used in the IRI is given by the

$$hmF2 = \frac{1490}{M(3000)F2 + \Delta M} - 176$$

(6.1)
\[
\Delta M = \frac{F_1(R_{12}) F_2(R_{12}, \Phi)}{f_0 F_2 / f_0 - F_3(R_{12})} + F_4(R_{12})
\] (6.2)

where \(R_{12}\) is the 12-month smoothed sunspot number, \(\Phi\) is the modified dip latitude, \(F_1\), \(F_2\), \(F_3\), and \(F_4\) are empirical coefficient functions to account for solar activity [Bilitza et al., 1979].

Given the wide-ranging use of the IRI, validation studies are necessary to ensure its appropriate use and to evolve the model over time.

\noindent \textbf{6.1.1.2 Validation Dataset}

In order to validate the performance of the IRI peak parameters at high latitudes, we make use of a network of polar cap ionosondes, operated by the Canadian High Arctic Ionospheric Network (CHAIN), and an ionosonde operated by the United States military in Qaanaaq, Greenland.

CHAIN provides a unique opportunity to undertake an evaluation IRI performance during the minimum of solar cycle 23/24 and rising phase of solar cycle 24 [Jayachandran et al., 2009]. CHAIN operates 25 stations in the Canadian Arctic region that are each equipped with a dual-frequency Global Positioning System (GPS) receiver, five of which are collocated with a Canadian Advanced Digital Ionosonde (CADI). These
systems allow for the accurate estimation of TEC and bottomside electron density parameters in the Auroral Oval and Polar Cap regions [Themens et al., 2013]. Table 2 lists the geographic location of the 5 CHAIN CADI stations and also identifies the operational capacity of each station at the time of this study.

Table 2 CHAIN Station Geographic Locations and Status

<table>
<thead>
<tr>
<th>Station</th>
<th>Latitude (°N)</th>
<th>Longitude (°E)</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eureka</td>
<td>79.99</td>
<td>274.03</td>
<td>Operational</td>
</tr>
<tr>
<td>Resolute Bay</td>
<td>74.75</td>
<td>265.00</td>
<td>Operational</td>
</tr>
<tr>
<td>Pond Inlet</td>
<td>72.69</td>
<td>282.04</td>
<td>Operational</td>
</tr>
<tr>
<td>Cambridge Bay</td>
<td>69.12</td>
<td>254.97</td>
<td>Operational</td>
</tr>
<tr>
<td>Hall Beach</td>
<td>68.78</td>
<td>278.74</td>
<td>Operational</td>
</tr>
</tbody>
</table>

A map of the CHAIN ionosondes is provided in Figure 6.1. Only the Cambridge Bay, Pond Inlet, Resolute, and Eureka stations are used in this validation.
Figure 6.1 A map of the CHAIN GPS stations used in this and following studies. Red markers correspond to locations where ionosondes are collocated with GPS observations. Approximate lower boundaries of the auroral oval (MLAT 65N) and polar cap (MLAT 75N) are marked with dashed and dotted lines, respectively.

For this validation, over 120,000 virtual height ionograms have been manually scaled from CHAIN’s database of Cambridge Bay, Pond Inlet, Resolute, and Eureka CADI data. These manually scaled ionograms were inverted to determine real height electron density profiles through the use of the Polynomial Analysis (POLAN) method [Titheridge 1985; 1988].

CADI-derived M(3000)F2, used in the following analysis, is calculated by taking the ratio of the Maximum Usable Frequency at 3000 km (MUF(3000)) to the F-region peak critical frequency (foF2). The MUF(3000) is directly retrieved from the manually scaled
ionograms using the standard transmission curve technique of Smith [1939] with a secant correction factor of 1.116 [Wieder, 1955].

Ionograms are available in either one- or five-minute temporal resolution and 6 km altitude resolution from the CHAIN network, depending on the station and time of study. All data after the summer of 2009 is at one-minute temporal resolution. To reduce the amount of manual scaling required for this study, ionograms were only scaled in 10 minute resolution.

In order to define the IRI’s performance in modelling polar cap NmF₂ prior to the extended solar minimum of cycle 23/24, we also make use of a Digisonde in operation at Qaanaaq/Thule, Greenland (77.5°N, 290.8°E). Constant-operation data from this station have been gathered from the Global Ionospheric Radio Observatory (GIRO) Digital Ionogram Database (DIDBase) for the period between 2004 and 2011 [Reinisch et al., 2004]. These Qaanaaq ionograms have been autoscaled and inverted using the Automatic Real-Time Ionogram Scaler with True height (ARTIST) autoscaling program [Reinisch et al., 2005]. Due to a series of additional complications inherent in the autoscaling of high latitude ionograms, namely spread-F, Z-mode propagation, and mode-splitting due to Travelling Ionospheric Disturbances (TIDs) (see Section 5.2), we limit the use of this data to the NmF₂ portion of this study.
6.1.1.3 NmF2 Validation

NmF2, or foF2, is an extremely important parameter in the IRI model, as the majority of the IRI electron density profile is scaled to the density at the F-region peak. In addition to this, foF2 is a primary parameter in the estimation of the IRI’s NeQuick topside thickness and is also used in IRI hmF2 estimation. Proper modelling of NmF2 within the IRI model is thus integral to the model’s capability to be used as a baseline model in HF communications or positioning forecasting [Komjathy, 1998; Hernandez-Pajares et al., 2002]. In order to evaluate the IRI’s performance within the polar cap, we begin by comparing monthly median CADI-measured and IRI-modelled NmF2. An example of this comparison at the CHAIN station in Resolute is presented in Figure 6.2, where we have plotted contour plots of CADI and IRI NmF2 values, using both the URSI and CCIR coefficient options, for the period of 2008-2011.
Qualitatively from this figure, the URSI option fails to demonstrate equinox enhancements in daytime NmF$_2$, apparent in the ionosonde data, during the period studied. Examining the CCIR option, a seasonal phase shift is observed, particularly during the increasing phase of solar cycle 24, where the seasonal maximum in daytime CCIR NmF$_2$ appears to be delayed by over a month. On the whole, both models appear to consistently underestimate nighttime NmF$_2$ for all but summer periods. Over diurnal cycles, both the CCIR and URSI models appear to terminate the daytime enhancement in NmF$_2$ too soon in the day. This lack of a persistent daytime NmF$_2$ enhancement is likely the result of transport processes playing a much more significant role in F-region dynamics at high-latitudes, as compared to the low- and mid-latitudes where the majority of the model fitting was undertaken.
Figure 6.3 Percent differences between CADI and IRI NmF2 using both CCIR and URSI coefficient maps at Resolute.

In Figure 6.3 we present the percent differences between CADI and IRI NmF2 for the same period and station. URSI NmF2 demonstrates good agreement in summer months, particularly during the daytime, where agreement is generally within ~10%. During the summer nighttime, errors are found to remain within 25%. Performance during equinox periods is, however, not encouraging, as errors during equinox nighttime are found to exceed 65%, at times, and daytime errors never fall below 25%. This pattern of increased error during periods of little solar-driven production likely arises due to transport process, which could not be observed in the primarily midlatitude datasets used to generate the model.
Looking at the CCIR option, we again observe a pattern of improved agreement during periods of solar-production dominated dynamics. NmF$_2$ during summer daytime is found to be underestimated by no more than 20%, increasing to roughly 30% during summer nighttime periods. During the equinoxes, trends are found to be similar to those of the URSI option, where NmF$_2$ is underestimated by up to 60% during nighttime periods and 30% during daytime periods. In contrast to URSI observations, the CCIR option significantly over estimates the magnitude of winter diurnal variability, overestimating winter daytime NmF$_2$ by up to 65% while underestimating nighttime values by up to 35%.

In order to compare the performance of both IRI NmF$_2$ options, monthly RMS errors for both the URSI and CCIR options, for all CHAIN CADI stations, are presented in Figure 6.4.

![Monthly RMS Errors](image_url)

**Figure 6.4** Monthly RMS errors between CADI and IRI NmF2 using CCIR (dashed) and URSI (solid) coefficient maps at Resolute (black), Eureka (red), Pond Inlet (blue), and Cambridge Bay (green).
From this figure, it is clear that the CCIR model performs best during early winter and during summer periods. The URSI model performs best during the same periods but performs particularly well during summer periods, outperforming the CCIR option. In all other periods, the CCIR option outperforms the URSI option, doing particularly well in winter periods. Both options demonstrate notable error in equinox months, as neither of the options demonstrate the spring daytime enhancement in \( NmF_2 \) that is obvious in the CADI data. Also of note is an almost linear increase in equinox \( NmF_2 \) RMS errors with increasing solar activity over the period studied, particularly while using the URSI option. This could be the result of there being a characteristically different relationship between solar activity and peak electron density at high latitudes, as compared to low- and mid-latitudes. This will be investigated in later work.

We may also observe the largely similar error patterns at all stations, demonstrating statistically insignificant differences between each. In Figure 6.5 we present contour plots of the \( NmF_2 \) from the remaining three CHAIN stations. As can be seen, all four CHAIN stations demonstrate consistent solar cycle, seasonal, and diurnal behavior in \( NmF_2 \), where \( NmF_2 \) decreases with increasing latitude during the photoionization dominated summer daytime and \( NmF_2 \) increases with increasing latitude during the transport dominated winter nighttime.
Figure 6.5 CADI-derived NmF2 from the Cambridge Bay, Pond Inlet, and Eureka CHAIN stations.

In order to characterize the effect of the extended solar minimum on IRI performance within the polar cap, we present Qaanaaq/Thule Digisonde and IRI NmF2 data for the period between 2004 and 2011 in Figure 6.6, as well as percent differences in Figure 6.7.
Figure 6.6 Digisonde and IRI modelled NmF2 using URSI and CCIR coefficient maps at the Qaanaaq GIRO station between 2004 and 2011.
From these figures, we may note that the percent differences between IRI and Digisonde NmF$_2$ increase significantly during the solar minimum period. In particular, between 2004 and 2008 the magnitude of percent differences between URSI and Digisonde NmF$_2$ during the winter and equinox nighttime increase from 35% to more than 50%, while errors during summer months remain roughly constant. Errors between Digisonde and CCIR NmF$_2$, however, remain roughly consistent, steadily demonstrating fair (within
20%) agreement during early winter and fall equinox periods and underestimation of 40-50% during all other seasons. These results are consistent with the errors observed at the CHAIN stations.

![Monthly RMS Errors](image)

**Figure 6.8** Monthly RMS errors between Digisonde and IRI NmF$_2$ using URSI (solid) and CCIR (dashed) coefficient maps at Qaanaaq between 2004 and 2011.

Absolute RMS differences between IRI and Digisonde NmF$_2$ are presented in Figure 6.8. While, in terms of percent differences, the IRI appears to perform much better during periods of high solar activity, RMS errors between IRI and Digisonde NmF$_2$ in fact decrease significantly during the extended solar minimum. This is particularly observed during summer periods, where NmF$_2$ performance increases significantly over the course of the extended solar minimum, particularly while using the URSI option.

Both the CADI and Digisonde results demonstrate comparable error patterns in the IRI’s NmF$_2$ products.
6.1.1.4  **hmF2 Validation**

The modelling of hmF₂ can have significant implications on the accuracy of OTHR system design and thus must be evaluated prior to the IRI’s use as a potential model for OTHR applications. In this study we use hmF₂ inverted from CADI virtual height ionograms to determine the accuracy of IRI-modelled hmF₂ within the high latitude polar cap region. Equations 6.1 and 6.2 imply that the choice of foF₂ map (CCIR or URSI) within the IRI model could have a significant impact on IRI hmF₂ values, and thus, errors in foF₂ could further propagate as errors in hmF₂.

![Graph showing CADI-derived and IRI-modelled hmF2 using URSI, CCIR, and CADI foF2 at Resolute between 2008 and 2011.](image)

**Figure 6.9** CADI-derived and IRI-modelled hmF₂ using URSI, CCIR, and CADI foF₂ at Resolute between 2008 and 2011.

In Figure 6.9 we present the hmF₂ measured by CADI and that modelled by the IRI using the CCIR and URSI foF₂ maps, as well as IRI-modelled hmF₂ using CADI foF₂ values.
ingested into the model. It is easy to see from this figure that the choice of foF2 can have a significant impact on the IRI-modelled hmF2, at times resulting in differences between hmF2, using CCIR or CADI foF2 exceeding 60 km.

In Figure 6.10 we present the percent differences between CADI and IRI hmF2 at Resolute for each of the options of Figure 10.

**Figure 6.10** Percent differences between CADI and IRI hmF2 at Resolute for each of the options of Figure 10.

In Figure 6.10 we present the percent differences between CADI-derived and IRI-modelled hmF2 using all three foF2 options. IRI hmF2 derived using the CCIR foF2 option underestimates equinox daytime hmF2 by upwards of 25% during the deepest phase of the extended solar minimum. During summer periods, hmF2 is slightly underestimated by roughly 5% to 10%. During nighttime periods, with the exception of the summer, hmF2 is slightly overestimated by between 3% and 9%. Using the URSI foF2 map, we find that those errors observed in the spring and summer daytime during the solar minimum are less pronounced, decreasing to within 15%. Overall, the URSI option demonstrates
appreciable improvement over the use of the CCIR foF\textsubscript{2} option. All three options overestimate winter and equinox nighttime hmF\textsubscript{2} by approximately 10\%, producing a less obvious semi-annual variation. These observations are consistent with the observations of Oyeyemi et al. [2010] and Magdaleno et al. [2011] at their Sondrestrom (66.98°N, 309.06°E) and College (69.9°N, 212.2°E) stations.

We may also note that the ingestion of CADI foF\textsubscript{2} into the IRI produces little improvement in hmF\textsubscript{2} results and even increases errors at times. What little improvement that is observed is largely constrained to spring equinox periods. This result can be easily explained by examining the effect of foF\textsubscript{2} in Equations 6.1 and 6.2. From this relationship, it is easy to show that an increase in foF\textsubscript{2} will result in a corresponding increase in modelled hmF\textsubscript{2}. During the spring equinox, the CCIR and URSI options tend to overestimate the diurnal variability of foF\textsubscript{2}. Since the IRI tends to overestimate hmF\textsubscript{2} variability during this period as well, the decreased diurnal variability of CADI foF\textsubscript{2} values corrects some of this overestimation. Agreement during these periods between CADI hmF\textsubscript{2} and that modeled by the IRI with CADI foF\textsubscript{2} ingested into the model, suggests that M(3000)F\textsubscript{2} maps likely perform best during these periods. In most nighttime periods, however, the CCIR and URSI options tend to underestimate foF\textsubscript{2} but overestimate hmF\textsubscript{2}; thus, ingesting the higher foF\textsubscript{2} from the CADI data has the effect of further overestimating hmF\textsubscript{2}. During the daytime, both options tend to underestimate the persistence and intensity of daytime foF\textsubscript{2} enhancements; thus, significant differences arise during these periods.
These results can also be examined through the use of RMS errors, which are presented in Figure 6.11. One may observe that the IRI performs best during the summer while performing its worst at the equinoxes for all both IRI foF$_2$ options. These errors at the equinoxes are a direct result of the IRI’s overestimation of the magnitude of diurnal variability during these periods. Looking at the CADI-ingested RMS results, errors are observed to be at a minimum during the spring and at their worst during the summer. Errors using all three foF$_2$ options appear to decrease as solar activity increases in 2010.

Assuming that Equations 6.1 and 6.2 are correct, these results, particularly those featuring CADI foF$_2$ ingestion, imply that there are significant errors within the IRI M(3000)F$_2$ map, which are most significant during the extended solar minimum period. Namely, based on the overestimation of hmF$_2$ observed in the CADI-ingested IRI results, it is likely that the IRI significantly underestimates MUF(3000)F$_2$ during the summer daytime.
and the nighttime of the remaining seasons, where errors are largest during the extended solar minimum.

### 6.1.1.5 $M(3000)F_2$ Validation

To verify the hypothesis identified in Section 6.1.1.4, we have undertaken an evaluation of IRI $M(3000)F_2$ using CADI-derived values. Monthly median values of $M(3000)F_2$ from both CADI and the IRI at Resolute between 2008 and 2011 have been plotted in Figure 6.12.
Figure 6.12 CADI-derived and IRI-modelled M(3000)F2 at Resolute between 2008 and 2011.

The problem of underestimated hmF2, identified in the previous Section of this study, is clearly evident in this figure. As one can see there is a striking error in the IRI’s M(3000)F2 maps, particularly in summer periods, where the model demonstrates the opposite diurnal behavior of the CADI observations. Percent differences for this data is presented in Figure 6.13, where we can now quantify a substantial underestimation in summer and spring daytime M(3000)F2.
Figure 6.13 Percent differences between CADI and IRI M(3000)F2 at Resolute between 2008 and 2011.

From this figure we observe underestimation of up to 22% during summer and spring daytime periods. Outside of this period M(3000)F2 values generally agree to within 5%. Agreement during winter and fall nighttime periods is in conflict with the hypothesis proposed in Section 6.1.1.4, as we would expect underestimation similar to that of the summer daytime during these periods based on the hmF2 results of Section 6.1.1.4; thus, this suggests a potential issue in the foE parameterization during these periods. In terms of solar activity, the errors observed through this comparison appear to increase with the deepening of the extended solar minimum, consistent with the hypothesis of Section.
6.1.1.4. This is characterized by a strong increase in observed summer daytime maximum M(3000)F₂ during the extended solar minimum while the IRI demonstrates far more marginal solar activity-based variability.

These results are largely consistent with the hmF₂ results observed in Section 6.1.1.4 (with the exception of the winter nighttime) and have significant implications for the HF communications forecasting community. As the IRI uses the CCIR M(3000)F₂ maps to directly model this parameter, these results reflect a significant mis-modeling of high latitude M(3000)F₂ by the CCIR model. This CCIR model is used, with some adjustments, in a variety of important HF communications forecasting models. The implications of these errors with respect to these HF forecasting models will be investigated in a subsequent study.

6.1.2 F₂-Region Bottomside

The bottomside of the F-region in the IRI is modeled with the following function established by Ramakrishnan and Rawer [1972]

\[
N(h) = NmF₂ \frac{\exp(-x^{B₁})}{\cosh(x)}
\]  \hspace{1cm} (6.3)

where \( x = (hmF₂ - h)/B₀ \), \( B₀ \) is the thickness between hmF₂ and the altitude where \( N = 0.24NmF₂ \), and \( B₁ \) controls the curvature of the profile shape between hmF₂ and the
anchor point at $N = 0.24NmF2$. For versions of the IRI prior to 2012, two options for B0 were available: the Table Option, and the model of Gulyaeva [1987]. The Table Option, as its name would suggest, is simply a reference table of B0 values derived from ionosonde profiles under various solar cycle, local time, and season conditions. B0 is determined from this model through interpolation. The Gulyaeva [1987] model uses a relationship between hmF2 and $h_{0.5}$, the height where the electron density has decreased to half the value of NmF2, to determine B0. We invite the reader to consult Bilitza [1990] for a full description of these options. For B1, the two models use an assumed constant value of 3.0, except under exceptional circumstances outlined in Bilitza [1990].

In IRI2012, a new model for B0 and B1 by Altadill et al. [2009] was introduced. This new B0 and B1 model represents these parameters using spherical harmonics in a magnetic dip and magnetic local time coordinate system for its spatial and diurnal representation and a Fourier expansion in Day of Year (DoY) to represent seasonal variations.

In the following section, we undertake a validation of the IRI’s two B0 options using CADI observations. The experimental B0 thickness parameter that is used in the following analysis, was retrieved using a least squares fit of Equation 6.3 to the CADI bottomside true height electron density profiles down to 0.24 NmF2, or to NmF1 if an F1-layer is present.
6.1.2.1 Validation

The IRI’s B0 bottomside thickness parameter, although not as important with respect to positioning applications as the topside thickness parameter, has been suggested to be included in the IRI’s topside scale factor algorithm [Coïsson et al., 2009], which would make accurate B0 estimation crucial to correctly modelling the topside ionosphere within the IRI. In addition to this, B0 is incredibly important in the IRI’s application in HF communications and OTHR. In Figure 6.14 we present B0 estimated using CADI measurements through Equation 6.3, as well as that modelled by the IRI Table and Gulyaeva options over the CHAIN Resolute station.

![Figure 6.14](image)

**Figure 6.14** CADI-derived and IRI-modelled B0 at Resolute using both the Gulyaeva and Table options.
As can be clearly seen, there are striking differences in the dynamics of the B0 parameter demonstrated by each model option. In this figure we observe a clear diurnal structure in B0, which is largely not modelled by the IRI Table option. During summer and equinox periods B0 is largest during the daytime and lowest during the nighttime, where the magnitude of diurnal variability is greatest during the equinoxes. During the winter, the phase of the B0 diurnal variability changes significantly, where B0 is found to be largest during the nighttime and lowest during the daytime. This feature is weakly present in both IRI options, where the magnitude of this variability is well represented by the Table option but either significantly overestimated or not represented at all by the Gulyaeva option. Summer and equinox daytime B0 demonstrates a strong coupling with solar activity, where values are lowest during the extended solar minimum and increase significantly during the rising phase of solar cycle 24. Also observed are strong equinox daytime enhancements, which are not represented in either of the IRI options.
In Figure 6.15 we present the percent differences between CADI-derived and IRI modelled B0 at the Resolute station between 2008 and 2011. Because the Gulyaeva option poorly models the observed winter reversal in the diurnal variations of B0, we find that this option underestimates nighttime winter B0 more than 30% during the solar minimum period. Both IRI B0 options overestimate equinox nighttime B0 by 20% to 35%. The absence of diurnal variability in the Table option leads to underestimation of B0 by up to 30% during the summer and spring daytime periods and overestimation by between 20% and 30% during winter and fall nighttime periods. In general, both options appear to have difficulty modelling B0 during 2009, where B0 is overestimated by up to 50% and 30%, using the Gulyaeva and Table options, respectively. This is, perhaps, due to the abnormally low solar activity during 2009. Also, the absence of a strong coupling
between equinox daytime B0 and solar activity in the IRI model, leads to underestimation by up to 25% and 30% by the Gulyaeva and Table products, respectively.

The recent publication of IRI2012 cites an improved B0 representation following the harmonic function methodology of Altadill et al. [2009] [Bilitza et al., 2010]. Although such an approach would likely resolve the primary issues observed in this study, namely the over-simplification of B0 temporal variations in the current IRI Table option, the model does not include a significant database from high latitude regions; thereby, it remains to be seen whether such IRI improvements can correctly model ionospheric parameters in these regions without the expansion of the baseline dataset to include more high latitude observations. This concern is highlighted in the Gulyaeva results presented above, where diurnal and seasonal structures are present but do not represent the dynamics of the high latitude region, likely due to the largely mid-latitude database used in its creation.

### 6.1.3 Topside

The default topside specification within the IRI is the NeQuick topside function, which was added to the IRI in its 2007 version [Bilitza and Reinisch, 2008]. While other topside model options exist, we here only discuss the NeQuick topside, as it is the default and most popular option at this time.
6.1.3.1 Formulation

The IRI and NeQuick topside is represented by a semi-Epstein layer function with height-varying scale thickness. The model function is given as

\[ N(h) = \text{sech}^2 z = \frac{4 \cdot NmF2}{(1 + \exp(z))^2} \exp(z) \]  
\[ z = \frac{h - hmF2}{H} \]  
\[ H = H_o \left[ 1 + \frac{rg(h - hmF2)}{rH_o + g(h - hmF2)} \right] \]

where \( r = 100, \ g = 0.125, \) and the topside scale thickness at the F2-peak (\( H_o \)) is given by

\[ H_o = k \cdot B2_{bot} \]
\[ k = 3.22 - 0.0538 f_o F2 - 0.00664 hmF2 + 0.113 \frac{hmF2}{B2_{bot}} \]

where \( B2_{Bot} \) is the NeQuick bottoms ide scale thickness at the F2 peak, and R12 is the 12-month smoothed sunspot number. \( B2_{Bot} \) for this model is given as

\[ B2_{bot} = \frac{0.385 NmF2}{\left( \frac{dN}{dh} \right)_{\text{max}}} \]
\[
\ln \left( \frac{dN}{dh} \right)_{\text{max}} = -3.467 + 0.857 \ln(foF2)^2 + 2 \ln(M(3000)F2) \quad (6.10)
\]

where \(foF2\) is the peak critical frequency of the F-layer, \(M(3000)F2\) is the propagation factor, and \(B2\)Bot is the bottomside thickness parameter. The latter of the above parameterizations is empirically derived based on the work of Mosert de Gonzales and Radicella [1990]. The first relationship is analytically derived assuming that the semi-Epstein function can properly represent the shape of the F-region bottomside.

### 6.1.3.2 Validation Dataset

For a preliminary validation of the IRI topside, we made use of full electron density profiles from the northern face of the Resolute ISR (RISR-N). The northward-looking face of the Resolute Incoherent Scatter Radar (RISR-N) is a deployment of the AMISR class of ISRs located in Resolute, Canada (74.73°N, -94.91°E). See Bahcivan et al. [2010] for system details.

RISR-N measurements are calibrated using two techniques. During summer daytime periods plasma line measurements, which provide a sensitive measure of Langmuir waves and hence electron densities, are used as an absolute measure of electron density. In other periods manually scaled \(NmF2\) measurements from an ionosonde system at Resolute are used, when available, to provide a robust, accurate density calibration.

Given that the RISR-N system calibration may change with viewing direction, a further
calibration step is required wherein long-term (at least several day) averages are used to normalize the calibration from different viewing directions. It is anticipated that with a proper calibration dataset, densities should be accurate to better than 10%; nonetheless, only peak-relative electron densities are used in this study.

An example of a calibration comparison between RISR-N and the Resolute ionosonde system for a four-day dataset at the end of September 2011 is shown in Figure 6.16.

![Figure 6.16 Example of ionosonde – RISR-N NmF2 (top) and hmF2 (bottom) comparison after calibration. RISR-N mean (black) and median (blue) curves for elevation angles greater than 60° are compared to CADI (red) NmF2 and hmF2 measurements. Right panels show histograms of the ratio of RISR-N to CADI NmF2](image-url)
(top) and hmF2 deviation (bottom). Mean and median ratios for the electron density are
~0.99 +/- 0.02 (standard deviation of ~20%).

The mean ratio of the ISR-ionosonde density is 0.99 +/- 0.02, and the densities track each
other well on both short and long time scales. The peak height of the F region, measured
by the two instruments, generally agrees within a standard deviation of ~15 km. This
excellent agreement is found despite the fact that the instruments use very different
techniques and are not probing a common volume.

6.1.3.3 Validation

As the default IRI topside option is, in fact, the NeQuick topside model, we will here
provide a brief validation of the model in the context of the IRI, which is later used as a
foundation for further elaboration in Section 6.2.2, where a full validation of the model is
undertaken in the Polar Cap and Auroral Oval and problem areas are diagnosed.

The topside ionosphere can sometimes account for more than 75% of ionospheric TEC,
making the accurate modelling of this region crucial to using the IRI in positioning
applications [Belehaki and Tsagouri, 2002]. In this study, we make use of ISR electron
density profiles at Resolute in order to evaluate the IRI’s performance in modelling
electron density in the lower topside (below 650 km).
In Figure 6.17 we present the seasonal means of F2 peak-normalized topside electron density contours retrieved from the Resolute ISR for the period between September 2009 and August 2010. F2 peak-normalized refers to normalizing the profiles to the F2 peak density and converting them to peak relative altitude prior to evaluating the seasonal means. Examining this figure, one will observe strong diurnal variability in topside electron density throughout the year studied, with the exception of the winter period. Diurnal variability is greatest during the spring period.
Figure 6.18 IRI-modelled, monthly median, peak normalized, topside electron density (as percentage of NmF2) at Resolute between September 2009 and August 2010.

Figure 6.18 presents the seasonal mean trends produced using the IRI with the URSI foF₂ map option. In contrast to the ISR observations, the IRI topside demonstrates little to no diurnal or seasonal variability. The percent differences between IRI-modelled and ISR-measured topside profiles are presented in Figure 6.19.
Figure 6.19 Percent differences between IRI and ISR monthly median, peak normalized, topside electron density at Resolute between September 2009 and August 2010.

It is clear from this figure that the IRI demonstrates a seasonal agreement with ISR data, where there is good agreement in the summer and winter months, but poor agreement during the equinoxes. This is characterized by overestimation in the fall and underestimation in the spring. The increasing magnitude of these disagreements with increasing altitude suggests that these errors are due in large part to an error in IRI-modelled topside thickness.

To illustrate this seasonal error trend, we have fit these mean topside ISR and IRI profiles to the NeQuick semi-Epstein layer model (Equations 6.4 - 6.6). In this way, we retrieve a characteristic shape parameter, $H_o$, for both datasets. The results of this analysis are presented in Figure 6.20, where seasonal mean trends in $H_o$ are presented for the same period as Figure 6.17, Figure 6.18, and Figure 6.19.
Figure 6.20 Topside scale parameter (Ho) derived from ISR (solid black) and IRI topside profiles using URSI foF2 (solid red) and CADI-ingested foF2 and hmF2 (dashed red) at Resolute between September 2009 and August 2010.

This figure demonstrates that the IRI is underestimating the seasonal trend in topside shape parameter, while also significantly underestimating its diurnal variations in the equinox periods. This disagreement ranges from 6 km at nighttime and 20 km during daytime in the spring period and ranges from 5 km during the daytime to 15 km at nighttime in the fall period. There is however, a high degree of agreement between IRI-modelled and ISR-derived topside scale factor during the summer and winter months, where thicknesses are within 5 km at all times. Also, there appears to be strong agreement between ISR and IRI annual means.
The thickness parameter used in the IRI’s implementation of the NeQuick topside is dominated by the calculation of the empirical function $k$ and of $B2Bot$, the characteristic shape parameter associated with a semi-Epstein F2 bottomside. As one can see from Equations 6.7 – 6.10, there is a strong dependence on $foF_2$, $M(3000)F_2$, and $hmF_2$ in these parameterizations. One may thereby assume that errors in one or more of these parameters, as was illustrated in Sections 6.1.1.3 and 6.1.1.4, should have a profound impact on the IRI topside profile. In order to deduce where the errors in the IRI topside are coming from, we ingest CADI $foF_2$ and $hmF_2$ into the IRI and compare the resulting $H_0$ values with the corresponding ISR-derived values. It should be noted at this point that the IRI uses the CCIR $M(3000)F_2$ maps in the calculation of the $ln(dN/dh)_{max}$ term no matter the parameters ingested into the model. The results of this ingestion are presented in Figure 6.20, where we may note that there is no appreciable improvement in the IRI topside after ingestion. These results suggest that errors in IRI topside profiles above Resolute are likely the result of errors in the parameterization of the topside scale factor and likely arise due to a lack of topside sounder data in these regions, particularly the lack of data covering complete diurnal and seasonal cycles. This is obvious as the annual means from both ISR and the IRI compare exceptionally well while the IRI fails to adequately represent diurnal and seasonal patterns.

Other studies have also demonstrated issues with the IRI’s NeQuick topside at high latitudes. In the sub-auroral trough region, Xiong et al. [2013] showed that the IRI overestimates near peak and lower-topside electron density by an average of 20%. Other,
less localized studies, such as Zhu et al. [2015] and Gordiyenko and Yakovets [2017], have shown significant issues in the model at virtually all latitudes. These limitations highlight a need to develop a new topside thickness parameterization for use at mid and high latitudes.

6.1.4 TEC

While the IRI does not explicitly model TEC, it is often used and validated as a TEC model [Adebiyi et al., 2014; Kenpankho et al., 2013; Komjathy et al., 1998]. Outside of high latitude regions, this practice should generally be avoided, due to significant plasmaspheric contributions to TEC above the IRI’s 2000km altitude limit. At high latitudes, however, no such plasmaspheric content exists and the integrated electron density up to 2000km can be taken as the TEC. While TEC is the integrated electron density and thus provides limited information on the specific components of the model that may be problematic, it can still be used as a diagnostic of significant issues in the representation of horizontal structuring.

For the purpose of evaluating the representation of TEC in the IRI, we will examine the impact of the user’s choice of foF2 map (CCIR or URSI), as well as the role of solar activity and radio propagation indices (sunspot number and IG index, respectively) in driving IRI TEC variability. We use the NeQuick topside model and Table bottomside thickness options for this study. These options were selected because they are the IRI-
2007 defaults and thereby the most popular choices for topside and bottomside specification.

While we use IRI-2007 for this study, the newest version of the IRI, IRI-2012, features an improved bottomside thickness parameterization described in Altadil et al. (2009). It should be noted that this change in bottomside thickness parameterization is the sole difference between IRI-2007 and IRI-2012 that will have an impact on the IRI’s TEC representation. The impact of our choice of bottomside thickness representation, and thus of the version of the model used, is discussed in Section 6.1.4.4.

This examination of IRI TEC performance uses GPS TEC observations from CHAIN (see Figure 6.1), calibrated with the technique of Themens et al. [2015], to evaluate the performance of the IRI in representing TEC over diurnal, seasonal, and solar cycle time scales. For this purpose, we first begin with a discussion of the IRI’s representation of solar cycle variability and the implications of high temporal smoothing in the performance of the IRI in representing this solar cycle variability. The work presented in this section is based on results of the author’s paper, Themens et al. [2016].

6.1.4.1 Solar Activity Variability: IG vs. IG12

To assess the solar cycle variability of IRI TEC, we first present the monthly median diurnal variability of TEC derived from the IRI and measured by GPS at a sub-auroral
location (Edmonton), two auroral locations (Sanikiluaq and Iqualuit), and a polar cap location (Pond Inlet) in Figure 6.21. We have here used the URSI foF2 map option as a starting point for our analysis. There are several features in these figures that are worthy of thorough investigation, but for the moment, we will focus on the representation of solar cycle variability.
Figure 6.21 Contours of monthly median diurnal vTEC behaviour between 2009 and 2015 at Edmonton (a), Sanikiluaq (b), Iqaluit (c), and Pond Inlet (d) using GPS (left), IRI with IG12 (center), and IRI with monthly IG (right).
For all four locations presented in Figure 6.21, we note significant enhancements in observed TEC, centered about December, 2011, and March, 2014, that are not represented by the IRI. These TEC enhancements correspond to large, short-term increases in solar activity on the order of ~100% (2011) or ~70% (2014) that persisted for only approximately three months each time. The monthly IG index and the IG12 index, derived from global ionosonde observations, are gathered from the UK Solar System Data Center (http://www.ukssdc.ac.uk/) for the period between 2009 and 2015 and are presented in Figure 6.22.

Figure 6.22 Monthly IG index (solid) and IG12 (dashed) between 2009 and 2015.
These enhancements are clearly visible in the monthly IG index but fail to be represented in the IG12 due to the high level of smoothing; in fact, these enhancements are spread out and shifted to later in the year by the IG12.

One should also note that these two enhancements represent comparable levels of solar activity while the observed TEC response is strikingly different between the two events at high latitudes. This difference in ionospheric response to the two comparable solar activity enhancements can easily be explained by the solar inclination at these sites during these periods. The first event is during the winter, when most of these sites are experiencing polar nighttime or drastically reduced solar illumination, while the second enhancement is centered about March, when there is an appreciable sunlit period at all sites. The latitudinal dependence of this enhancement is thus most pronounced during the winter, where polar nighttime limits the solar control of plasma density at high latitudes to an indirect relationship resulting from daytime plasma being transported into the polar cap via anti-sunward convection [Carlson, 1994; MacDougall and Jayachandran, 2007].

Looking at the IRI-modeled TEC of Figure 6.21, we see that the IRI fails to represent any of these solar activity enhancements in TEC and qualitatively underestimates the long term, solar cycle variability in TEC, resulting in a systematic underestimation of TEC during high solar activity periods. This leads us to consider the drivers of solar activity within the IRI.
There are four main mechanisms through which solar activity indices are used to drive IRI electron density parameters. The first, and most significant to the overall TEC representation, is the use of IG12 to linearly interpolate ionospheric peak critical frequency ($f_{oF2}$) between low- and high-solar activity $f_{oF2}$ maps. This relatively simple representation has shown reasonable performance at mid-latitudes [Kumar et al., 2015; Zakharenkova et al., 2015] but has not been assessed in the high latitude region. The second is through a weak relationship within the Bradley-Dudeney equation [Bilitza et al., 1979], used to calculate the F2 region peak height ($h_mF2$), on 12-month-smoothed sunspot number ($R_{z12}$). The third mechanism is through the bottomside thickness parameter ($B0$) used to control the thickness of the bottomside F2 layer (Bilitza et al., 1979). Finally, $h_mF2$, $f_{oF2}$, and $R_{z12}$ are used in the parameterization of the topside thickness used by the IRI to specify the electron density above the F2 peak. Through $f_{oF2}$, this also makes topside thickness dependent on IG12.

The IRI has the capability to ingest user-defined IG indices and sunspot number. In place of the IRI’s IG12 indices, we have instead ingested monthly IG indices into the IRI and recalculated the IRI TEC. For consistency we have also ingested monthly smoothed sunspot number; however, sunspot number did not seem to have a significant impact on IRI-modeled TEC. Differences between default IRI TEC and that with monthly sunspot number ingested never exceeded 0.2 TECU. The TEC resulting from ingesting monthly IG and sunspot number is presented in the third column of Figure 6.21 for the stations described above. The use of a monthly IG index to drive the model’s solar cycle variability qualitatively recovers at least a portion of the TEC enhancements in the IRI.
representation at the sub-auroral location; however, at higher latitudes, this correction is less successful, particularly during 2014. In fact, it appears that, at all locations, the IRI treats both the 2011 and the 2014 events virtually identically, despite the differences in solar illumination between those events, which likely drives the significant differences between the events as observed in the GPS results. Whether these observations result from bottomside, peak, or topside parameter errors will be discussed in Section 6.1.4.4.

**Figure 6.23** Overall RMS errors at all stations between 2009 and 2015 using IG12 (dashed) and monthly IG index (solid) with the IRI CCIR (red) and URSI (black) options. Standard errors of the differences are provided as error bars.

Overall, the use of monthly IG, rather than the IRI’s IG12, results in an improvement in TEC representation at all sites. This is demonstrated in Figure 6.23, where RMS errors calculated over the entire period of observation at each site are presented using both the
URSI and CCIR foF2 options. The use of monthly IG over IG12 leads to a systematic improvement in IRI performance at all locations. The largest impacts are seen at sub-auroral and auroral latitudes, where solar inclinations facilitate larger solar control of ionospheric density. Within the polar cap, it appears that the CCIR option outperforms or performs similarly to the URSI option; however, in auroral and sub-auroral regions, the URSI option outperforms the CCIR option on average over the experiment time range. One should take care when interpreting Figure 6.23, as data availability was not the same at all stations; thus, station-to-station patterns may be representative of sampling differences and not latitude trends. For example, Edmonton, Talyoak, and Resolute are all missing data for a large portion of 2014. Errors are largest in 2014 at all other stations; the lack of data during that period may explain the anomalously low errors at those stations as compared to their neighbours. To avoid this sampling issue when comparing station-to-station results, one must look at the errors on a monthly basis.
Figure 6.24 Differences in monthly RMS errors between the use of IRI with IG12 and monthly IG index for the URSI option. Positive values imply an improvement through the use of the monthly IG index.

For the purpose of assessing the station-to-station impact of using Monthly IG over IG12, we have also plotted, in Figure 6.24, the difference in monthly RMS errors in IRI TEC calculated using monthly IG and IG12 for all locations. As expected, the largest impacts of changing the solar activity index are felt at the sub-auroral locations, where improvements of up to 6 TECU are found with respect to IG12 results. These improvements are largely isolated to the two periods of short-term enhancement in solar activity. Polar cap stations experience far smaller improvements of up to 3 TECU. One should also note that the use of monthly IG also leads to an increase in observed errors of up to 1.6 TECU, generally during the equinoxes. The worse performance during the equinoxes through the use of monthly IG indices can possibly be explained by the fact that the IG index is derived from largely mid- and low-latitude ionospheric observations.
and includes a seasonal variability element of higher amplitude than that observed at high latitudes. The 12-month smoothing damps the seasonal variations and thus leads to better performance during the equinox maxima in NmF2. This highlights a problem with using globally-averaged ionospheric indices to drive ionospheric variability in all regions. Increases in errors while using monthly IG are marginally less prevalent while using the CCIR coefficients, particularly in the polar cap region during solar minimum, but otherwise exhibit nearly identical behaviour to their URSI counterpart. While these increases in error exist, they are significantly outweighed by reductions in error during periods of short-term solar activity enhancement, which can be seen in the overall RMS results of Figure 6.23.

### 6.1.4.2 Seasonal Variability

To assess the performance of the IRI in representing seasonal variability in TEC, we make use of monthly median TEC measured by GPS and compare it to that modelled by the IRI. This comparison is presented in Figure 6.25, with RMS errors between IRI-modeled and GPS-measured TEC presented in Figure 6.26. We are here using the URSI option for simplicity, as results were qualitatively similar between URSI and CCIR models.
Figure 6.25 Monthly median GPS (left), IRI with IG12 (middle), and IRI with monthly IG index (right) vTEC versus latitude.
Figure 6.26 Monthly IRI with IG12 (red) and IRI with monthly IG index (blue) RMS errors for all stations (arranged by latitude from top left to bottom right). GPS RMS standard deviations (black) are also included for each station.
As discussed in Section 6.1.4.1, there is a drastic improvement in the IRI representation of TEC at sub-auroral latitudes while using monthly IG indices, particularly during the short-term enhancements in solar activity of December, 2011, and March, 2014. More interestingly, it is apparent that the IRI is significantly in error during summer and equinox periods at high solar activity at all stations. In particular, we see in Figure 6.25 that the IRI is significantly suppressing TEC in the polar cap region during the equinoxes after 2010, while observations show no such drastic decrease in TEC; in fact, GPS observations only show a slight decrease in TEC with increasing latitude, reaching a minimum in the auroral oval region before increasing again as one progresses into the polar cap.

Despite the observed errors during the summer and equinoxes, the IRI appears to perform exceptionally well in representing mean winter TEC at all stations with monthly RMS errors less than GPS RMS standard deviation even during solar maximum. This underestimation during the summer and good agreement during the winter manifests as a significant underestimation of seasonal TEC variability in the IRI that gets exceedingly worse as solar activity increases and as one goes to higher latitudes.

Comparing across all stations, we see that, for IG12, comparable errors are found at all stations during summer and equinox periods at high solar activity. During low solar activity, errors appear to increase with decreasing latitude for all seasons. During winter periods errors were generally lowest in the auroral region. Maximum errors in TEC for all stations reach between 13.5 and 14.5 TECU. For Monthly IG, we again generally see
the lowest RMS errors in the auroral region. For polar cap stations, maximum errors are reduced to ~11.5 TECU, while maximum auroral and sub-auroral errors are reduced to 8.5 to 9.5 TECU.

6.1.4.3 Diurnal Variability

The analysis above allows us to assess the presence of seasonal biases in IRI-modeled TEC; however, it lacks information about the IRI’s performance with respect to modeling diurnal TEC variability. To examine the IRI’s representation of diurnal variability, we present the slopes of fits between URSI IRI and GPS monthly median diurnal TEC variability in Figure 6.27.
Figure 6.27 Slopes calculated by fitting monthly diurnal variations between GPS and IRI vTEC for IG12 (solid) and monthly IG index (dashed).

We see here that the IG12 IRI overestimates the amplitude of diurnal variability in the sub-auroral zone during the winter, with slopes reaching up to 1.4, and underestimates it
in that zone during summer and equinox periods, with slopes as low as 0.6. For the auroral region, there is very good agreement (slopes near unity) between the IRI and GPS observations during summer and winter periods, but we again observe significant underestimation of the magnitude of diurnal variability during equinox periods, with slopes as low as 0.6. In the polar cap region, we see good agreement between the IRI and GPS diurnal variability during summer periods but also see progressively worse performance in winter and equinox periods as we go to higher latitudes. Slopes reached as low as 0.2 at the Eureka station.

For Monthly IG, we see the overestimation of diurnal variability in sub-auroral regions, observed with IG12 during the winter, change to good agreement; otherwise, the change to Monthly IG appears to improve performance during the sudden enhancements in solar activity, discussed in Section 6.1.4.1, at all stations but also leads to greater underestimation of diurnal variability during fall equinox periods at all auroral and polar cap stations at low solar activity.
Figure 6.28 Pearson correlation coefficients of monthly diurnal variations between GPS and IRI vTEC for IG12 (solid) and monthly IG index (dashed).

In Figure 6.28, we present the Pearson Correlation Coefficients for the fits used to produce Figure 6.27. As you can see, outside of the polar cap, correlations did not
decrease below 0.95. For polar cap stations, we see correlations sharply decrease during summer periods as we go to higher latitudes and solar activities. Correlations reached as low as 0.4 at Eureka during the summer of 2011. This is not necessarily surprising as diurnal variability in the polar cap during the summer is minimal; thus small errors in diurnal patterns can lead to very poor correlations. The use of Monthly IG offers no appreciable improvement over IG12 in terms of diurnal correlation.

6.1.4.4 Bottomside vs. Topside

We have presented here evidence of a significant discrepancy between IRI-modelled and GPS-measured TEC, particularly at high solar activity. The question remains: where are these errors propagating from in the IRI’s parameterization of the electron density profile? To answer this question, we examine the relative contributions of bottomside and topside TEC to the observed TEC and compare these to that modelled by the IRI. To do this, we have integrated vertical electron density profiles, retrieved from the Resolute CADI system, to produce a measure of bottomside TEC. We then take the difference between this bottomside TEC and the GPS-measured TEC to arrive at an estimate of the topside TEC. Bottomside and Topside TEC at Resolute, as measured by CADI and GPS and modelled by the IRI, is presented in Figure 6.29 and Figure 6.30, respectively.
Figure 6.29 Contours of monthly median diurnal variations of bottomside TEC calculated using CADI (left), the IRI with the URSI option (center-left), the IRI with the CCIR option (center-right) and the IRI with CADI-ingested NmF2 (right) at the Resolute CHAIN station.
First, examining Figure 6.29 we see that the IRI performs reasonably well (errors less than 1 TECU) in representing bottomside TEC, particularly using the URSI option, which only slightly underestimates bottomside TEC during equinox periods. Examining now Figure 6.30, we see a very different situation, where the GPS-CADI observations
demonstrate strong enhancements in TEC during the equinoxes that are significantly underrepresented or absent in the IRI representation. Also, we see that the IRI significantly underestimates the solar cycle and diurnal variability in topside TEC. From Figure 6.29 and Figure 6.30, we can conclude that the observed discrepancies between overall TEC measured by GPS and modelled by the IRI is almost entirely isolated to the topside portion of the electron density profile. This means that these errors would also be present in the newest version of the IRI (IRI-2012), as there have been no changes to the IRI’s topside between the IRI-2007 and IRI2012 versions.

In Figure 6.31 and Figure 6.32, we present the monthly median TEC and RMS errors at Resolute for the bottomside and topside, respectively. These figures clarify the performance of the IRI’s bottomside and topside representations.
Figure 6.31 Monthly median bottomside TEC (top) and monthly RMS errors in bottomside TEC (bottom) for the IRI URSI (blue), IRI CCIR (red), IRI with CADI-NmF2 (dashed black) and GPS (black) at the Resolute CHAIN location.

From Figure 6.31, we see that the bottomside is very well represented by the IRI, where RMS errors are generally less than 1 TECU and solar cycle variability is well tracked, as there is no statistically significant solar activity trend in bottomside RMS errors.
Figure 6.32 Monthly median topside TEC (top) and monthly RMS errors in topside TEC (bottom) for the IRI URSI (blue), IRI CCIR (red), IRI with CADI-NmF2 (dashed black) and GPS (black) at the Resolute CHAIN location.

In Figure 6.32, we see another situation entirely, where significant errors are present during the summer and equinox periods, getting worse as solar activity increases. RMS errors in topside TEC range from as low as 1 TECU during winter periods to as much as 7 TECU during the equinoxes at high solar activity.
So, what is the root cause of these discrepancies in the IRI’s topside TEC representation? There are two mechanisms through which the IRI could produce significant errors in its topside TEC: through errors in the peak electron density (NmF2), from which the entire electron density profile is scaled, or through errors in the topside thickness parameterization. To evaluate the contribution of each of these mechanisms, we have ingested CADI-derived NmF2 into the IRI model at Resolute and compared the resulting bottomside and topside TEC to our observations. The results of this ingestion are presented in the fourth column of Figure 6.29 and Figure 6.30 and as the dashed lines in Figure 6.31 and Figure 6.32. Examining first Figure 6.29, we see that the ingestion has drastically improved the bottomside TEC representation, as one would expect since we have forced the IRI to use measured NmF2 in its bottomside TEC calculation. Looking at Figure 6.31, we see that this culminates in an overall reduction in bottomside TEC error; however, errors still persist during summer periods, where bottomside thickness errors begin to contaminate the IRI product [e.g. Section 6.1.2]. Interestingly, it appears that errors in the bottomside thickness parameterization were counteracting NmF2 errors during the summer periods, as is evidenced by the increase in summer RMS errors while using ingested NmF2. It is possible that the use of IRI-2012 and its improved bottomside parameterization could provide a measure of improvement over our use of IRI-2007; however, the magnitudes of these errors are not significant with respect to observed errors in topside TEC. Observing Figure 6.30, we see that ingestion has resulted in the recovery of some of the equinox enhancements seen in the GPS data; however, these improvements are marginal and significant errors still persist. Examining Figure 6.32, we see that NmF2 errors only appear to have accounted for roughly a third of the total
observed topside TEC errors at high solar activity. During low solar activity, the monthly averages of the NmF2-ingested IRI TEC demonstrate overestimation of topside TEC during the winter and fall, as well as underestimation during the spring and summer. This suggests analogous errors in the topside thickness parameterization and is consistent with the low solar activity observations of Section 6.1.3.3. As solar activity increases, these errors transition to good agreement in the winter and significant underestimation in the summer and equinoxes. These observations demonstrate that there must be a significant misrepresentation of the topside thickness in the IRI at high latitudes and highlight the need for the improvement of this parameterization in this region.

6.2 The NeQuick

The NeQuick, whose development is led by Dr. Sandro Radicella, is perhaps the second most popular empirical model of the ionosphere. The NeQuick was designed to be a computationally inexpensive representation of ionospheric electron density with hope of being integrated as a replacement of the Klobuchar [1987] model in GNSS ionospheric correction. As of the NeQuickG version the model, it is now used as the ionospheric correction model for the European GNSS service, GALILEO, and is the defacto model used by the European Space Agency for ionospheric electron density specification [Radicella, 2009]. The following validations are based on the author’s paper, Themens et al. [2017b].
6.2.1 Formulation

The NeQuick shares many of the same features/methodologies as the IRI, where both models share the same foF2, topside, and M(3000)F2 models and use a Bradley-Dudeney parameterization for hmF2. The main point where these models diverge is in their representation of the bottomside and their choice of shape functions. Unlike the IRI, the NeQuick uses a sum of semi-Epstein (hyperbolic secant squared) layers to represent each component of the bottomside profile. The generic expression for a semi-Epstein layer is given by Equation 6.4.

In the NeQuick, the altitude of each ionospheric layer is given as 120km for hmE, (hmF2+hmE)/2 for hmF1, and the Bradley-Dudeney parameterization for hmF2. The thicknesses of these layers are given as follows

\[
BE_{bot} = 5 \tag{6.11}
\]
\[
BE_{top} = \max([0.5(hmF1 - hmE), 7.0]) \tag{6.12}
\]
\[
B1_{bot} = 0.5(hmF1 - hmE) \tag{6.13}
\]
\[
B1_{top} = 0.3(hmF2 - hmF1) \tag{6.14}
\]

where \(BE_{bot}\) and \(BE_{top}\) are respectively the thickness below and above the E-region peak and \(B1_{bot}\) and \(B1_{top}\) are respectively the thickness below and above the F1-layer.
parameterization for the scale thicknesses of the topside and F2-region bottomside in the NeQuick is described in detail in Section 6.1.3.1.

### 6.2.2 Validation

The particular components of the NeQuick that are evaluated in the current study include the F2 bottomside thickness and topside thickness. While the bottomside thickness parameterization has not been altered since the model’s inception in Di Giovanni and Radicella [1990], the topside has evolved over the years, beginning first as a constant scale height model before becoming the present-day parameterization of Coïsson et al. [2006], which features a varying scale height with asymptotic behavior at high altitudes. This newest version of the topside model was developed with a limited dataset of topside sounder data, little of which came from high latitude regions. While, like the IRI, the NeQuick model can be considered reasonably accurate at mid latitudes, its application at high latitudes remains untested. That said, it is expected that the NeQuick model suffers similar limitations to those of the IRI [Themens et al., 2014; Themens and Jayachandran, 2016] due to their common use of the same critical frequency (foF2) maps, their similar hmF2 parameterizations, their common use of the CCIR Maximum Usable Frequency (MUF) maps, and their shared use of the NeQuick topside parameterization.

Previous work, outlined in Section 6.1, has shown significant shortcomings in the use of the CCIR foF2 and MUF(3000)F2 maps at high latitudes. These works also pointed to
potential issues in the NeQuick topside parameterization at these latitudes; however, no direct examinations of the source of these issues were undertaken. The results of Bjoland et al. [2016] and those of Sections 6.1.3 and 6.1.4.4 suggest that the NeQuick topside parameterization is significantly underestimating the seasonal variability of the topside thickness, resulting in significant errors in the overall topside representation and in IRI-derived Total Electron Content (TEC).

The present study focuses on the NeQuick’s representation of the F-layer, particularly the topside thickness parameter, as it is an integral part of both the NeQuick and IRI topside parameterizations. Through this work we attempt to identify the specific problem areas resulting in the errors presented in Sections 6.1.3 and 6.1.4, as well as in Bjoland et al. [2016], and offer recommendations to the NeQuick team for future model adjustments.

6.2.2.1 Bottomside Validation

Prior to examining the topside in more detail, it is important to first consider the bottomside, which has a significant impact on the topside thickness through its role in Equations 6.7 and 6.8. To evaluate the use of the NeQuick bottomside parameterization at high latitudes, we make use of a Canadian Advanced Digital Ionosonde (CADI) operated by the Canadian High Arctic Ionospheric Network (CHAIN) at Resolute, Canada (74.75N, 265.00E) [Jayachandran et al., 2009]. From this ionosonde, we can:
B1) calculate the expected B2Bot from measured foF2 and M(3000)F2 values (using Equations 6.9 and 6.10),

B2) do the same using CCIR-modeled foF2 and M(3000)F2 values,

B3) analytically calculate the maximum derivative of the vertical electron density profile to ultimately derive B2Bot from the NeQuick parameterization function (using Equation 6.10), or

B4) use a least squares fit of the semi-Epstein layer function to the ionosonde-derived electron density to derive a measured B2Bot value.

Some examples of peak-relative bottomside electron density profiles, derived from each of the above methods, are provided in Figure 6.33.
Figure 6.33 Example peak-relative bottomside electron density profiles for a) January 2, 2009, at 11:00UTC, b) January 14, 2009, at 00:30UTC, c) June 14, 2009, at 14:00UTC, and d) June 1, 2009, at 2:30UTC. The black solid line corresponds to the measured profile, the orange squares correspond to the best fitted NeQuick profile (i.e. method B4), the green triangles correspond to the profile calculated using method B3, the red stars correspond to the profile calculated using method B1, and the blue diamonds correspond to the profile calculated using method B2. Corresponding B2Bot values for each of the methods is provided in the legend of each image in km.

Figure 6.33a demonstrates somewhat of an ideal scenario, where the semi-Epstein function clearly fits the measured profile very well and only relatively small errors are seen in the profile generated using fitted \((dN/dh)_{\text{max}}\). The small disagreement in Figure 182
6.33a highlights the strong sensitivity of the NeQuick parameterization to even slight departures from the idealized semi-Epstein shape, where seemingly innocuous differences in the electron density profile result in significant differences in the profile slope and thus result in significant errors in the densities generated through method B3. Figure 6.33b demonstrates a situation where the semi-Epstein shape almost perfectly matches the measured shape in the near-peak portion of the profile, as demonstrated by the strong agreement between methods B3 and B4. It is important to note that despite this agreement, both profiles significantly diverge from the measured profile beyond the region of fitting (i.e. below 0.6NmF2). Figure 6.33c demonstrates a worst-case scenario, where the F-region density decreases so quickly beyond the peak region that the semi-Epstein shape is incapable of properly fitting the measured bottomside profile. Finally, Figure 6.33d demonstrates a situation similar to Figure 1b but where the method B3 profile fails to match the measured profile. In all four of the above situations, methods B1 and B2 fail to even remotely correctly represent the measured bottomside profile, generally significantly underestimating the measured electron density. It is only in Figure 6.33c that method B1 outperforms the other approaches.
Figure 6.34 Monthly median B2Bot calculated by methods B4, B3, B1, and B2, respectively from left to right.

In Figure 6.34, we apply all four of the above B2Bot calculation methods and plot the resulting monthly median values for all available data at Resolute. For method (B2), we use CCIR-derived foF2 and M(3000)F2 to come to a modeled B2Bot estimate. As both the IRI and NeQuick use the CCIR foF2 and MUF(3000)F2 maps, the results of the following study should be representative of the performance of both models.

From this figure, it becomes clear that methods of B2Bot calculation using Equations 6.9 and 6.10 lead to significantly underestimated B2Bot values regardless of the source of foF2 and M(3000)F2 values.
Figure 6.35 Monthly RMS errors in B2Bot for situations B1 (dashed black line), B2 (dashed red line), and B3 (solid black line) at Resolute.

This can be seen in Figure 6.35, where RMS errors from methods (B1) and (B2) are largely comparable and much greater than those from method (B3). These results suggest that Equation 6.10 must be re-parameterized.

Interestingly, even the use of ionosonde-derived \((dN/dh)_{max}\) with Equation 6.9 exhibits some appreciable errors, particularly during the summer nighttime, winter, and low solar activity periods. These errors imply that the high latitude F-layer does not adequately follow a semi-Epstein shape to ensure the validity of Equation 6.9. This is further illustrated in Figure 6.36, where we present the fitted B2Bot plotted against the variable term on the right-hand side of Equation 6.9.
Figure 6.36 Plot of fitted B2Bot vs. the NmF2 to (dN/dh)max ratio (black points). The blue line represents the expected relationship given Equation 6.9.

As one can see, there is a significant departure from expectation, largely associated with winter and nighttime periods (see Figure 6.34). At these latitudes, we often observe a very strong vertical gradient in electron density in the lower F2 region with a relatively broad peak thickness. When fitting for B2Bot, the NeQuick profile function was at times observed to be incapable of fitting the curvature of the bottomside, where the thickness required to match the near-peak region would produce an incorrect (underestimated) electron density gradient in the lower portion of the F2 region profile. It should be noted that the range of densities used for the fitting of B2Bot was limited to the region of densities greater than the F1 peak density and less than hmF2. If no F1 ledge was reported, the lower limit was set to 0.6NmF2. Examples of some of the aforementioned situations can be seen in Figure 6.33.
Wang et al. [2010] previously suggested the need to modify the constant factor, 0.365, of Equation (4) in order to improve the modeled bottomside thickness; however, because this relationship is analytically derived, changing this relationship would only serve to create an inconsistency in the semi-Epstein layer parameterization and would only produce correct B2Bot values at the cost of poorly capturing the curvature of the bottomside and potentially exacerbating bottomside profile errors as a whole. These results suggest that either a new profile shape function is necessary or a spatially-dependent, height-varying term should be added in place of the current constant scale factor, B2Bot.

Overall, the above results suggest that Equation 6.10 should be revised to account for potential regional variations in such a relationship, or at the very least, that a new parameterization is necessary. Also, the results of Figure 6.36 suggest that an alternative profile shape function may be necessary for applications at high latitudes.

The question remains, do these issues materialize in the topside thickness as well? To address this issue, we examine the NeQuick topside parameterization in its use of modeled and measured ionospheric characteristics (dN/dh, hmF2, foF2, MUF(3000), etc…), as well as the use of the parameterization with measured bottomside thickness.
6.2.2.2 Topside Validation

To examine the performance of the NeQuick topside model as it applies to the high latitude ionosphere, we make use of Incoherent Scatter Radar (ISR) data from the Resolute (RISR-N) and Poker Flat (PFISR) ISRs. The topside thickness from these ISRs is calculated by fitting ISR electron density profiles to the NeQuick topside function, fitting for $H_0$. In Figure 6.37, we present the monthly median $H_0$ values at RISR-N calculated:

T1) from directly fitting the electron density profile using Equations 6.4, 6.5, and 6.6,
T2) by using B2Bot, fitted from the bottomside electron density profile using Equations 6.7 and 6.8,
T3) by using ISR-measured $(dN/dh)_{\text{max}}$, $hmF2$, and $NmF2$, with Equations 6.7, 6.8, and 6.9, and
T4) by using the full NeQuick parameterization with IRI values of $foF2$, $M(3000)F2$, and $hmF2$ in Equations 6.10, 6.9, 6.8, and 6.7.
In Figure 6.37, we see a strong seasonal variation in ISR-derived topside thickness with a muted diurnal variation and increasing trend with increasing solar activity. Figure 6.37 shows that the use of IRI-derived parameters with the full NeQuick parameterization results in significantly underestimated $H_0$ values during summer periods, overestimation during the winter, and thus, a damped seasonal variation, as compared to observation. The use of measured $(dN/dh)_{max}$ results in a much more realistic seasonal variation in $H_0$; however, it also results in a significant overestimation of winter topside thickness at low solar activities and an opposite solar activity trend, with $H_0$ decreasing with increasing
solar activity. This result is somewhat interesting for ionosonde users who intend to use
the NeQuick topside function to extrapolate the topside from bottomside ionograms at
high latitudes. We may also note the significant overestimation of \( H_0 \) when using fitted
B2Bot. This implies that the relationship for \( k \) already takes into account a large portion
of the \((dN/dh)_{\text{max}}\) underestimation errors and that correcting B2Bot will likely not
improve the \( H_0 \) performance.

Figure 6.38 Same as Figure 6.36 but for PFISR.

In Figure 6.38, we present the results for the PFISR system, which is located within the
auroral oval region. The larger and more continuous dataset at PFISR allows us to further
examine the performance of the NeQuick topside for ionogram-derived topside electron density extrapolation, as well as to assess the performance of the parameterization in the auroral oval region and examine the qualitative features of the topside thickness in this region.

Examining the qualitative behavior of the topside thickness in the auroral oval, we only see two major differences from the variations presented in Figure 6.37 for the polar cap region; namely, we see seasonal maxima in daytime thickness during the equinoxes and stronger diurnal variations in thickness, strongest during the equinoxes.

In terms of the NeQuick’s performance, the larger dataset at PFISR highlights some remaining issues with the use of the NeQuick topside profile function with ionosonde- or profile-derived parameters: 1) we more clearly see an opposite solar cycle variation in the situation (T3) $H_0$, which culminates in an overestimation of daytime $H_0$ at solar minimum, 2) seasonal variability in situations (T3) and (T4) $H_0$ is damped with respect to that fitted from ISR topside profiles, 3) in situation (T2), an unexpected maximum is found during winter nighttime periods, and 4) $H_0$ is significantly overestimated by the NeQuick parameterization during winter periods, regardless of the parameters used. These results are all consistent with the observations of RISR-N; however, the IRI appears to perform better within the auroral oval region.
Figure 6.39 Percent errors in monthly median $H_0$ at PFISR for methods T2, T3, and T4, respectively from left to right.

The differences in relative performance can be seen in Figure 6.39, where we present the percent differences between modeled and fitted $H_0$. From this figure we see overestimation of $H_0$ by situation (T2) during winter periods by over 150%. Situations (T3) and (T4) are each characterized by underestimation during summer periods by up to 60%, overestimation by up to 50% during winter daytime periods, and slightly better performance during the equinoxes. The main difference between situations (T3) and (T4) largely culminate in situation (T3)’s tendency to increasingly overestimate daytime $H_0$ at low solar activities by up to 100%.
The observed decreasing trend in $H_0$ with solar activity in situations (T2) and (T3) is likely due to the negative $f_0F2$ term of Equation 6.7. Since the IRI tends to systematically underestimate $f_0F2$ at high latitudes, particularly at high solar activity, we do not see this decreasing trend in the IRI $H_0$ values. Interestingly, we also see characteristically different behavior in the polar cap (RISR) compared to the auroral oval (PFISR). At RISR, the IRI underestimates $H_0$ during the summer and over estimates it during the winter. Also, we see only very minor diurnal variations in fitted topside thickness, which matches the variability of the observations. At PFISR, however, we see strong performance by the IRI during the summer with overestimation during virtually all other periods. We also see the development of a pronounced diurnal variability with increasing solar activity.

It should be noted that white areas in Figure 6.37 and Figure 6.38 correspond to data gaps. In the fitted $H_0$ plots, white areas correspond to periods where either there was no ISR data or the quality of the topside data was insufficient for fitting. This second case occurred predominantly during the winter nighttime at solar minimum, when the ISR signal-to-noise was very low in the topside, and corresponds to profiles that did not have reliable data up to 150km above $h_mF2$. In the other plots of these figures, white areas only correspond to periods where there was no ISR data available or when an $h_mF2$ and $N_mF2$ estimate could not be reliably derived from the ISR profile.
In Figure 6.40, we present the monthly RMS errors in topside thickness as derived using measured and IRI-modeled parameters. Despite the tendency for the IRI to underestimate diurnal and seasonal variability, we see in Figure 6.40 that the IRI, in fact, outperforms the use of measured bottomside parameters by a significant margin, particularly during solar minimum. This, however, by no means implies that the IRI parameters produce good agreement with observation, as it is still producing errors of -50% to 50%. Rather, these results imply that not only does the bottomside (dN/dh)\text{max} parameterization need to
be adjusted for regional variations, the NeQuick topside parameterization itself needs to be adjusted to address its aforementioned shortcomings. Despite this, given the significant issues seen in the B2Bot parameterization, the NeQuick topside parameterization performs far better than initially expected, where the k adjustment parameter corrects for a large portion of the observed B2Bot errors; in fact, it is interesting to note that RMS errors in $H_0$ are maximized during winter periods, as opposed to the summer maxima for B2Bot.

7 Introduction to the E-CHAIM Models

Since the creation of the IRI, and similarly the NeQuick [Nava et al., 2008], electron density models, a plethora of data have become available for use in empirical modeling, namely that from new ionosonde and Incoherent Scatter Radar (ISR) deployments in the arctic regions and from radio occultation (RO)-based electron density inversion. These new data sources allow for the modeling of spatial scales that were not available to previous models, and satellite data, in particular, promise to improve the representation of the ionosphere in regions of sparse ground instrument coverage, such as in the arctic regions and over the oceans.

There have been several attempts to create global alternatives to the IRI, some of which were sufficiently successful to warrant their inclusion in the IRI as direct alternatives or replacements to the original IRI model [Shubin, 2015; Altadill et al., 2013; Altadill et al.,
Coïsson et al. [2006] used topside sounder data to develop the NeQuick topside model, which is now the default topside specification in the IRI, since IRI2007. Altadill et al. [2009] used ionosonde data with a spherical harmonic expansion in local time and modified dip angle to create a new bottomside thickness representation, which was later included in IRI2012. Shubin [2015] used radio occultation, ionosonde, and topside sounder data to create a new global hmF2 model, which is slated to replace the IRI’s traditional hmF2 model in IRI2016 [Bilitza et al., 2017]. Other global models include the non-linear models of Hoque and Jakowski [2012] and [2011], which provide hmF2 and NmF2, respectively. All of these models are global representations that highlight significant progress within the empirical ionospheric modeling community over the past decade; however, their ability to model the high latitude ionosphere is still suspect, particularly due to the lack of high latitude data incorporated into these models during fitting and the tendency for global models to emphasize the representation of the much higher amplitude, equatorial and low latitude ionospheric variabilities. As such, we believe that, particularly for the high latitude region, a regional modeling approach is warranted.

Such regional modeling approaches have been attempted in the past. For example, Karpachev et al. [2016] developed a model of ionospheric critical frequency (foF2) for the Main Ionospheric Trough (MIT) region between 45°N and 75°N and between 40°S and 80° using INTERCOSMOS-19 (IK-19) topside sounder and Challenging Minisatellite Payload (CHAMP) in situ data. Their model demonstrated a consistently better representation of the MIT compared to that generated by the IRI; unfortunately,
given the MIT focus of the Karpachev et al. [2016] model, only night time (18 – 06 LT) winter (Nov - Feb) values are provided.

E-CHAIM is intended as a replacement for the use of the IRI and NeQuick models at high latitudes. To this end, the model represents ionospheric electron density in the region above 50°N geomagnetic latitude. The model is composed of several sub-models, each representing a key feature in the ionospheric electron density profile. Like the IRI, NmF2 and hmF2 are chosen as the anchor point of the profile, with all other components representing characteristics with respect to the F2 peak density and height. Each of these sub-models feature a spherical cap harmonic expansion in the new Altitude-Adjusted Corrected Geomagnetic (AACGM) coordinates of Shepherd [2014], calculated at 350km altitude, for the representation of the horizontal structure of the modelled parameter. All further references to geomagnetic coordinates in this text refer to these 350km AACGM coordinates. These spherical cap harmonics are similar to those used in Weimer [1996], Liu et al. [2014], and Yamazaki et al. [2015]. The order and degree of this expansion is determined experimentally, based on the amount, distribution, and quality of available data. The seasonal variability is modelled by a Fourier expansion and solar cycle variability is modelled via a function of solar F10.7 cm flux and IG ionospheric index.
8 Modeling the Ionospheric Peak Density: NmF2

This chapter discusses the development and validation of E-CHAIM’s peak electron density (NmF2) model. The first phase of this model is a climatological (quiet-time) representation of NmF2, which will act as a reference point upon which to build the rest of the model. The model dataset is presented in Section 8.1 while the quiet-time model parameterization, validation, and application is presented in Sections 8.2, 8.3, and 8.4, respectively. Following this, a second model was developed to represent the storm behaviour of the high latitude ionosphere. This storm model is presented in Sections 8.5 and 8.6. The author’s Themens et al. [2017a] paper forms the bulk of the work presented in this chapter.

8.1 Data

To represent the quiet-time behaviour of NmF2 at high latitudes, we have gathered ionogram data from 82 ionosondes. These ionosondes include data gathered from the Canadian High Arctic Ionospheric Network (CHAIN) available at http://chain.physics.unb.ca [Jayachandran et al., 2009], from the Global Ionospheric Radio Observatory (GIRO) available at http://giro.uml.edu/ [Reinisch and Galkin, 2011], from the now decommissioned Space Physics Interactive Data Resource (SPIDR), which was available at http://spidr.ionosonde.net/spidr, from the World Data Center for Solar-Terrestrial Physics (WDC for STP, Moscow) available at

The location of these ionosondes is presented in Figure 8.1. Stations included in the model fitting are marked with filled red circles, while stations used for validation are marked with X’s. Note that some validation station locations coincide with the location of a station used in the model fitting. This is due to there being previous stations at those locations prior to the stations used for validation (ex: Resolute). For ionograms processed with the ARTIST autoscaling software (e.g. those from GIRO), only data with a quality control index of 70 or greater were included in the fitting dataset.

![Figure 8.1](image)

**Figure 8.1** Plot of the global distribution of ionosondes used in the creation of the NmF2 phase of E-CHAIM (red dots). The dashed line corresponds to the lower boundary of the model at 50°N geomagnetic latitude. The dotted line corresponds to 65°N geomagnetic latitude. Black X’s represent stations that were not included in the model fitting and were instead used in the model validation.
Radio Occultation (RO) data is also gathered from the CHAMP, Gravity Recovery and Climate Experiment (GRACE), and Constellation Observing System for Meteorology, Ionosphere, and Climate (COSMIC) missions, for all occultations above 45°N geomagnetic latitude. This data was gathered from the CDAAC data portal at http://cdaac-www.cosmic.ucar.edu/cdaac/. NmF2 and hmF2 were processed from these RO electron density profiles in a similar manner to that used in Shubin et al. [2013] and Shubin [2015].

For the purpose of the quiet-time model, only data corresponding to periods with Kp index less than 3.5 were included in the fitting dataset for the quiet-time model. Overall, over 28 million data points are used in the fitting of the quiet-time NmF2 portion of the E-CHAIM model, spanning seven solar cycles between 1931 and 2016.

### 8.2 Quiet-time Model Parameterization

As mentioned in Section 7, the quiet-time NmF2 model is fitted by multiple linear regression (e.g. ordinary linear least squares) to a spherical cap harmonic function with Gauss coefficients tied to a Fourier expansion in day-of-year (DoY). The explicit parameterization of the model is as follows:

\[
\log(NmF2) = G + \sum_{l=0}^{L} \sum_{m=0}^{\min(l,M)} \left[ A_{lm} \cos \left( \frac{\pi m}{180} \lambda \right) + B_{lm} \sin \left( \frac{\pi m}{180} \lambda \right) \right] P_{lm}(\eta) \quad (8.1)
\]
\[ \eta = \cos \left( \left( 90 - \varphi \right) \frac{\pi}{45} \right) \] (8.2)

\[ A_{lm}, B_{lm} = (\gamma_{lm} F_1 + \delta_{lm} F_2) \cdot \sin^2 \left( \frac{\pi \cdot \text{DoY}}{365.25} \right) + (C_{lm} F_1 + D_{lm} F_2) \] (8.3)

\[ c_{lm}, d_{lm} = \sum_{c=1}^{K} \alpha_{lm}^c \cos \left( \frac{2\pi c \cdot \text{DoY}}{365.25} \right) + \beta_{lm}^c \sin \left( \frac{2\pi c \cdot \text{DoY}}{365.25} \right) \] (8.4)

\[ G = F10.7 \cdot (a_1 \cos(\chi) + a_2 \sin(\chi)) + \sqrt{F10.7} \cdot (a_3 \cos(\chi) + a_4 \sin(\chi)) + IG \cdot (a_5 \cos(\chi) + a_6 \sin(\chi)) + a_7 F10.7^2 + a_8 IG^2 \] (8.5)

\[ F_1 = F10.7_{81} \quad F_2 = (F10.7_{81})^{\frac{1}{1.9}} \] (8.6)

where \( \lambda \) is magnetic local time, \( \varphi \) is geomagnetic latitude, DoY is the day of year, \( F10.7_{81} \) is the 81-day smoothed F10.7 solar flux gathered from the NGDC portal at ftp://ftp.ngdc.noaa.gov/STP/GEOMAGNETIC_DATA/INDICES/KP_AP, IG is the monthly ionosonde-derived IG index gathered from the UKSSDC, and \( \chi \) is the solar zenith angle. \( \alpha, \beta, \gamma, \delta, \) and \( a_{1-8} \) are fitting coefficients. The maximum degree (L) and order (M) for the spherical cap harmonic expansion were optimized via trial and error to be five and four, respectively. The Fourier expansion in DoY was chosen such that up to quintenial variations are represented (K = 5). This relatively high seasonal resolution was chosen such that short term (2-3 month) equinox enhancements, observed in polar cap NmF2, could be adequately represented by the model. This expansion results in 488 coefficients. Overall, these values were chosen to minimize the number of artifacts in the NmF2 representation while providing a realistic representation of the spatial gradients present in the climatological high latitude ionosphere. This proved particularly
challenging when trying to realistically represent the spatial extent of auroral region enhancements in NmF2 while avoiding the creation of artifacts in regions of little data, such as the Arctic Ocean region.

To represent diurnal variability, the model is actually composed of 24 separate models (e.g. 24 separate sets of Equations 8.1 – 8.6) fitted to data binned in UTC. To get the NmF2 at a given point in time, linear interpolation between the models is used. This approach is similar to that used in Shubin [2015] with their full spherical harmonic model. Attempts were made to use local time or geomagnetic local time as a longitudinal coordinate for a single model, such as that used in Altadill et al. [2009]; however, because of the coupled solar and geomagnetic control of NmF2 at these latitudes and the prevalence of UTC-dependent structures at high latitudes [Sojka et al., 1991], fits to such a single model were unsatisfactory.

Despite using 24 separate models, magnetic local time was selected as the longitudinal coordinate to partially account for within-hour variability. This geomagnetic latitude-magnetic local time coordinate system was chosen due to it providing a slight improvement in performance over similar pure-geographic, geographic latitude-local time, and pure-geomagnetic coordinate systems, which largely represented the differences in how well these systems could capture the within-hour variability of the dataset.
The functions of F10.7 flux and the additional group of G terms, used in the above model parameterization, were selected purely experimentally, via trial and error, and were chosen based on maximizing the fit correlation and minimizing the fit root mean square (RMS) error.

8.3 Quiet-time NmF2 Validation

To validate the quiet-time portion of the E-CHAIM model, we have selected four validation datasets that were not included in the original fitting of the model; thus, we here test the predictive capability of the model and not just the model’s ability to fit the dataset. The stations chosen are: Eielson, Alaska, (64.66°N, 212.91°E) a station that is generally near the equatorward boundary of the Auroral Oval; Cambridge Bay, Nunavut, (69.12°N, 254.88°E) a station that is generally at the poleward boundary of the Auroral Oval; Resolute, Nunavut, (74.75°N, 265.00°E) a Polar Cap Station; and Alert, Nunavut, (82.60°N, 297.40°E) another Polar Cap station but one closer to the region of data absence in the Arctic Ocean and to the geomagnetic and geographic poles. These four stations, represented as X’s in Figure 8.1, provide reasonable coverage of the latitudinal domains expected to be represented by the model. It should be noted here that all data used for model validation in this study have been manually scaled to ensure the quality and accuracy of the validation datasets.
The validation of the quiet-time model primarily examines the model’s capability to represent monthly median NmF2 and will include comparisons to the URSI foF2 maps of the IRI model. To that end, we present the monthly median NmF2 from all four validation sites in Figure 8.2. Please note that different time periods are examined by each validation station, purely as a result of data availability restrictions; nevertheless, these stations capture the solar minimum period of cycles 23-24, the solar maximum period of cycle 24, and the solar minimum period of cycles 22-23.
Figure 8.2 Ionosonde-measured (left column), E-CHAIM-modeled (middle column), and IRI-modeled (right column) NmF2 for the Resolute (A), Alert (B), Eielson (C), and Cambridge Bay (D) validation stations.

Purely qualitatively, we see a significant improvement in the representation of NmF2 by the E-CHAIM model. At all stations, we see a significant improvement in the
representation of equinox, daytime NmF2, particularly at high solar activity. To a lesser extent, we see a significant improvement in the representation of NmF2 during the summer daytime at solar minimum. In addition to this, the use of 81-day smoothed F10.7 flux and monthly IG index allows the E-CHAIM model to partially represent the daytime NmF2 enhancements associated with short term increases in solar activity centered about December, 2011, and March, 2014, which are not captured by the IRI [Themens et al., 2016]. The magnitudes of the E-CHAIM-modeled enhancements are still, however, somewhat smaller than that observed by ionosonde, particularly at Cambridge Bay, leading the authors to suspect these enhancements are far more complex in origin than initially expected. Simulations with physics-based models will be necessary in order to properly assess the mechanism through which we observe these strong enhancements in ionospheric electron density.

The apparent improvement, seen in Figure 8.2, is reinforced when comparing the RMS errors between the models in their representation of monthly median foF2, which are presented in Figure 8.3. foF2 was chosen for this comparison instead of NmF2 simply for its greater familiarity and simpler physical interpretation.
Figure 8.3 Monthly RMS errors in E-CHAIM (solid line) and IRI (dashed line) foF2 at the Resolute (top left), Alert (top right), Cambridge Bay (bottom left), and Eielson (bottom right) validation stations.

Clearly from this figure we see a dramatic improvement in the representation of foF2 by the model at each of the polar cap stations, particularly at solar minimum and during the equinoxes, where improvements can be by as much as 1 MHz or more. At all stations, the E-CHAIM quiet-time model is either better than or matches IRI performance. These improvements diminish significantly as we progress to lower latitudes, which is somewhat expected, as more data was available for the IRI fitting at those latitudes; that said, the improvements seen are still significant, particularly for low solar activity periods.
Figure 8.4 Contours of foF2 from E-CHAIM (A) compared to that generated by the IRI (B), for representative months in 2013 at 00 (top row), 08 (middle row), and 16 (bottom row) UTC.
In terms of spatial variability, we present quiet-time foF2, generated by the E-CHAIM and IRI models, for 2013 in Figure 8.4. 2013 was chosen arbitrarily for illustration purposes. From this figure, we can easily distinguish the winter MIT and subsequent polar hole separated by an enhancement in foF2 in the auroral oval, not resolved by the IRI. During the summer we see a strong enhancement in foF2 extending from the dayside into the polar cap, which is also not represented in the IRI. Aside from structures we expect to see at high latitudes, such as the MIT and polar hole, one should note that without comparisons to datasets with large spatial coverage, truly assessing which model provides a more comprehensive representation of horizontal structures cannot be accomplished using this Figure. In lieu of these types of observations, we must interpret this Figure in conjunction with previous studies of the IRI performance in these regions and with the comparisons provided previously in this study. From Figure 8.4, one sees that the E-CHAIM generally produces higher densities in the polar cap region with respect to those of the IRI with the exception of the nightside, where densities are lower. Given the IRI’s cited tendency to underestimate foF2 in the polar cap during the equinox daytime and overestimate it during the equinox nighttime [e.g. Section 6.1.1.3], this comes as no surprise. Also of interest is the E-CHAIM nighttime foF2 at 08 UTC, which is bisected by an enhancement extending from the dayside to the lower boundary of the model. In the IRI we see this bisection at all times of the day but only during and around winter periods. It is also interesting to note here that, had we used a single model in local time or magnetic local time for the model’s longitude coordinate in the spherical cap expansion, we would not see such differences between the E-CHAIM outputs at the
various UTC times and would rather see the same structures simply rotated about the pole location.

![Graph showing correlations and slopes for linear fits of IRI (dashed) and E-CHAIM (solid) NmF2 to that from the Resolute ionosonde for each local time hour.](image)

**Figure 8.5** Correlations (top) and slopes (bottom) for linear fits of IRI (dashed) and E-CHAIM (solid) NmF2 to that from the Resolute ionosonde for each local time hour.

Given the results of Section 6.1.4, where the IRI demonstrated significant issues in representing TEC at high solar activity within the domain of this model, we feel it is important to also examine the seasonal and solar cycle variability of the E-CHAIM model. In Figure 8.5 we present the slopes and correlations of linear fits between modelled and measured NmF2 at each local time hour at Resolute. In effect, these fits represent how well the models perform in their representation of seasonal and solar cycle variability at each of these times of day. If seasonal and solar cycle variability is well matched, we expect slopes of unity and strong, near unity, correlation coefficients. At
Resolute, we see a consistent pattern of sub-unity slopes from both models, indicating a tendency to underestimate seasonal and solar cycle variability. Despite this, we see a significant improvement in both slope and correlation through the use of the E-CHAIM model over the IRI at nearly all local times. Resolute is taken as a representative location for consistency with previous studies: however, comparable features were seen at each of the other validation locations.

### 8.4 Solar Activity Trends and Seasonal Anomalies

In the process of fitting the E-CHAIM NmF2 model, we have acquired and processed an unprecedentedly robust dataset of ionosonde observations. This dataset allows us to examine climatological ionospheric behaviour that may not have been accessible previously. In particular, we shall here examine the behaviour of experimental NmF2 in three latitudinal regimes on seasonal and solar cycle time scales and assess the performance of E-CHAIM’s fit to this dataset. Ultimately, we here hope to identify features that may require additional consideration in the model formulation that could be added in later versions of the model.

In each of the following subsections, we discuss and provide plots of the seasonal and solar cycle behaviour of NmF2 recorded by ionosondes and fitted by E-CHAIM for a particular latitudinal regime.
8.4.1 Upper Mid Latitudes

To examine the solar cycle behaviour of NmF2 at upper mid latitudes, we present median NmF2 from all ionosonde and E-CHAIM observations within a 5° bin in AACGM latitude centered about 52.5°N for December (Figure 8.6a), September (Figure 8.6b), and June (Figure 8.6c) plotted against local time and solar flux. During December, one will note a prominent daytime maximum in NmF2 centered about local noon with a sharp morning increase and more gradual evening decline. This behaviour is as expected and occurs during all solar activity periods. For the equinox period in September, we note similar behaviour but with much more gradual transitions between night and day at high solar activity compared to low solar activity periods. This can be attributed to a much more dramatic nighttime increase in NmF2 with solar activity during the equinoxes, as compared to winter periods, which feature only a very weak nighttime increase with increasing solar activity. Finally, for summer periods, we see dramatically different local time behaviour in NmF2, where there is a daytime maximum centered prior to local noon and a second maximum just prior to local midnight. This nighttime maximum in NmF2 is still the subject of research; however, most studies believe the dominant mechanism for its production to be equatorward nighttime neutral winds, which drive plasma along magnetic field lines to higher altitudes and thus a region of reduced recombination rate [Chen et al., 2015] or to be the result of a sudden drop in electron temperature at sunset driving downward electron diffusion [Evans, 1965]. Zonal neutral winds act in a similar manner to raise or lower the ionosphere, which can accentuate or suppress this maximum depending on the magnetic declination and thus geographic location.
Figure 8.6 Contours of median NmF2 from ionosonde observations (top) and E-CHAIM (bottom) between 50°N and 55°N AACGM latitude plotted against local time and F10.7 solar flux for a) December, b) September, and c) June. Please note that white space in the above plots corresponds to an absence of data.

For all three periods presented in Figure 8.6, E-CHAIM demonstrates a good fit to the ionosonde dataset, where all of the aforementioned behaviours are present within the model.
For a better indication of seasonal variability and winter/semi-annual anomaly behaviour, we have included a figure similar to Figure 8.6 but plotted in terms of season (in place of local time) for local midnight and local noon periods in Figure 8.7.

![Figure 8.7](image)

**Figure 8.7** Contours of median NmF2 from ionosonde observations (top) and E-CHAIM (bottom) between 50°N and 55°N AACGM latitude plotted against month and F10.7 solar flux for a) local noon, and b) local midnight.

At local noon, we note clear semi-annual or winter anomaly behaviour at all levels of solar activity and no apparent saturation of NmF2 at high solar activity. At local midnight, NmF2 reaches a maximum during the summer solstice and shows what appears to be a slight asymmetry of higher NmF2 during the spring equinox at moderate-to-high solar activities and a sudden enhancement in NmF2 during the fall equinox at extreme-high solar activity. Also, one will note a tendency for a saturation of NmF2 at high solar
activities during summer nighttime conditions. All of these features are well captured by E-CHAIM.

8.4.2 Auroral Oval

To examine the behaviour of NmF2 within the Auroral Oval, we present plots similar to those produced for upper mid latitudes (Figure 8.6 and Figure 8.7) but for a 5º bin centered about 72.5ºN AACGM latitude. We begin in the same manner as Section 8.4.1, with a plot of median NmF2 from all ionosonde and E-CHAIM observations within this bin for December (Figure 8.8a), September (Figure 8.8b), and June (Figure 8.8c) plotted against local time and solar flux.
Figure 8.8 Contours of median NmF2 from ionosonde observations (top) and E-CHAIM (bottom) between 70°N and 75°N AACGM latitude plotted against local time and F10.7 solar flux for a) December, b) September, and c) June.

From this figure, we note a narrow daytime enhancement in NmF2 during winter months associated with the very short period of daytime solar illumination at F-region altitudes during this period. We also note a stronger winter nighttime increase in NmF2 with increasing solar flux, as compared to the mid latitude case of Section 8.4.1. During the
September equinox period, we see behaviour similar to that of the mid latitude case. Interestingly, during the summer, we no longer see the midnight enhancement in NmF2 that is observed at mid latitudes in Section 8.4.1 and instead see behaviour more characteristic of simple photoionization.

For seasonal behaviour, we present the same plot as Figure 8.7, but instead for our Auroral Oval bin, in Figure 8.9.

![Figure 8.9](image.png) Contours of median NmF2 from ionosonde observations (top) and E-CHAiM (bottom) between 70°N and 75°N AACGM latitude plotted against month and F10.7 solar flux for a) local noon, and b) local midnight.

Similar to the mid latitude case, we see clear semi-annual anomaly behaviour during daytime periods. Interestingly, we begin to see the development of semi-annual anomaly behaviour during nighttime periods at high solar activity in the auroral oval region. Also, we see solar activity saturation of NmF2 at virtually all times.
8.4.3 Polar Cap

To examine the behaviour of NmF2 within the Polar Cap, we present plots similar to those produced for the previous two sections but for a 5° bin centered about 87.5°N AACGM latitude. We begin in the same manner as Section 8.4.1, with a plot of median NmF2 from all ionosonde and E-CHAIM observations within this bin for December (Figure 8.10a), September (Figure 8.10b), and June (Figure 8.10c) plotted against local time and solar flux.
Figure 8.10 Contours of median NmF2 from ionosonde observations (top) and E-CHAIM (bottom) between 85°N and 90°N AACGM latitude plotted against local time and F10.7 solar flux for a) December, b) September, and c) June.

During the winter, when photoionization is largely negligible at these latitudes, convection of dense plasma from lower latitudes forms the dominant source of plasma to the polar cap. Due to the anti-sunward behaviour of this and the thermospheric winds, we here still observe an appreciable diurnal variation in NmF2 during the winter in the polar
cap, centered about a few hours after local noon. During the equinoxes, we see a far more broad diurnal variation as compared to that at lower latitudes but still with diurnal maximum centered about or slightly after local noon. Finally, during the summer, we see only a week diurnal variation and a depletion in NmF2 in the morning sector, similar to that seen during winter periods.

For seasonal behaviour, we again present the same plot as Figure 8.7 but now for our polar cap bin in Figure 8.11.

**Figure 8.11** Contours of median NmF2 from ionosonde observations (top) and E-CHAIM (bottom) between 85°N and 90°N AACGM latitude plotted against month and F10.7 solar flux for a) local noon, and b) local midnight.

From Figure 8.11, we see that, in the polar cap, we no longer see semi-annual anomaly behaviour at low solar activity. During daytime periods, we see a single seasonal maximum in NmF2 at low solar activity and it is only at solar activities above ~150 sfu
that semi-annual anomaly behaviour begins to develop. During nighttime periods, we again see only a single maximum at low solar activity, but unlike during daytime periods, there is an apparent gradual propagation of the seasonal maxima toward the equinoxes as solar activity increases.

8.5 NmF2 Perturbation Model Parameterization

In order to match the functionality of the IRI, we have also included an ionospheric storm correction to the quiet-time model. Empirical storm models, such as those of Araujo-Pradere and Fuller-Rowell [2002], Araujo-Pradere et al. [2002], and Li et al. [2016], are commonly used with simple geomagnetic indices to model the general ionospheric response to increased geomagnetic activity. While these models cannot be expected to represent all of the storm behaviour observed at high latitudes, given its highly dynamic nature and coupling to the solar wind, they can provide information on the overall large-scale ionospheric response due to Joule heating [Foster et al., 1983].

The main driving parameters for this portion of the model were determined experimentally via trial and error and were selected as

\[ G = \left[ \text{Dst}', e^{-\frac{ap'}{30}}, e^{\frac{AE'}{700}} \right] \] (8.7)
where $G$ is a three-element array, $Dst'$ is the integrated hourly Dst index from the Kyoto World Data Center (WDC) for Geomagnetism (available at http://wdc.kugi.kyoto-u.ac.jp/), $ap'$ is the integrated three-hour ap index from the NGDC portal, and $AE'$ is the integrated hourly AE index gathered from the Kyoto WDC for Geomagnetism. The geomagnetic activity indices, used here, are integrated forms of the Dst, Ap, and AE indices, where integration is done in the same manner as Wu and Wilkinson [1995] with persistence factors of 0.95, 0.75, and 0.95, respectively. Persistence factors were determined purely by trial and error using the values provided in Perrone et al. [2001] and Wu and Wilkinson [1995] as starting points. The functions, $G$, of the above geomagnetic indices were arrived at purely through extensive trial and error.

Using a spherical cap harmonic expansion to represent horizontal variability in ionospheric storm response and the sine and cosine of the dipole tilt angle as a seasonal term, we have the following for the storm model parameterization

$$\log \left( \frac{NmF'2}{NmF2} \right) = \sum_{l=0}^{L} \sum_{m=0}^{\min(l,M)} \left[ A_{lm} \cos \left( \frac{\pi m}{180} \lambda \right) + B_{lm} \sin \left( \frac{\pi m}{180} \lambda \right) \right] P_{lm}(\eta) \quad (8.8)$$

$$A_{lm}, B_{lm} = \sum_{d=1}^{3} \left[ a_{lm} \sin \theta + \beta_{lm} \cos \theta + (\gamma_{lm} \sin \theta + \delta_{lm} \cos \theta) \sqrt{F10.7_{81}} \right] G_d \quad (8.9)$$

where $\lambda$ is magnetic local time, $\phi$ is geomagnetic latitude, $F10.7_{81}$ is the 81-day smoothed F10.7 solar flux, and $\theta$ is the magnetic dipole tilt angle. $\alpha, \beta, \gamma, \text{ and } \delta$ are fitting coefficients. $G_d$ refers to the $d$th element of the $G$ array, given by Equation 8.7.
The maximum order and degree of the expansion was set to five and three, respectively, for this portion of the model. The reduced degree of the spherical cap expansion is a consequence of the reduction in the quality of ionosonde data during geomagnetic storm events, which tended to exaggerate noise in the storm output. There are 360 coefficients for each of the 24 hourly UTC models.

8.6 Storm-time Performance and Examples

To demonstrate the performance of the storm/perturbation model, we have scaled ionograms from the Resolute, Cambridge Bay, Pond Inlet (72.69N, 262.04E), and Eureka (79.99N, 274.10E) stations. For these stations, we compare measured and modeled NmF2 for a moderate, and fairly long-lived, storm between May 21 and June 7, 2010, in Figure 8.12.
Figure 8.12 (top row of each quadrant) Ionosonde-measured (black), E-CHAIM modeled (blue), and IRI-modeled (red) NmF2 around the May 29th, 2010 geomagnetic storm at the Resolute (top left - RESC), Eureka (top right - EURC), Cambridge Bay (bottom left - CBBC), and Pond Inlet (bottom right - PONC) validation stations. (middle row of each quadrant) Differences between observations and the E-CHAIM (blue) and IRI (red) modeled observations for the corresponding stations. (bottom row of each quadrant) Kp index for the periods presented.
It is clear from Figure 8.12 that the E-CHAIM perturbation model performs well in representing NmF2 during this storm at all four stations, with the exception of Eureka at storm onset, where we see a delay in the modeled storm response from both E-CHAIM and the IRI. We clearly see a strong negative phase response in observed NmF2 that is captured by both the E-CHAIM perturbation model and the IRI. Interestingly, the IRI also performs reasonably well during this storm, despite significant issues during the quiet periods preceding and following the storm; in fact, the IRI performs better during the storm period than during quiet periods. This, however, implies that the IRI is underestimating the storm response of the ionosphere with respect to quiet periods (e.g. the difference between storm and climatological NmF2). In terms of improvement over the climatological E-CHAIM model, the perturbation model results in a 33% improvement at Resolute, a 22% improvement at Eureka, an 18% improvement at Cambridge Bay, and a 30% improvement at Pond Inlet.
For the auroral region, we also present NmF2 at Eielson for a pair of disturbed periods in June and October, 2013, in Figure 8.13. For the June period, we again see a tendency for the IRI to underestimate quite-time NmF2 and see much better performance from the E-CHAIM model. Total RMS errors during these periods (listed in the figure headers) show a 30-35% improvement over the IRI through the use of E-CHAIM, mostly coming from quiet-time improvements. During storm periods, however, both models converge to similar density representations. For the October period, we see much of the same. There are, however, a few interesting features in the October period:

1) a relatively weak increase in geomagnetic activity on October 7th appears to generate a strong ionospheric response that is not captured by either model;
2) E-CHAIM shows slightly better performance during the Kp 7+ geomagnetic storm on October 2nd, where E-CHAIM matches the main phase and early recovery phase to within 0.5e11 while the IRI overestimates by up to 3e11. During the remainder of the storm recovery phase, E-CHAIM shows only minor improvements over the IRI of about 0.5e11 before becoming virtually identical to the IRI during the post recovery quiet period;

3) neither model captures the enhanced (above climatological) daytime density proceeding the recovery from the October 2nd storm; and

4) E-CHAIM correctly captures the quiet-time density following the last storm of the period.

Features 1 and 3, to some extent, highlight the limitations of empirical storm models in their application at high latitudes; nonetheless, these first order adjustments constitute a significant improvement over the climatological density. For the periods at Eielson, the perturbation model resulted in 24% improvements over the climatological E-CHAIM model. Further improvements to storm NmF2 representations, to account for features like patches, will likely need to come in the form of data assimilation, perhaps using an empirical background model, such as E-CHAIM.
9 Modeling the Peak Height of the Ionosphere: hmF2

The electron density of the complete E-CHAIM model will be parameterized with respect to both the peak density and height of the ionosphere, which necessitates the development of an accurate hmF2 model. This chapter discusses the development and validation of the E-CHAIM hmF2 model and is based on the author’s work, published in Themens et al. [2017a].

9.1 Data

The bulk of the ionosonde data included in the following model is derived from POLAN- or ARTIST-inverted ionograms [Titheridge, 1988; Huang and Reinisch, 2001]. Due to the additional difficulty in processing hmF2 from ionograms, particularly at high latitudes, additional quality control measures were necessary in order to ensure outliers and poorly processed data were removed [Moskaleva and Zaalov, 2013]. To this end, in addition to the quality control measures applied to the foF2 data, when original ionograms were available, suspect auto-scaled data were verified visually by an experienced ionogram scaler and corrected, if necessary.

In addition to directly inverted ionosonde hmF2 data and GPS RO data, we have also included hmF2 data inferred via the Bradley-Dudeney method, as provided by the SPIDR
system [Bilitza et al., 1979]. This data was found to be consistent with that derived from ionogram inversion and was thus included in the model fitting.

![Figure 9.1 Plot of the global distribution of ionosondes used in the creation of the hmF2 phase of E-CHAIM. The dashed line corresponds to the lower boundary of the model at 50°N geomagnetic latitude. The dotted line corresponds to 65°N geomagnetic latitude. A map of the ionosonde stations used for the hmF2 portion of the model is presented in Figure 9.1, where red filled circles mark the locations of stations included in the fitting and X’s mark the locations of stations used in the validation of the model. Note that some validation station locations coincide with the location of a station used in the model fitting. This is due to there being previous stations at those locations prior to the stations used for validation (ex: Resolute). Overall, ~15 million data points are used in the fitting of the hmF2 model, spanning seven solar cycles. Similar to the quiet-time NmF2 model, only data corresponding to Kp values less than 3.5 are included in the fitting dataset for the quiet-time hmF2 model.]}
9.2 hmF2 Model Parameterization

The hmF2 model features a similar parameterization to that for climatological NmF2 with some adjustments to take into account different solar activity scaling, a greater solar zenith angle and dipole tilt dependence, and a significant reduction in the amount of high quality data available for model fitting. The explicit parameterization is given as:

\[
hmF2 = G + \sum_{l=0}^{L} \sum_{m=0}^{\min(L,M)} [A_{lm} \cos\left(\frac{\pi m}{180} \lambda\right) + B_{lm} \sin\left(\frac{\pi m}{180} \lambda\right)] P_{lm}(\eta) \tag{9.1}
\]

\[
\eta = \cos\left(90 - \varphi\right) \frac{\pi}{45} \tag{9.2}
\]

\[
A_{lm}, B_{lm} = (\gamma_{lm} F_1) \cdot \sin^2\left(\frac{\pi \cdot \text{DoY}}{365.25}\right) + (C_{lm} F_1) \tag{9.3}
\]

\[
C_{lm}, D_{lm} = \sum_{c=1}^{K} a_{lm}^c \cos\left(\frac{2\pi c \cdot \text{DoY}}{365.25}\right) + \beta_{lm}^c \sin\left(\frac{2\pi c \cdot \text{DoY}}{365.25}\right) \tag{9.4}
\]

\[
G = F_{10.7} \cdot (a_1 \cos(\chi) + a_2 \sin(\chi)) + \sqrt{F_{10.7}} \cdot (a_3 \cos(\chi) + a_4 \sin(\chi)) + IG \cdot (a_5 \cos(\chi) + a_6 \sin(\chi)) + a_7 F_{10.7}^2 \cos(\chi) + a_8 IG^2 + \cos(\chi) \cdot [a_9 \sin(\theta) + a_{10} \cos(\theta)] + \sin(\chi) \cdot [(a_{11} \sin(\theta) + a_{12} \cos(\theta))] \tag{9.5}
\]

\[
F_1 = F_{10.7}^{0.275} \tag{9.6}
\]

where \(a, \beta, \gamma,\) and \(a_{1-12}\) are fitting coefficients. The solar activity scaling term \(F_1\) was chosen based on trial and error by iterating through powers of \(F_{10.7_{81}}\) between 0.1 and 230.
4.0 in steps of 0.025. Due to the increased noise, decreased reliability, and poor
distribution of hmF2 data, particularly, at high latitudes, the hmF2 model features a
drastically reduced spatial order and degree of three and three (L = 3, M = 3),
respectively, but the same, quintenial, seasonal Fourier Expansion (K = 5) as the NmF2
model. This model has 122 coefficients for each of the 24 UTC hourly models. Despite
the drastically reduced spherical cap expansion, the model still demonstrates significantly
more spatial structuring than the IRI2007 hmF2 model, which can be seen in Figure 9.2,
where we present maps of the E-CHAIM- and IRI2007-modeled hmF2 for 2013, similar
to Figure 8.4.
Figure 9.2 Contours of hmF2 from E-CHAIM (A) compared to that generated by the IRI (B), for representative months in 2013 at 00 (top row), 08 (middle row), and 16 (bottom row) UTC.
In Figure 9.2, there is drastically more structuring in the E-CHAIM representation than in the IRI maps, despite the relatively low order and degree of the spherical cap expansion. Please note that IRI2007 was used here as it features an identical hmF2 parameterization as IRI2012. While IRI2016 features the Shubin [2015] and Altadill et al. [2013] hmF2 models, at the time this work was completed, IRI2016 had only just been released. Regardless of this, the authors felt that, for consistency with the analysis of Section 6.1, the use of the pre-2016 model was sufficient here. Future studies will examine comparisons between the new IRI2016 hmF2 parameterizations and that of the E-CHAIM model.

There are several interesting differences between the E-CHAIM and IRI spatial patterns in modeled hmF2. The IRI demonstrates a local maximum in hmF2 over the geographic pole during summer periods, not seen in the E-CHAIM results. The E-CHAIM shows an enhancement of this sort; however, its enhancement is located closer to the geomagnetic pole and occurs primarily during winter and equinox periods. E-CHAIM also shows a deep minimum/trough in hmF2 in the auroral region, particularly during the winter, while the IRI demonstrates no such feature. In general, E-CHAIM produces lower dayside hmF2 compared to the IRI and higher nightside hmF2. Also, at 08 UTC during spring and summer periods, the IRI shows an enhancement in hmF2 at local noon that stretches from the pole to mid latitudes. In E-CHAIM, a comparable enhancement is seen; however, this enhancement only extends to the auroral region, with a subsequent hmF2 minimum/trough at lower latitudes.
9.3 Model Validation

To validate the hmF2 portion of the model, we use hmF2 values derived from ionosondes at Resolute, Cambridge Bay, and Eielson. Comparisons between observed, E-CHAIM, and IRI hmF2 are presented in Figure 9.3.

![Figure 9.3](image)

Figure 9.3 Ionosonde-measured (left column), E-CHAIM-modeled (middle column), and IRI-modeled (right column) hmF2 for the Eielson (A), Cambridge Bay (B), and Resolute (C) validation stations.

Immediately evident here is a drastic improvement over the IRI in the hmF2 representation at Eielson by E-CHAIM during daytime periods, where the IRI appears to consistently overestimate hmF2. This improvement is likely associated with E-CHAIM’s representation of an auroral hmF2 trough, demonstrated in Figure 9.2, which is represented in the IRI at lower latitudes and as a much broader structure. During the
nighttime at Eielson, we see a slight improvement through using E-CHAIM; however, both models are reasonably comparable during these periods. At Cambridge Bay and Resolute, we see a much deeper seasonal nighttime minimum in the winter. Also, observations show a daytime hmF2 maximum during the winter at low solar activity transitioning into a maximum in each equinox at higher solar activity. The IRI shows this single maximum in the winter at low solar activity but does not make the transition into maxima in each equinox. E-CHAIM shows much stronger performance in representing these features, particularly at Resolute.

Over the periods presented in Figure 9.3, the overall RMS errors for hmF2 produced by E-CHAIM are roughly the same for all three sites at between 13.0km and 13.5km; however, the IRI demonstrates RMS errors of 15.7 km at Resolute, 18.5 km at Cambridge Bay, and 22.2 km at Eielson. The seasonal patterns in these RMS errors are presented in Figure 9.4. At Resolute, we see only a marginal improvement in hmF2, mostly concentrated in winter and summer periods, and E-CHAIM is otherwise largely comparable to the IRI. At Cambridge Bay, we see much more significant improvements, particularly during winter periods, where errors are reduced by as much as 20 km. At Eielson, we again see significant improvement over the IRI by as much as 20 km, this time during equinox periods, which are largely associated with the IRI’s overrepresentation of daytime, equinox maxima. Overall, both models perform reasonably well at the test locations; however, the E-CHAIM shows significant improvements over the IRI2007 representation in the auroral oval region.
Figure 9.4 Monthly RMS errors in E-CHAIM (solid line) and IRI (dashed line) hmF2 at the Resolute (top), Cambridge Bay (middle), and Eielson (bottom) validation stations.
10 The Topside Model

The current standard for topside specification within the International Reference Ionosphere (IRI) and NeQuick empirical electron density models, as detailed in Section 6.1.3.1, is the NeQuick2 topside parameterization. This parameterization is a semi-Epstein layer function with a varying topside scale height given by Equation 6.6, where the scale height exhibits roughly linear behaviour near the F-region peak and asymptotic behaviour at several thousand kilometers above the F-region peak. While the IRI has two other topside options (IRI2001 and IRI2001-cor), the NeQuick option was recommended as the standard option based on an evaluation with Alouette and ISIS topside sounder data [Bilitza, 2009].

Figure 10.1 a) Plot of NeQuick scale height vs. altitude for the following r and g combinations (r,g): red (20,0.2024), black (100,0.125), blue (100,0.5), and green (100,1.0). Dotted lines represent the asymptotes for the corresponding curve. b) Plot of...
NeQuick scale height vs. altitude subsection of the domain presented in a), where dotted lines now represent the corresponding $H = Ho + g(h-h_{\text{max}})$ curve.

An illustration of the behaviour of the scale height function given by Equation 6.6 is provided in Figure 10.1. For a given $Ho$ value, here chosen to be 60km for illustration purposes, the $g$ parameter controls the rate of increase of the scale height near the F-region peak. To illustrate this, we have plotted dotted lines corresponding to $H = Ho + g(h-h_{\text{max}})$ in Figure 10.1b. The $r$ parameter controls the scale height at the asymptote, which is given by $H_{\text{asymptote}} = (r+1)Ho$. In Figure 10.1a, the asymptotes derived from this relationship are plotted as dotted lines.

Unfortunately, the choice of $r$ and $g$ constants has never been explicitly discussed in the literature, even in the original study that proposed the current NeQuick parameterization (e.g. Equation 6.6) [Radicella and Leitinger, 2001]. Given the significant dependence of the topside model on the choice of these parameters, the lack of previous discussion regarding their significance, and the fact that the choice of these parameters has not changed since the original publication of the model (despite changes to Equation 6.6 over the years), it is imperative that these parameters be adequately investigated if one wishes to properly evaluate and re-parameterize the NeQuick topside model.

The choice to parameterize $Ho$ as a function of separate parameters, $B2Bot$ and $k$, rather than directly parameterizing $Ho$, originates from the original NeQuick topside model that used $H=k \cdot B2Bot$ and allowed for the simple calculation of the total electron content as a function of only $NmF2$, $B2Bot$, and $k$ [Radicella and Zhang, 1995]. As this is no longer a
necessary functionality of the model, a direct parameterization of $H_0$ may also be in order.

This chapter explores the performance of the NeQuick topside parameterization in its use above 45°N geomagnetic latitude using an unprecedented amount of topside observations and proposes a new parameterization of the model that could be used with the IRI and NeQuick. In Section 10.2, we first assess the main challenges that must be addressed if one is to re-parameterize the model, namely: 1) the limitations of the $H_0$ parameterization in its ability to capture the variability of $H_0$, and 2) the limitations of the current topside function to capture the shape of the topside electron density profile. In Section 10.3.1, we address issue (2) by assessing the existing values of $r$ and $g$ and proposing new values. In Section 10.3.2, we re-fit the NeQuick parameterization of $H_0$, first in a manner that preserves the form of the original model and later in a new form. Finally, in Section 10.3.3, we present a proposed parameterization for the topside representation in E-CHAIM. The main body of work presented in this chapter is based on the author’s work in Themens et al. [2017b] and [2018].

10.1 Data

This study uses all available topside observations above 45°N geomagnetic latitude from several instruments including: Incoherent Scatter Radars (ISRs), GPS Radio Occultation (RO) satellites, and topside sounders.
10.1.1 Incoherent Scatter Radar (ISR)

For this study, ISR data was gathered from the Madrigal Database at [http://isr.sri.com/madrigal/](http://isr.sri.com/madrigal/). Typically, this data has not been used for topside model representations due to its limited altitude coverage and limited spatial coverage. In this study, we are mainly concerned with the adjustment of the NeQuick model, which is a single parameter function requiring only the specification of the topside thickness $H_0$. The fact that this is a single parameter model allows us to make effective use of data with even limited vertical extent, assuming that the NeQuick function itself can adequately represent the curvature of the topside electron density profile. To make use of the ISR electron density profiles, we undertook the following procedure in order to ensure the quality of the dataset:

1) Data has been binned in altitude bins, with thicknesses ranging from 10km in the bottomside to 50km in the topside, over 15 minute intervals, using only data from beams above 45 degrees elevation.

2) Probability distribution functions are created for each bin and then fitted with a Gaussian to determine the electron density and standard deviation of the measurements in each bin.
This procedure is undertaken in order to account for the tendency of ISR measurements to exhibit a positive bias in statistical averages of low signal-to-noise measurements. In these situations, the distribution of electron density measurements in these bins is often artificially truncated at zero and thus a simple average would be positively biased when the density is smaller than the noise width of the measurements. Overall, this procedure resulted in the identification of 960,450 electron density profiles from between 1963 and 2016 for inclusion in the model fitting. The locations of the ISRs used in this study can be found in Figure 10.2, where ESR refers to the EISCAT Svalbard Radar, PFISR refers to the Poker Flat ISR, and RISR refers to the northern face of the Resolute ISR.

Figure 10.2 A map of the locations of the ISRs used in this study with measurement density from all data sources superimposed (density is presented in 2° x 2° latitude x longitude bins).

10.1.2 Radio Occultation (RO)
This study also makes use of RO electron density profiles from the Gravity Recovery and Climate Experiment (GRACE) and FORMOSAT-3/COSMIC missions. GRACE, a pair of satellites launched at an orbital altitude of 500km in the same 89° inclination orbital plane, has provided RO data between 2007 and 2016. COSMIC is a constellation of six Low Earth Orbit (LEO) satellites with orbital altitudes of 700km to 800km and inclinations of 72°. All satellites, except satellite FM3, have been operational since launch in 2006. Satellite FM3 experienced power issues in 2010 and is no longer operational. COSMIC and GRACE data are available from the COSMIC Data Analysis and Archive Center (CDAAC) database at http://cdaac-www.cosmic.ucar.edu/.

LEO satellites determine electron density profiles up to their orbit altitude using occulting GPS TEC measurements with the Abel inversion technique [Schreiner et al., 1999]. The revised procedure of Pedatella et al. [2015] was used for the COSMIC and GRACE observations. These techniques can be susceptible to the effect of horizontal ionospheric gradients, which can be exacerbated at high latitudes [Pedatella et al., 2015]. To accommodate the limitations of LEO RO electron density observations, quality control of these datasets was undertaken in a manner similar to that of Shubin et al. [2013]. In addition to these quality control measures, we have also omitted profiles where the horizontal track of tangent points is longer than three times the vertical extent of the profile and profiles where negative electron densities are found above 100km altitude. ~735,000 profiles were identified for use in fitting the topside model.
10.1.3 Topside Sounder

We also make use of all available topside sounder profiles from above 45°N geomagnetic latitude. The missions included in this dataset are the Alouette 1 and 2 missions, the International Satellites for Ionospheric Studies (ISIS) 1 and 2 missions, and the Intercosmos-19 (IK-19) mission. Information about the orbit, operational period, and number of profiles included in this study are presented in Table 3.

**Table 3** Summary of the characteristics of the topside sounder missions used in this study.

<table>
<thead>
<tr>
<th>Mission</th>
<th>Number of Profiles</th>
<th>Operational Period</th>
<th>Orbit Altitude (km)</th>
<th>Orbit Inclination (°)</th>
<th>Orbital Period</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alouette 1 (a,b,c)</td>
<td>3806, 15250, 16584</td>
<td>1962 - 1971</td>
<td>1000</td>
<td>80</td>
<td>~105min</td>
</tr>
<tr>
<td>Alouette 2</td>
<td>3768</td>
<td>1965 - 1972</td>
<td>500 - 3000</td>
<td>80</td>
<td>~120min</td>
</tr>
<tr>
<td>ISIS 1</td>
<td>7974</td>
<td>1969 - 1980</td>
<td>500 - 3500</td>
<td>88</td>
<td>~130min</td>
</tr>
<tr>
<td>ISIS 2</td>
<td>28732</td>
<td>1971 - 1979</td>
<td>1400</td>
<td>88</td>
<td>~115min</td>
</tr>
<tr>
<td>IK-19</td>
<td>400</td>
<td>1979 - 1982</td>
<td>500 - 995</td>
<td>74</td>
<td>~100min</td>
</tr>
</tbody>
</table>

For the Alouette and ISIS data, only data that was manually interpreted by the Canadian Communications Research Center (CRC) and had a quality control index better than five were included in the model fitting. The Alouette and ISIS data are available at https://spdf.sci.gsfc.nasa.gov/pub/data/ and the IK-19 data is available at http://www.izmiran.ru/projects/IK19/. Overall, ~77,000 topside sounder profiles were included in the fitting of the topside model.
Contours of measurement density are superimposed over the ISR location map in Figure 10.2. The number of measurements is presented in 2° x 2° bins.

10.2 Limitations of the Current NeQuick Topside Model and its Parameterization

Using the data listed in Section 10.1, we here present an evaluation of the performance of the NeQuick topside model, both in terms of its ability to correctly model the $H_0$ scale height and its ability to properly capture the curvature of the topside electron density profile.

10.2.1 $H_0$ parameterization

The variable component of the NeQuick topside model is its $H_0$ parameter, given by Equations 6.7 – 6.10. To assess the performance of this parameterization, we have fitted the entirety of the dataset from Section 10.1 to the NeQuick topside model (Equations 6.4 – 6.6) in terms of $H_0$ using non-linear least squares. In this way, we have derived $H_0$ for each profile. We then calculated the expected $H_0$ from Equations 6.7 – 6.10 using CCIR $f_0F_2$ and $M(3000)F_2$, as well as $hmf2$ from IRI2007. The use of this older $hmf2$ parameterization is to remain consistent with the $hmf2$ model used in the original fitting of the NeQuick topside model. This $hmf2$ should be consistent with that in the NeQuick,
which also uses a Bradley-Dudney-type hmF2 parameterization [Nava et al., 2008; Bilitza et al., 1979]. Given the IRI’s transition away from Bradley-Dudney-type hmF2 models to new, standalone, hmF2 models (beginning in IRI2016), it is expected that there may be some inconsistencies in its use of the NeQuick topside model. That however, is a topic for a different study and goes beyond the scope of the current work.

**Figure 10.3** Plots of monthly median $H_0$ derived from measurements (black), from the traditional IRI/NeQuick (red), and from refitting the NeQuick $k$ parameterization (green) for 00LT, 06LT, 12LT, and 18LT local time hours.

Calculating $H_0$ in this way, we present a comparison between measurement-derived and NeQuick-modeled $H_0$ at various local times throughout the period of study. In Figure
10.3, we present monthly median measurement-derived and IRI/NeQuick-derived $H_o$ at several different local times. This figure demonstrates some significant issues in the current NeQuick $H_o$ parameterization: 1) a tendency for the model to underestimate the amplitude of solar cycle variability, resulting underestimation at high solar activity by as much as 60km, and 2) a tendency to underestimate month-to-month variability, particularly during the nighttime. While solar cycle variations are easy to extract from this figure, the sheer extent of the dataset does not allow for an easy interpretation of the smaller-scale variations, such as seasonal variations.
Figure 10.4 A) Plots of monthly RMS $H_0$ errors for the traditional IRI/NeQuick (red) and for the refitted NeQuick k parameterization (green) for 00LT, 06LT, 12LT, and 18LT.
local time hours between 2007 and 2016. B) Same as Figure 10.3, but only for 2007-2016.

For the purpose of identifying features at shorter time-scales, we present a subset of the time range of Figure 10.3, namely from 2007 to 2016, in Figure 10.4 along with plots of the associated monthly RMS errors. It should be noted that the relatively weak solar maximum of this cycle (2012 - 2015) corresponds to the IRI/NeQuick’s best performing solar maximum. Similarly, the anomalously weak solar minimum of this cycle (2007 - 2010) corresponds to the IRI/NeQuick’s worst performing solar minimum. Based on this, the following evaluations during this period should be taken as a best-case scenario during solar maximum and a sort of worst case scenario at solar minimum. In Figure 10.4, we can more easily identify the solar activity and seasonal issues seen in Figure 10.3. More specifically, we see significantly damped seasonal variability in IRI/NeQuick-derived $H_0$ at all local times, with an almost complete lack of seasonal variability during midnight periods, despite ~10km seasonal variations in the measured values. At all local times, the IRI almost universally overestimates monthly median $H_0$, sometimes by as much as 25km. The exception to this is good agreement during solar maximum summer, particularly during dawn and dusk periods; however, given the results of Figure 10.3, this is likely limited to the relatively weak solar maximum of cycle 24, where stronger solar maxima see significant underestimation by the IRI/NeQuick-generated values.
Figure 10.5 a) Distributions of $H_0$ differences from measurement-derived values for the traditional IRI/NeQuick (red) and the refitted k parameterization (black). b) Mean (solid) and standard deviation ranges (dashed) of $H_0$ vs. geomagnetic latitude. c) Mean (solid) and standard deviation ranges (dashed) of $H_0$ vs. magnetic local time. d) Mean (solid) and standard deviation ranges (dashed) of $H_0$ vs. integrated AE index.

The distribution of the observed model-derived errors is also presented in Figure 10.5a. From this figure, we see an overall tendency for the model to overestimate $H_0$ by a median of 7.8km with lower quartile of -0.86km and upper quartile of 15.2km. The distribution also appears to be somewhat skewed toward large underestimation of $H_0$, possibly associated with the observed strong negative trend in geomagnetic activity errors.
To examine the spatial and geomagnetic activity variations of the observed errors in model-derived $H_0$ we have plotted the mean errors with respect to Geomagnetic Latitude, Magnetic Local Time (MLT), and integrated Auroral Electrojet (AE) index in Figure 10.5. The integrated AE index that is used here is calculated in the same manner as described in Section 8.5 and follows the methodology of Wu and Wilkinson [1995] and Perrone et al. [2001]. In terms of variations with respect to geomagnetic latitude, we see a tendency for the IRI to overestimate $H_0$ below 70°N by an average of ~10km and standard deviation of ~10km. Above 70°N we see relatively good agreement in the mean, with average underestimation of 0km to 3km; however, standard deviations increase to ~15km. In terms of MLT variability, we see mean overestimation during daytime periods of 8km – 9km with improvements during dawn and dusk periods, which is consistent with our observations from Figure 10.4. Most interesting in Figure 10.5 is the observed trend in errors with respect to integrated AE index, where we see a strong linear error trend that reaches an average underestimation of up to 40km at 1000nT. Of course, a median model such as the IRI/NeQuick should not be expected to be able to represent geomagnetic activity but the linearity of this trend suggests that these errors may be easily corrected through the inclusion of an integrated AE index term in the IRI/NeQuick’s $H_0$ parameterization.

10.2.2 Does $H_0$ need new parameters?
In the previous section, we demonstrated a propensity for significant errors in model-derived $H_0$, which are consistent with previous studies. The question remains, are these errors a consequence of there being a limited fitting dataset available for the development of the NeQuick model or do they arise from the inability of the parameters in Equation 6.8 to properly represent the variations in $H_0$? To assess this, we have re-fitted the NeQuick parameterization of Equation 6.8 to our entire dataset of $H_0$. The resulting model is given by the following

$$k = 3.927 - 0.2093f_oF_2 - 0.002979hmF_2 - 0.01311\frac{hmF_2}{B_2Bot} + 0.006925R_12$$

(10.1)

The results of using this new parameterization are presented as the green curves in Figure 10.3 and Figure 10.4 and as the black curves in Figure 10.5. Clearly from these figures, there is a significant improvement in $H_0$ modeled using the re-fitted parameterization. In terms of solar activity variability, we see significantly improved performance during low solar activity periods. Spatially, we see significantly reduced errors at low latitudes. Despite these improvements, we still see significant errors during the winter and an almost complete lack of seasonal variability during nighttime periods. These results demonstrate that the current form of the $k$ parameterization of Equation 6.8 is insufficient in its capacity to capture the variability of $H_0$, particularly during nighttime periods.

10.2.3 Topside Shape Errors
While limitations in the ability of the model to capture H\textsubscript{\textalpha} variability are a significant concern, we feel it necessary to examine the suitability of the model function itself (Equations 6.4 – 6.6) in representing the shape of the topside electron density profile. To assess this suitability, we have calculated the differences between IRI-derived electron density profiles and those measured by the various instruments discussed in Section 10.1. For this, we have generated probability distributions/histograms of the model errors for each altitude above the F-region peak. To ensure that we are solely examining the performance of the topside model, and are not also seeing contamination due to errors in hmF2 or NmF2, we have normalized all of the profiles to NmF2 and used hmF2-relative altitude.

![Normalized distributions of absolute electron density errors, as a percentage of NmF2, vs. altitude for the traditional IRI/NeQuick. Right) Normalized distributions of the corresponding relative errors in electron density vs altitude.](image)

**Figure 10.6** Left) Normalized distributions of absolute electron density errors, as a percentage of NmF2, vs. altitude for the traditional IRI/NeQuick. Right) Normalized distributions of the corresponding relative errors in electron density vs altitude.

In Figure 10.6 we present contour plots of these error distributions, normalized separately for each altitude. On the left, we plot the distribution of NeQuick model errors,
normalized to NmF2 (i.e. \( \text{NeQuick/NmF2} - \text{Measured/NmF2} \)). On the right, we have plotted the distribution of relative errors in electron density (i.e. \( \text{NeQuick} - \text{Measured} \)/\( \text{Measured} \)). Clearly, from this figure, it is evident that there are significant issues in the NeQuick topside electron density representation. In terms of NmF2-relative errors, we see a pattern of significant overestimation in the lower topside characterized by median overestimation errors of up to 8% of NmF2 and 1-sigma errors of up to 20% of NmF2. At higher altitudes, median errors decrease significantly to within 1% of NmF2 and 1-sigma errors decrease to ~6% of NmF2; however, the distribution of these errors becomes highly skewed toward underestimation. In terms of relative errors (e.g. Figure 10.6, right), the situation at high altitudes becomes somewhat concerning, where we see median relative underestimation by 42% at 500km above hmF2. Overall RMS errors for the IRI/NeQuick were found to be 10.34% of NmF2.

While the above error distributions provide more insight into the performance of the IRI/NeQuick topside model, they do not provide specific insight into the performance of the shape functions given by Equations 6.4 – 6.6. To examine this performance, we present a similar figure to Figure 10.6, where we now present error distributions for the best-fitted NeQuick profile (fitting only for \( H_o \)) in Figure 10.7.
Figure 10.7 Same as Figure 10.6 but for the best-fitted NeQuick parameterization \((r,g) = (100,0.125)\).

As one would expect, we see significantly reduced distribution spread and somewhat better median errors than the corresponding IRI/NeQuick profiles; however, one can clearly see some significant remaining issues, namely we still see a trend of overestimation in the lower topside and underestimation at higher altitudes. This is consistent with the results presented in Zhu et al. [2015]. The overall RMS error for this best fit NeQuick model is 4.50 % of NmF2. It should be noted that the discontinuity in the distributions at 200km - 250km is a result of the differences in the sampling ranges of the various instruments used in this study. Below 200km - 250km, the error distribution functions see a significant contribution from ISR observations. Above this altitude there is substantially less ISR data and the dataset is dominated by RO and topside sounder data.
Despite the above issue, this discontinuity can actually provide some measure of insight into the performance of Equations 6.4 – 6.6. As we are using least squares, if the model function cannot adequately reflect the shape of the profile to be fitted, the fitting approach will minimize fitting errors by overestimating in one region and underestimating in another (a consequence of only fitting for $H_0$). In this case, we see overestimation in the near-peak region and underestimation at higher altitudes. For the data with a greater vertical extent, the underestimation and overestimation regions are broader leading to an upward translation of the region of transition between underestimation and overestimation. If the topside model was fully capable of capturing the shape of the topside ionosphere, and the datasets do not exhibit systematic differences from one-another, we would not expect to see such a discontinuity in the error distributions presented in Figure 10.6. In the following section, we demonstrate this to be the case, assuaging the concern that the datasets may be inconsistent with one-another.

The vertical error trends in Figure 10.6 and Figure 10.7, combined with the observed issues in $H_0$ modeling presented in Sections 10.2.1 and 10.2.2, demonstrate a need to re-examine the formulation of the NeQuick topside model, both in terms of the $H_0$ parameterization and the choice of r and g constants.

10.3 Refitting and re-Parameterizing the NeQuick
The errors presented here, and in past studies, suggest a need to re-fit the NeQuick. As a first step to doing so, we first attempt to correct the model shape errors presented in Section 10.2.3.

10.3.1 New r and g constants

As the r and g constants of the NeQuick model have not been changed since the original implementation of the model and have not been publicly discussed in the past, we see optimizing these parameters as the main opportunity to correct the observed limitations of the current NeQuick topside shape. As a first step to optimizing these parameters, we have calculated the NmF2-normalized RMS errors for all the available data while varying r in steps of 1.0 and g in steps of 0.01. Similarly, we have also calculated the mean correlation between measured and modeled topside profiles varying through the same values of r and g.
The results of these tests are presented in Figure 10.8, where the X marks the location of the current NeQuick’s r and g constants and the + marks the location corresponding to the RMS error minimum and correlation maximum on the corresponding plot. As one can see, there is a modest improvement in NmF2-normalized error and correlation between the current NeQuick and the corresponding minimum/maximum. Despite these being relatively modest improvements, we will later show that these corrections can in fact completely correct the shape errors observed in Section 10.2.3. The RMS error minimum corresponds to \((r,g) = (20,0.20)\) while the correlation maximum corresponds to \((20,0.18)\). As a second assessment of these parameters, we have set \(r\) to 20 and re-fitted the NeQuick parameterization to our entire dataset for \(H_0\) and \(g\) using nonlinear least squares.
The corresponding mean g value from this fitting was found to be $g = 0.2024 \pm 0.043$. For the remainder of this study, the new r and g constants are taken as $(r,g) = (20,0.2024)$.

Revisiting Figure 10.1, we see that the traditional NeQuick model’s $(r,g)$ of $(100,0.125)$ results in a scale height that remains within 10% of the linear case up to 6000km above the F-region peak. As the data used in fitting these models generally comes from below 2500km above the F-region peak, the traditional parameterization would imply that only a linear scale height variation with altitude with a single constant rate of change ($g = 0.125$) would be necessary to fit the dataset. This does not, in fact, appear to be the case. Examining the red curve of Figure 10.1, corresponding to the optimized values for $(r,g)$ of $(20,0.2024)$, we see that the new parameterization quickly diverges from linearity at altitudes within the fitting dataset range.

**Figure 10.9** Same as Figure 10.6 but for the best-fitted new NeQuick model function using $(r,g) = (20,0.2024)$. 
To assess the performance of these new r and g constants, we have fitted Equations 6.4 – 6.6, for $H_0$, to our dataset using the new r and g constants. In Figure 10.9 we present error distributions for this best fitted model using new r and g values, similar to Figure 10.6 and Figure 10.7. Immediately evident from this figure is the near complete correction of the NeQuick’s near-peak overestimation, it’s high altitude underestimation, and the discontinuity at the boundary between ISR and RO/topside sounder dominated datasets. The overall RMS error for this new NeQuick model is 3.6% of NmF2. While this is a relatively small improvement, the tendency for the errors of the traditional model to constitute a systematic bias in different altitude regimes makes such an improvement significant nonetheless, particularly when such a model is used as a background or basis set for data assimilation frameworks [Yin and Mitchell, 2014; Bust and Mitchell, 2008] or to calculate the projection shell height for LEO topside TEC observations.

### 10.3.2 Re-fitting $H_0$

Now that the shape of the topside profile function has been updated/corrected, we may now re-fit the $H_0$ parameterization of Equations 6.7 and 6.8. As a baseline, we first re-fit the $H_0$ parameterization to the same parameters as the original NeQuick model (e.g. re-fitting the k parameterization). The RMS error in $H_0$ for this parameterization was found to be 11.74km. The results of fitting to the traditional parameterization are presented in Figure 10.10.
Figure 10.10 Same as Figure 10.5 but the black curves now represent the new $H_0$ parameterization of Equation 10.2 and the red curves represent the refitted traditional NeQuick k parameterization using the new $(r,g) = (20,0.2024)$.

Similar to the results from Figure 10.5, the use of the traditional parameterization of $H_0$ results in an underestimation of $H_0$ at high latitudes, underestimation of $H_0$ during dawn and dusk periods, and a linear pattern of underestimation with respect to geomagnetic activity. To attempt to resolve the observed error patterns, we have made three changes to how $H_0$ is parameterized: 1) we have directly fitted $H_0$, 2) we have removed all dependence on B2Bot, and 3) we have added additional terms to the parameterization.

The new best fitted parameterization is given by the following relationship
\[ H_o = 20.120 - 3.734 f_0F2 - 0.0993hmF2 + 0.226R12 + 23.636 \cos \chi + \\
0.639 \sin \chi + AE'(0.0691 \sin \varphi - 0.0442 \cos \varphi) - 1.227 \sin \left(\frac{LT-\pi}{12}\right) + 4.821 \cos \left(\frac{LT-\pi}{12}\right) \] (10.2)

where \( \chi \) is the solar zenith angle, \( AE' \) is the integrated AE index, \( \phi \) is the geographic latitude, and LT is the local time. Ap index was also attempted in place of AE in order to retain compatibility with the parameters already within the IRI; however, we found that Ap could not adequately represent the variability of \( H_0 \) in very active (Kp 7 or greater) conditions. The use of geographic latitude and local time instead of their corresponding geomagnetic counterparts was to accommodate the limitations of consistently available parameters within the NeQuick and IRI models. RMS errors for this newly fitted \( H_0 \) model were found to be 10.18km. We have chosen to refit \( H_0 \) directly here for several reasons: 1) fitting \( H_0 \) directly resulted in slightly lower RMS errors than re-fitting \( k \) with the same parameterization, and 2) as mentioned in Section 6.2.2 using improved B2Bot can have a negative impact on the performance of \( H_0 \) parameterizations; thus, fitting \( H_0 \) directly will reduce the impact of significant B2Bot errors and new B2Bot parameterizations, such as that of Alazo-Cuartas and Radicella [2017], on the topside model.

The performance of this new \( H_0 \) parameterization is presented as the black curves in Figure 10.10. Clearly, the resulting \( H_0 \) exhibits virtually no mean errors in latitude, reduced mean errors in MLT variability, and only marginal mean errors with increasing
geomagnetic activity. In Figure 10.11, we present error distributions for each altitude using the new parameterization.

**Figure 10.11** Same as Figure 10.6 but for the new NeQuick parameterization of Equation 10.2 with \((r,g) = (20,0.2024)\).

This figure demonstrates a significant improvement over the original results of the IRI that were presented in Figure 10.6. Median \(\text{NmF2}\)-normalized errors remain below 3% of \(\text{NmF2}\) at all altitudes with 1-sigma errors consistently below 10% of \(\text{NmF2}\). Overall RMS errors were found to be 7.16% of \(\text{NmF2}\). It should be noted that this is a regional alternative parameterization for the NeQuick topside and should not be used in regions below 50°N geomagnetic latitude.

**10.3.3 A Topside for E-CHAIM**
As with the other components of the E-CHAIM model, the E-CHAIM topside, developed here, is built around a spherical cap harmonic expansion in Altitude Adjusted Corrected Geomagnetic (AACGM) coordinates for horizontal spatial variability, a Fourier expansion for seasonal variability, and a function of F10.7 cm solar radio flux for solar cycle variability. As we are here developing a topside model for sole use with E-CHAIM, we have increased flexibility in the number and type of parameters available for use with the model, as compared to what is available when we were redeveloping the topside model to be used with both the NeQuick and IRI. It is also our intent to develop a topside model not dependent on ionospheric characteristics, such as hmF2 and foF2, so that future updates of models for those characteristics will not affect other components of E-CHAIM. Taking advantage of the above stated flexibility and avoiding the use of ionospheric characteristic terms, we propose the following model of a similar form to that used for the other components of E-CHAIM:

\[
H_0 = G + \sum_{l=0}^L \sum_{m=0}^{\text{min}(l,M)} \left[ A_{lm} \cos \left( \frac{\pi m}{180} MLT \right) + B_{lm} \sin \left( \frac{\pi m}{180} MLT \right) \right] P_{lm}(\eta) \quad (10.3)
\]

\[
A_{lm}, B_{lm} = (\gamma_{lm} F_1 + \delta_{lm} F_2) \cdot \sin^2 \left( \frac{\pi \cdot \text{DoY}}{365.25} \right) + (C_{lm} F_1 + D_{lm} F_2) \quad (10.4)
\]

\[
C_{lm}, D_{lm} = \sum_{c=1}^4 \alpha_{lm} \cos \left( \frac{2\pi c \cdot \text{DoY}}{365.25} \right) + \beta_{lm} \sin \left( \frac{2\pi c \cdot \text{DoY}}{365.25} \right) \quad (10.5)
\]

\[
G = F10.7 \cdot (a_1 \cos(\chi) + a_2 \sin(\chi)) + \sqrt{F10.7} \cdot (a_3 \cos(\chi) + a_4 \sin(\chi)) + a_5 \cos(\chi) + a_6 \sin(\chi) + \sin(\chi) \cdot (a_7 \sin \theta + a_8 \cos \theta) + \cos(\chi) \cdot (a_9 \sin \theta + a_{10} \cos \theta) + a_{11} \sin \theta + a_{12} \cos \theta \quad (10.6)
\]
where $MLT$ is AACGM local time, $L = 5$, $M = 5$, $F1$ is 81-day smoothed F10.7 flux, $F2$ is integrated AE index, $DoY$ is the day of year, $\theta$ is the dipole tilt angle, $\chi$ is the solar zenith angle, $a_{1.12}$, $\alpha_{lm}^c$, $\beta_{lm}^c$, $\gamma_{lm}$, and $\delta_{lm}$ are fitting coefficients, and

$$\eta = \cos\left( (90 - \varphi) \frac{\pi}{45} \right)$$

(10.7)

where $\varphi$ is AACGM latitude. MLT is used for the longitudinal coordinate to account for the limited spatial coverage of the available data set in a similar manner to Altadill et al. [2013]. The solar zenith angle terms partially accommodate for the short comings of this type of parameterization in representing UTC-dependent variations. RMS errors in $H_0$ using this parameterization are 9.05km. Overall RMS errors in electron density are 6.62% of NmF2. A comparison of the performance of this $H_0$ parameterization with respect to that of Equation 10.2 is presented in Figure 10.12 in a similar manner as used in Figure 10.5 and Figure 10.10.
Clearly, there are slight improvements in the use of the E-CHAIM parameterization over the new NeQuick parameterization, particularly with respect to diurnal variations, likely attributed to variations better represented via a geomagnetic coordinate system, and at high latitudes. Interestingly, the morning and evening periods where the NeQuick parameterizations appear to underestimate topside thickness are coincident with enhancements in quiet time ion temperature at high latitudes [Yamazaki et al., 2016], attributed to ion-neutral heating. Such changes in ion temperature, while directly influencing the topside thickness, are not reflected in the limited set of parameters used to drive the NeQuick models, namely NmF2, hmF2, and M(3000)F2. Biases in modeled $H_0$ are shown to remain within 1km in MLT and MLat. Biases only exceed 1km in
conditions with integrated AE index above 700nT and never exceed 5km. This improvement corresponds to vertical profile errors illustrated in Figure 10.13 below.

![Figure 10.13](image)

**Figure 10.13** Same as Figure 8.23 but for the E-CHAIM topside model.

### 10.4 Behaviour of the E-CHAIM Topside

Now that we have a reliable topside thickness model, we can begin examining the behaviour of high latitude topside thickness, which is largely unexamined in the literature aside from some more recent studies using topside sounder data [Benson et al., 2015]. Generally, the spatial behaviour of topside electron density has been examined using in situ satellite measurements or, as is the case in Benson et al. [2015], topside sounder data. All of these studies look at the absolute electron density without normalization to the F2-peak; thus, these studies have largely failed to examine the change in the spatial behavior of the topside profile shape. With the luxury of an unprecedentedly comprehensive
topside profile dataset, we will here take a moment to examine the spatial behaviour of the topside shape both seasonally and under disturbed geomagnetic conditions.

We begin this examination by presenting examples of the spatial characteristics of topside thickness at high latitudes, for quiet conditions, in Figure 10.14.

**Figure 10.14** Contours of quiet time topside thickness ($H_0$) from E-CHAIM for representative months in 2009 (A) and 2013 (B) at 00 (top row), 08 (middle row), and 16 (bottom row) UTC. Note the change in contour colour bar.
Evident is the tendency for the topside thickness to have a maximum in the morning sector and a second local maximum in the evening sector at high latitudes. This, perhaps, runs contrary to expectation; however, the work of Yamazaki et al. [2016] demonstrates ion temperature enhancements during the morning at high latitude locations, which could lead to topside thickness enhancements. In that study, these temperature enhancements are attributed to ion-neutral frictional heating, even during quiet periods, through simulations with the TIEGCM. We may also note that these maxima are enhanced during UT periods where midnight is located in the American sector. During the winter, aside from an enhancement in the auroral oval region, topside thickness appears to be greatest during nighttime periods and at a minimum during daytime periods. This behaviour is consistent with ionosonde observations of the near-peak bottomside thickness [Themens et al., 2014].

In terms of latitude, we see enhancements in the auroral oval that tend to be located near the expected location of the auroral electrojets, which would agree with an ion-neutral frictional heating production mechanism; however, these maxima are asymmetric, favouring the morning sector. While this is seen in ISR ion temperature, it is not reproduced by the TIEGCM, suggesting that another mechanism may be at play. Kosch and Nielsen [1995] demonstrate such an asymmetry in joule heating associated with the auroral electrojets and attribute its presence to the behaviour of ionospheric conductivity. Other than the auroral oval, we see enhanced topside thickness coincident with the MIT, which also tends to be coincident with increased ion temperature. We also note that the
topside thickness, in general, tends to increase with increasing latitude, with the exception of a local minimum at the geomagnetic pole.

To illustrate the behaviour of topside thickness during a geomagnetic storm we present snapshots of the topside thickness state during a Kp = 7 geomagnetic storm between May 30\textsuperscript{th} and June 2\textsuperscript{nd}, 2013, in Figure 10.15. This storm reaches a minimum Dst of -124nT at 8:00UTC on June 1\textsuperscript{st}, 2013.
Figure 10.15 Contours of the topside thickness progression of the June 1st, 2013, geomagnetic storm. The number at the top of each plot indicates the number of hours past 00UT May 30th, 2013. Plots are provided every eight hours at 00UTC, 08UTC, and 16UTC on each day, such that each row of the above plot can be taken as a day between May 31st (top) to June 3rd (bottom). The above plots are provided in AACGM latitude (radius), and magnetic local time (angle counter clockwise from the bottom of each plot).
During the geomagnetic storm, we see an increase in the topside thickness across the entire domain. During the early phase of the storm, we particularly see a strong enhancement at the auroral oval, roughly collocated with auroral electrojets. These enhancements near the auroral electrojets, like the quiet time behaviour, seem to favour the westward electrojet region. Thicknesses are observed to reach a maximum at the peak of the storm before slowly returning to quiet time values during the storm recovery phase. Interestingly, at 16 UT, the region of maximum thickness extends to much lower latitudes than during other periods, reaching the upper mid-latitude region, and rotates toward magnetic midnight. This rotation toward magnetic midnight also occurs during quiet periods and must not be taken as a feature distinct of geomagnetic storms.

The overwhelming tendency of the topside thickness to reach its diurnal maximum in the morning sector could be one reason why the traditional NeQuick topside thickness parameterization performed poorly during morning and evening periods. As the NeQuick topside thickness is largely just controlled by foF2 and M(3000)F2, whose CCIR formulations mostly follow solar zenith angle, it is unreasonable to expect that such a parameterization could reproduce variations that are a quarter cycle out of phase with the CCIR foF2 and M(3000)F2 parameters. In order to accommodate the observed trend of topside thickness peaking in the morning sector, one would have to add a separate local time term to the NeQuick parameterization.
11 The Bottomside Model

The bottomside of E-CHAIM is, perhaps, the most challenging component of the model. The bottomside is highly vertically structured and consists of layers with very different or independent physical dynamics. This difference in the behaviour of the ionosphere’s various bottomside layers and the tendency of some of these layers to only exist part of the time (F1-layer/ledge and E-region) necessitates that one takes separate considerations for each of these regions.

In light of this challenge, one would be tempted to try to re-fit current standards like the IRI and NeQuick; however, each of these formulations have been found to exhibit undesirable behaviour that we here wish to avoid.

11.1 The IRI Bottomside

The IRI bottomside is composed of five sub-regions illustrated in Figure 11.1. This decomposition of the bottomside is very complex and was designed to be flexible enough to accommodate the modeling of the bottomside globally, where different geographic locations demonstrate vastly different bottomside behaviour. We have already discussed the F2-bottomside portion of the IRI’s bottomside parameterization in Section 6.1.2, so we will limit our discussion here to the F1-region and below.
Figure 11.1 Illustration of the six components of the IRI bottomside profile. 1) The Topside, 2) the F2 bottomside, 3) the F1-layer/ledge, 4) An intermediate region between the F1-layer/ledge and the E-region valley, 5) the E-region valley, and 6) the bottom of the E-region and the D-region.

For the F1-layer/ledge, the IRI uses the foF1 parameterization of Ducharme et al. [1971, 1973] that represents foF1 as a nonlinear model of polynomial functions of magnetic dip latitude, solar zenith angle, and 12-month smoothed sunspot number ($R_{12}$). $hmF1$, the height of the F1-layer/ledge, is taken as the altitude of the F2-bottomside function at the electron density corresponding to foF1. As the F1-layer/ledge is not always present, this formulation also requires the use of an F1-layer/ledge occurrence trigger. For this purpose, the IRI uses the model of Scotto et al. [1998], which is illustrated in Figure 3.3. If an F1-layer/ledge is determined to be present, the electron density in the region of the
F1-layer/ledge is given by the same function as the F2 bottomside, but with a modified height term given in Bilitza [2003].

The remaining E-region is modeled in the same manner as the F1-layer/ledge with yet another different height term. $h_mE$, the height of the E-region peak, is taken as a constant value of 105km, and $f_oE$ is given by the parameterization of Kouris and Muggleton [1973,a,b]. This $f_oE$ parameterization is given as a product of four functions, each dependent on the solar zenith angle, 12-month smoothed F10.7 flux, and geodetic latitude.

Between the F1-layer/ledge and the E-region, the valley and intermediate regions are each simply modelled by yet more different height terms in Equation 6.3. In this way, the IRI bottomside is essentially a piece-wise continuous function with equal densities forced at the boundary of each region. An illustration of the IRI bottomside model with a comparison to ionosonde data at the Ramey ionosonde in Puerto Rico is presented in Figure 11.2. We present this figure to illustrate one of the more severe limitations of the IRI bottomside; namely, the fact that the vertical gradient of the electron density at the boundary of each region is not continuous. This poses significant challenges for users interested in ray tracing, as such applications require that the electron density profile has a continuous derivative. In Figure 11.2, you will note this tendency in the top panel, where we see a discontinuous and artificial “kink” in the IRI electron density profile at the boundary between the F1- and F2-layers.
Figure 11.2 Bottom) Examples of the IRI electron density profile function fitted to ionosonde observations at Ramey, Puerto Rico. Top) A zoom in to the F1-F2 transition for various F1-layer/ledge thicknesses. Taken from Reinisch and Huang [2000].

Due to the observed issue of discontinuous bottomside vertical gradients, we have chosen to avoid the use of the IRI bottomside formulation for our E-CHAIM bottomside.
11.2 Revisiting the NeQuick Bottomside

The NeQuick bottomside has already been discussed briefly in Section 6.2.1 due the model’s topside dependence on the bottomside; however, we will here take a moment to reflect on that formulation strictly in the context its implications for the bottomside shape.

As mentioned in Section 6.2.1, the NeQuick bottomside is given as the sum of three semi-Epstein layer functions, each with different topside and bottomside scale thicknesses. The NeQuick uses the foF1 model of Leitinger et al. [2005] for its foF1, which parameterizes foF1 as a piece-wise function of foE. foE in the NeQuick is then modeled using the work of Titheridge [1996]. As mentioned, in Section 6.2.1 hmE in the NeQuick is taken as a constant value of 120km and hmF1 is set as the average of hmF2 and hmE.

Unlike the IRI, the NeQuick bottomside model has a continuous vertical gradient profile, despite using discontinuous thickness changes at the layer peaks. This continuity is guaranteed, despite the thickness changes at each layer peak, because the derivative is zero at the peak by construction (a property of a semi-Epstein layer function), regardless of the thickness. This gives the NeQuick an advantage over the IRI when one wishes to use the model for HF ray tracing; however, the NeQuick has a separate challenge that makes fitting this model to an electron density profile quite difficult; namely, the requirement of an foF1 and foE specification even when there is no F1 Layer/Ledge or appreciable E-region present.
11.3 E-CHAIM Bottomside

We have used the challenges and successes of the above two bottomside formulations to inform our choice of bottomside model for E-CHAIM. In this way we have come to a formulation that, we believe, is robust and avoids the limitations of the IRI and NeQuick bottomside models.

11.3.1 Data

For the bottomside portion of the E-CHAIM model, we make use of global ionosonde profiles from the GIRO and CHAIN networks. We are here restricted to a substantially smaller dataset than what was available for the NmF2 and hmF2 portions of the model for a number of reasons, as many of the older datasets, which did not use digital ionosondes, did not invert electron density profiles from their data, instead scaling only foF2, MUF(3000)F2, and foE. In addition to this, the bottomside electron density profile, as a whole, is quite sensitive to scaling errors; thus, in addition to the quality control measures used for the hmF2 portion of the model (see Section 9.1) we have further restricted the dataset to data that has an ARTIST quality control score of 90 or greater and to data that has been manually scaled. This quality control has resulted in a dataset of
just over 6.2 million profiles. The stations used in constructing this dataset are represented in Figure 11.3.

![Figure 11.3](image)

**Figure 11.3** Same as Figure 8.1 but for the stations used in the E-CHAIM bottomside shape model.

As one will note, the dataset has significant holes, particularly over central Canada and over the Arctic Ocean. To deal with this issue, the bottomside portion of our model is constructed in a magnetic latitude and magnetic local time coordinate system. In this way, we essentially “fill in” these data gaps by allowing each station to constrain the model along a zonal line at its magnetic latitude. Using such a coordinate system reduces the model’s capability to reproduce geographically-dependent structures, such as non-migrating tides, but this is an unfortunate necessary compromise, which has been used in the past for the IRI’s current hmF2 model [Altadill et al., 2009].
11.3.2 Bottomside Function

To define the shape of the bottomside electron density profile within E-CHAIM, we have chosen to create our own, unique, formulation. Unlike the data used in creating the IRI and NeQuick, we are making use of ionosonde data that does not explicitly provide profile parameters and instead simply directly provides the electron density profile. This makes us somewhat less confined to the use of the traditional method of using the F1-layer/ledge and E-region densities as anchor points for our profile function and, in fact, doesn’t allow us to use this type of approach. In our formulation, we model the bottomside as a single semi-Epstein layer with an altitude-varying scale height. The functions used to define this scale height are themselves taken as a sum of semi-Epstein layers given below.

\[ H' = H_{F2} + H_{F1} \left[ \text{sech}^2 \left( \frac{h - h_{mF1}}{0.5(h_{mF2} - h_{mF1})} \right) \right] + H_{E} \left[ \text{sech}^2 \left( \frac{h - h_{mE}}{45.0} \right) \right] \]  

(11.1)

\[ H = H' \cdot \left[ \frac{1}{1 + \exp \left( \frac{h - h_{mE} - 25.0}{5.0} \right) } \right] \]  

(11.2)

where \( H_{F2}, H_{F1}, \) and \( H_{E} \) are amplitude terms, \( H' \) is an intermediate scale height function, and \( H \) is the true scale height used by the model. You will note that the scale height function of Equation 11.1 is composed of two semi-Epstein layer functions, centered at \( h_{mF1} \) and \( h_{mE} \), with predefined thicknesses. These thicknesses were determined by optimizing RMS fitting error against a set of test thickness values. In Equation 11.2, we...
arrive at the true scale height, where the scale height function of Equation 11.1 is multiplied by a sigmoid function. This sigmoid function acts to suppress the scale height below the E-region, ensuring that the electron density tends to zero at the model’s lower boundary.

The above formulation does not require an iterative approach in order to fit to electron density profiles and is explicitly doubly differentiable. The choice to model the bottomside shape in the scale height domain, rather than in electron density directly, was made based on two fundamental reasons:

1) The scale height contribution of the E-region and F1-layer/ledge smoothly transitions between the presence and absence of these layers, such that the amplitude terms ($H_{F1}$ and $H_E$) of Equation 11.1 smoothly progress from zero in the absence of these layers to a non-zero value when these layers are present. In the electron density domain, the IRI requires a trigger in order to distinguish between the presence and absence of these layers and the NeQuick was forced to parameterize foF1 as a function of foE and use a cyclic set of models.

2) The scale height is highly sensitive to subtle changes in the bottomside profile shape. This implies that structures in the scale height domain will be more dramatic than those in the electron density domain and are thus easier to identify. One would think that using such a method would amplify noise in the profile function, but because ionosonde data is unique in how it is inverted using an already assumed set of functions, ionosonde profiles already highly suppress
noise. In addition to this, large errors in scale height produce only minor errors in the resulting electron density.

The use of this method is not without its own set of challenges, however. As we are not using the conventional foE and foF1 anchor points, we must determine the amplitude terms of Equation 11.1 by fitting electron density profiles to our model function. This process in itself is a significant challenge for multiple reasons:

1) The E-region and F1-layer/ledge can be ambiguous to a fitting algorithm (i.e. it doesn’t know which is which). To avoid this ambiguity issue, we pre-define hmE and hmF1 in our least squares fit to the electron density profiles.

2) The profile function is highly non-linear and contains correlated components, creating a situation whereby there are multiple attractors in the least squares fit. This is addressed by ensuring that the fitting is well pre-conditioned and by reducing the number of fitted parameters by pre-defining hmE, hmF1, E-thickness (45 km), and F1-thickness (0.5(hmF2-hmF1)) before fitting.

11.3.3 hmE and hmF1

As mentioned above, we have chosen to independently develop models for hmE and hmF1. Beginning with hmE, we attempted to fit our dataset of ionosonde-derived hmE values to a model similar to that done for E-CHAIM’s topside thickness. In so doing, we found RMS fitting errors hardly 0.5km improved over the standard deviation of the input
data. This implies that even the most sophisticated modeling approach would only amount to a marginal improvement over simply using the average of the input data (102.19 ± 5.02 km). Based on this result, we have decided to use this average value in place of an explicit model for \( h_{\text{mE}} \). It is interesting here to note that both the IRI and NeQuick use constant values for \( h_{\text{mE}} \) in their model; however, while the IRI’s value of 105 km falls within the error range of the average derived here, the NeQuick’s value of 120 km is substantially higher than the average by several standard deviations. This result may suggest a need to re-visit the NeQuick’s choice of \( h_{\text{mE}} \).

Unlike \( h_{\text{mE}} \), \( h_{\text{mF1}} \) demonstrated notable coherent variability, such that a model could provide a substantial improvement over the mean (175.31 ± 15.01km). For \( h_{\text{mF1}} \), we have used a similar framework to that used for the topside thickness but with a simplified AE index component. The full parameterization is given as

\[
h_{\text{mF1}} = G + \sum_{l=0}^{L} \sum_{m=0}^{\min(l,M)} F_{l,m} \left[ A_{l,m} \cos \left( \frac{\pi m}{180} \cdot MLT \right) + B_{l,m} \sin \left( \frac{\pi m}{180} \cdot MLT \right) \right] P_{l,m}(\eta) \tag{11.3}
\]

\[
C_{l,m}, D_{l,m} = \sum_{c=1}^{3} a_{l,m}^c \cos \left( \frac{2\pi c \cdot DoY}{365.25} \right) + \beta_{l,m}^c \sin \left( \frac{2\pi c \cdot DoY}{365.25} \right) \tag{11.4}
\]

\[
G = a_1 \cos(\chi) + a_2 \sin(\chi) + a_3 \chi + a_4 \chi^2 + a_5 \chi^3 \\
+ \sin(\chi) \cdot (a_6 \sin \theta + a_7 \cos \theta) + \cos(\chi) \cdot (a_8 \sin \theta + a_9 \cos \theta) + a_{10} \sin \theta + a_{11} \cos \theta \\
+ AE' (a_{12} \sin \varphi + a_{13} \cos \varphi) \tag{11.5}
\]
where $F1$ is 81-day smoothed F10.7 flux, $F2$ is integrated AE index, $DoY$ is the day of year, $\theta$ is the dipole tilt angle, $\chi$ is the solar zenith angle, $a_{1-12}$, $\alpha_{lm}$, and $\beta_{lm}$ are fitting coefficients, and $\eta$ is given by Equation 10.7.

To assess the validity of the E-CHAIP hmF1 model fit, we present a comparison between hmF1 derived from E-CHAIP, the NeQuick parameterization (using measured hmF2), and the “traditional” hmF1 parameterization [Bilitza, 1990], currently used by ICEPAC, given by

$$hmF1 = 165.0 + 0.6428 \cdot \chi$$  \hspace{1cm} (11.6)

A comparison of the median behaviour of hmF2 from these various models is presented in Figure 11.4. To illustrate the best possible representation based solely on solar zenith angle, we have also included a best fit to a quadratic function in solar zenith angle given by the following

$$hmF1 = 179.91 - 0.43804 \cdot \chi + 0.0056225 \cdot \chi^2$$  \hspace{1cm} (11.7)
Figure 11.4 Plot of mean hmF1 behaviour vs. AACGM latitude (top left), AACGM local time (top right), integrated AE index (bottom left), and solar zenith angle (bottom right). Ionosonde measured values are represented by black squares, E-CHAIM is represented by connect black stars, NeQuick is represented by connected blue stars, the “traditional” parameterization is represented by connected red stars, and a best fitted quadratic function in solar zenith angle is represented by connected green stars.

Examining first the patterns with respect to solar zenith angle, we see that the E-CHAIM and quadratic model fit the dataset’s solar zenith angle behaviour very well. Also evident is the apparent inability of a linear parameterization in solar zenith angle to represent the behaviour of hmF1, which appears to be largely unchanged except at large solar zenith angles. The NeQuick appears to perform well when the sun is low to the horizon but poorly during mid-day. In terms of latitude, the ionosonde data demonstrates a largely linear trend with hmF1 increasing toward the geomagnetic pole. This pattern is well
captured by the E-CHAIM fit. The best fitted quadratic model seems to underestimate the rate of increase with latitude, while both the NeQuick and the “traditional” parameterization are biased upward by ~8 km and ~30 km, respectively. In local time E-CHAIM and the ionosonde data demonstrate a local minimum at local noon and an asymmetric diurnal variation with higher morning hmF1 compared to the evening. As for the remaining models, the traditional approach appears to again be over estimating hmF1 by ~30 km and the NeQuick model, while performing well at in the morning and evening, overestimates hmF1 around local noon. Interestingly, we see here that the best fitted solar zenith angle curve underestimates the amplitude of the diurnal variation of hmF1 and performs better during daytime periods than during the evening and morning periods, likely because of the abundance of daytime data as compared to morning and evening data (fit heavily weighted toward local noon). The inability of this function to capture the diurnal variation of hmF1, suggests that hmF1 may have a significant dependence on neutral wind and composition behaviour, which may not simply follow a solar illumination-driven pattern. Finally, and perhaps most interesting, is the observed behaviour with respect to integrated AE index. Here we note a sudden increase in hmF1 from very quiet periods to more average periods with largely linear behaviour thereafter. Interestingly, the hmF2-based parameterization of the NeQuick seems to capture the trend with respect to AE index reasonably well, if one ignores the slight tendency for an upward bias. In fact, if one uses 105 km for hmE, in place of the NeQuick’s 120 km value, we see average AE index behaviour comparable to that of the E-CHAIM model. This is also true for the latitudinal behaviour of the NeQuick’s parameterization. This suggests that hmF2 may be a good target parameter when attempting to model hmF1. It
should be noted, however, that we have used measured hmF2 here, and thus the CCIR-based hmF2 of the NeQuick would not capture such variabilities, as it does not include a geomagnetic activity adjustment.

Table 4 RMS errors from each hmF1 modeling method tested.

<table>
<thead>
<tr>
<th>Method</th>
<th>RMS Error (km)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>15.01</td>
</tr>
<tr>
<td>E-CHAIM</td>
<td>9.47</td>
</tr>
<tr>
<td>NeQuick</td>
<td>17.89</td>
</tr>
<tr>
<td>NeQuick (hmE = 105km)</td>
<td>15.48</td>
</tr>
<tr>
<td>Traditional Model</td>
<td>32.32</td>
</tr>
<tr>
<td>SZA Quadratic Fit</td>
<td>14.43</td>
</tr>
</tbody>
</table>

Overall RMS errors from each method are listed in Table 4. One will note from this table that the RMS error of the traditional approach is significantly worse than the use of a constant mean value for hmF1 and that the NeQuick, using 120km for hmE also performs worse than a simple mean. In fact, if one used CCIR-derived hmF2 in the NeQuick parameterization, it can be presumed that these errors might have been much larger, given the errors in the IRI/NeQuick hmF2 presented in Section 6.1.1.4. Overall, E-CHAIM performs substantially better than the mean and the NeQuick. It also performs much better than functions that are based solely on solar zenith angle.

11.3.4 Fitting the Bottomside Function
Once an hmF1 and hmE model have been completed, we then begin fitting electron density profiles for the amplitude terms of Equation 11.1. This is done via nonlinear least squares with pre-specified values for the heights and thicknesses of the scale height layers. Ionosonde inversion errors are used to create a diagonalized measurement error covariance, the a priori covariances are also diagonalized, and all output amplitude values are constrained to be greater than zero. Examples of these fits to ionosonde-derived electron density profiles is provided in Figure 11.5.

Figure 11.5 Example ionosonde-derived electron density profiles at Cambridge Bay (a-c) and corresponding scale height profiles (d-e) for three situations: a,d) no E-region trace, b,e) no F1-layer/ledge trace, and c,f) a profile with all three layers present. d-f) also
demonstrate the various components of the scale height function, where the dotted line is the scale height function for the F1-layer/ledge, the dashed line is the constant term, the dash-dotted is the scale height function for the E-region, and the solid line is the final bottomside scale height function.

As you may note, our chosen profile function does an excellent job fitting ionosonde-derived electron density profiles under a robust set of conditions. That said, there are an enormous number of profile functions that will fit an electron density profile. The challenge is finding one that produces physically consistent parameter variability that can be easily fitted to a model. For example, we attempted to use a fourth order polynomial to represent the scale height in our model; however, multiple attractors and the fact that none of the parameters of the fit were tied to a physical phenomenon meant that the fitted parameters did not exhibit coherent variability that could be easily fitted to a model. Our approach, of course, is not without some caveats: the various layers will often overlap, causing unwanted correlations between the layer amplitudes. Nonetheless, through trial and error, we have found that simply fitting for the $H_E$, $H_{F1}$, and $H_{F2}$ parameters of Equation 11.1 produced satisfactorily coherent and model-able parameter variabilities. An example of the monthly median behaviour of these three parameters at the Dourbes ionosonde is presented in Figure 11.6. In this figure we clearly see the expected strong solar illumination control of $H_E$ and $H_{F1}$, as well as the expected inversion in diurnal $H_{F2}$ variability between winter and summer periods.
Figure 11.6 Contours of monthly median fitted $H_{F2}$ (top), $H_{F1}$ (middle), and $H_{E}$ (bottom) for the Dourbes ionosonde between 2006 and 2013.
11.3.5 Parameterization of the Bottomside Scale Height Amplitudes

To parameterize the bottomside scale height amplitudes ($H_E$, $H_{F1}$, and $H_{F2}$), we use the same methodology as that used for each of the other model components; namely, we have chosen to use three identical spherical cap harmonic expansions with Fourier expansions in day of year as Gauss coefficients. These parameterizations are functionally identical to that used for $hmF1$ (e.g Equations 11.3 – 11.5) but with a degree and order of $L = 5$ and $M = 5$. The choice to use identical parameterizations here was largely made for computational purposes to remove the need to build a new parameter basis set for each component of the bottomside model.

11.3.6 Validation

To validate the performance of the E-CHAIM bottomside model, we have gathered data from a selection of ISRs, which were not included in the fitting dataset, and compared model errors with respect to those found using IRI 2016. For the purpose of only examining the performance of the bottomside shape functions and parameterization, we have used measured $hmF2$ and $NmF2$ as inputs to E-CHAIM and the IRI in the following analysis. Also, to diminish the impact of the F2 peak, we have further normalized all of the resulting electron density profiles to $NmF2$ and $hmF2$. Figure 11.7 presents an example of the resulting normalized, peak-relative electron density profiles from the Millstone Hill ISR.

290
Figure 11.7 F2-peak normalized electron density profiles at the Millstone Hill ISR for various UT times between 2007 and 2015. Note that local time at Millstone Hill is UT – 4.8 hours.

In Figure 11.8 we present the differences between model and measured bottomside peak-relative density for the same Millstone Hill ISR dataset at four UT times (note Millstone local time is UT – 4.8 hours).
From this figure, we see that the E-CHAIM bottomside model demonstrates a tendency toward underestimation of the bottomside electron density, particularly at high solar activity and during summer daytime periods, indicative of an underestimation of the $H_F$ scale amplitude during these periods. The IRI, however, demonstrates a tendency to overestimate the F-region electron density during low solar activity periods, but otherwise, performs quite well. To better visualize the comparison of the performance between the E-CHAIM and the IRI we have plotted contours of the differences between IRI and E-CHAIM RMS errors in Figure 11.9. In this figure, negative values correspond to periods where E-CHAIM outperforms the IRI.
Figure 11.9 Differences between IRI and E-CHAIM bottomside electron density profile RMS errors. Negative values correspond to periods and locations where E-CHAIM outperforms the IRI.

From this figure, it is much easier to see that the IRI outperforms the E-CHAIM bottomside shape model in the F-region by a significant margin during evening periods at this location. Both models demonstrate comparable performance during other periods, where E-CHAIM generally performs better at low solar activity and the IRI performs better at high solar activity. Similar to Figure 11.9, we may also examine these performance differences at the Tromso ISR and RISR in Figure 11.10.
Figure 11.10 Same as Figure 11.9 but for RISR (left) and the Tromso ISR (right).

For the Tromso ISR, located within the auroral oval, we see significantly better performance from E-CHAIM over the IRI during daytime periods and generally comparable performance elsewhere. At RISR, we see improved performance from E-CHAIM during nighttime periods, but reduced performance compared to the IRI in the E-region during summer daytime periods. For a better impression of the overall performance of the E-CHAIM and IRI bottomside models, we present plots of the bottomside electron content (integral of the electron density up to hmF2) in Figure 11.11.
Figure 11.11 Bottomside electron content measured by the Millstone Hill (left) and Tromso (right) ISRs and that modeled by the IRI (red) and E-CHAIM (blue) between 2007 and 2015 at various UT times. Note that the Millstone local time is UT – 4.8 and that for Tromso is UT + 1.3.

In this figure, we see that both models perform reasonably well in their overall representation of the bottomside electron density when $N_{m}F_2$ and $h_{m}F_2$ are accurate. Errors in bottomside electron content, under these circumstances, did not exceed 1.0 TECU.

Overall, despite improvements over the IRI during winter and nighttime periods, E-CHAIM demonstrates a tendency to underestimate the thickness of the F-layer during summer daytime periods and, partially by extension, the density in the E-region. This issue could be attributed, at least in part, to the noisy nature of ionosonde-based electron density profile scale heights. Even in manually scaled datasets, noise in F2-peak scale
heights can exceed 15km. Ultimately, E-CHAIM's bottomside performs comparably to that of the IRI; however, future work should explore methods of improving these shortcomings. To that end, we will explore the possibility of modeling the bottomside entirely within the electron density domain and examine further refining data quality control measures in future versions of the model.

11.3.7 Challenges due to Precipitation

Despite the comparable performance between the IRI and E-CHAIM bottomside parameterizations, both models demonstrate a marked shortcoming in their ability to represent the bottomside electron density profile during periods of strong precipitation. For example, we have plotted electron density profiles from PFISR in Figure 11.12, which demonstrate reasonable performance by both models during periods not subject to significant precipitation and demonstrate comparably poor performance during periods of enhanced precipitation.
Figure 11.12 Example electron density profiles from PFISR without (top) and with (bottom) auroral precipitation structures. Solid lines correspond to measured profiles, dotted lines correspond to E-CHAIM, and dashed lines correspond to the IRI. Note that ISR hmF2 and NmF2 have been used here in place of the model values to facilitate comparison of just the profile shape.

We note that the IRI E-region, in the absence of significant precipitation performs quite well at all locations; however, during periods of precipitation, the IRI's auroral precipitation model sees any improvement over the climatology countered by worse performance during "false alarm" situations (i.e. situations where the IRI predicts auroral
precipitation but none is present). This trigger issue will be an important consideration in the adoption of such a precipitation model in E-CHAIM.

Particularly challenging in these situations is the fact that ionosondes are incapable of providing electron density profiles under strong precipitation conditions, where E-region densities are greater than those of the F-Region. This severely limits the datasets that can be used to model the bottomside under these conditions. In this way, the datasets that were used in the IRI and E-CHAIM for fitting the bottomside are biased against these conditions and, by construction, cannot represent these features. This is clearly seen in Figure 11.13, where we demonstrate the differences between the IRI and E-CHAIM bottomsides with respect to PFISR observations. Clearly, there is a marked challenge in both models to represent the precipitation-enhanced electron density that is regularly observed during night time periods at PFISR.
Figure 11.13 Model-to-measurement bottomside electron density errors for E-CHAIM (left) and the IRI (right) at PFISR between 2010 and 2013. Note local time at PFISR is UT – 9.8 hours.
12 Conclusions and Future Work

I have here presented the development, formalism, and validation of an empirical electron density model (E-CHAIMP) that I believe represents a substantial improvement over current standards in their application to high latitude regions. Despite this substantial body of work, there are still improvements and additions to the model that remain to be included, namely: the bottomside parameterization, while performing comparably to the IRI, should be further improved; and in light of this being a high latitude model, future work should strive to include an auroral E-region component to the model. Also, there remain a plethora of applications of the model that are waiting to be explored, like: including the model in the real-time IRI (IRI); exploring new physics of the high latitude ionosphere through attempting to identify new and interesting features revealed by E-CHAIMP; and applying the model in data assimilation and ray tracing frameworks for OTHR applications. The use of E-CHAIMP in data assimilation frameworks, in particular, represents one of the main reasons behind developing the E-CHAIMP model and will likely form the bulk of the model’s future applications.

12.1 Inclusion of an Auroral E-Region Model

As the domain of E-CHAIMP includes the auroral oval region, the model could not be considered wholly complete without the inclusion of an Auroral E-region model. For this purpose, we have discussed including the empirical auroral E-region model of Zhang and
Paxton [2008] with Dr. Zhang. Inclusion of the auroral E-region model should be a simple matter and an IDL code already exists. Validation and improvement of such an auroral E-region model can be accomplished through the use of our extensive ISR data set.

**12.2 hmF2 Storm Model**

You will note the absence of an hmF2 storm model in E-CHAIM. This arises due to the suspected poor quality of ionosonde hmF2 data during storm periods. During storms, there is a significant increase in the amount of ionospheric HF absorption. This absorption can result in the loss of the E-region ionogram trace and even partial loss of the F-region trace. The lack of these traces makes inverting such ionograms impossible, as inversion requires full knowledge of the profile below the F-peak; thus, if one were to invert hmF2 from these partial ionograms, it is likely that hmF2 will be overestimated due to not accounting for the signal retardation in the lower ionosphere. This can lead to an exaggeration of the storm-time increase in hmF2.

In order to accommodate the above issue, we will develop a storm-time hmF2 model that is instead based on ISR and GNSS RO data, which are not subject to this issue. The limitation in such a model is the greatly reduced spatial and temporal coverage of these alternative datasets; thus, a single MLat-MLT model function, similar to the E-CHAIM topside model, will be used.
12.3 Merging E-CHAIM with IRTAM

Incorporation of the E-CHAIM specification of the high-latitude ionosphere in the IRI, so as to produce an improved standard of the global ionospheric density distribution would be highly desirable; however, differences between the IRI and E-CHAIM are unbridgeable at the model junctions along E-CHAIM’s lower boundary. Use of such a hybrid model in scenarios that require density definitions across the models’ boundary (e.g. for ray tracing applications) would then be undesirable. Retraining the IRI expansions of \( \text{foF}_2 \) and \( \text{hmF}_2 \) in the E-CHAIM formalism can be considered, but this would require fundamental refitting of the entire IRI model, due to the internal dependence on \( \text{foF}_2 \) maps, which may be prohibitive.

There is, however, an attractive option of incorporating E-CHAIM maps of \( \text{NmF}_2 \) and \( \text{hmF}_2 \) in the IRI-based Real-Time Assimilative Model (IRTAM) [Galkin et al., 2012], which periodically generates updated IRI coefficients to transform the IRI into a better match with available measurements provided by contributing ionosondes of the Global Ionosphere Radio Observatory (GIRO) [Reinisch and Galkin, 2011]. By design, IRTAM gradually returns to the background IRI in those regions where sensor data are not available for assimilation. The E-CHAIM synthesized data grids can override the IRI background at high latitudes during IRTAM’s spatial expansion stage of analysis. IRTAM will then protect such a combined grid from discontinuity artifacts across
junctions in the same way it operates with GIRO data, by applying its iterative 2D
smoothing neural network algorithm and expanding the grid into the Jones-Gallet Gk
basis [ITU-R, 2009]. The resulting IRTAM will thus include all sensor data, the
improved E-CHAIM high latitude model, and the background IRI over no-coverage areas
in a single global specification that is continuous in time and space.

Including the E-CHAIM model in the IRTAM, as well as other assimilation frameworks,
will be pursued in future work.

12.4 New Physics

Finally, E-CHAIM provides a visualization of the spatial and temporal characteristics of
high latitude electron density that has not been previously available. This provides a
unique opportunity to pursue new physics, where unexpected behaviour of electron
density in E-CHAIM could be reproduced by physics-based models in order to infer the
physics resulting in those unexpected features. Some examples of this include the
morning sector and auroral oval maximum in topside thickness (see Section 10.4), the
short-term enhancement in NmF2 and TEC observed to be associated with short term
variations in solar flux (see Sections 6.1.1.3, 6.1.4.1, and 8.3), the saturation effect of
electron density with respect to solar flux (see Section 8.4), and the storm time behaviour
of each component of the model.
12.5 Further Validation

Like any model, future work will heavily involve further validating E-CHAIM against new data sources. The first of these validations will involve the use of Enhanced Polar Outflow Probe (ePOP) radio occultation measurements to independently validate the E-CHAIM hmF2, NmF2, and topside parameterizations. Other future validation datasets include in situ measurements from the CHAMP, GRACE, SWARM, and Defense Meteorological Satellite Program (DMSP) satellite missions, which can be used to test the combined performance of the NmF2, hmF2, and topside parameterizations.
Bibliography


D. Bilitza (1990), International Reference Ionosphere 1990, NSSDC 90-22, Greenbelt, Maryland.


CCIR (1982), CCIR atlas of ionospheric characteristics, Recommendations and Reports of the CCIR, 6(340-4), Geneva.


Evans, J. V. (1965), Cause of the mid-latitude evening increase in foF2, J. Geophys. Res., 70 (5), 1175-1185.


Hunsucker, R.D., and J.K. Hargreaves (2003), The high-latitude ionosphere and its effects on radio propagation. Cambridge University Press.


Lunt, N., L. Kersley, G. J. Bishop, A. J. Mazzella, and G. J. Bailey (1999b), The effect of
the protonosphere on the estimation of GPS total electron content: Validation using

Ma, G., and T. Maruyama (2003), Derivation of TEC and estimation of instrumental

Ma, R., J. Xu, W. Wang, and W. Yuan (2009), Seasonal and latitudinal differences of the
saturation effect between ionospheric NmF2 and solar activity indices. J. Geophys.

Ma, X. F., T. Maruyama, G. Ma, and T. Takeda (2005), Determination of GPS receiver
differential biases by neural network parameter estimation method, Radio Sci., 40,


MacDougall, J. W., and P. T. Jayachandran (2001), Polar cap convection relationships


Ionospheric peak height behavior for low, middle, and high latitudes: A potential
empirical model for quiet conditions – Comparison with IRI-2007 model. J. Atmos.

Maltseva O.A., N.S. Mozhaeva, and T.V. Nikitenko (2013), Comparison of model and

normalization and preliminary modeling results of total electron content during a
midlatitude space weather event, Radio Sci., 36, 351–361.

"GPS and Ionosphere" in W. Ross Stone (ed.), The Review of Radio Science 1996-

electron content, maximum electron density, and equivalent slab thickness at a
10.1016/j.jastp.2005.02.024

Mazzella, A. J., Jr. (2009), Plasmasphere effects for GPS TEC measurements in North


Rodger, A.S., R.J. Moffett, and S. Quegan (1992), The role of ion drift in the formation of ionization troughs in the mid- and high-latitude ionosphere – a review. Journal of Atmospheric and Terrestrial Physics, 54(1), 1-30.

Rush, C.M., M. PoKempner, D.N. Anderson, and F.G. Stewart (1982), The use of theoretical models to improve global maps of foF2, NTIA Report 82-93


Shubin V.N. (2015), Global median model of the F2-layer peak height based on ionospheric radio-occultation and ground-based Digisonde observations, Adv. Space Res. 56, 916-928, doi:10.1016/j.asr.2015.05.029


Watson, C. (2016). GPS Total Electron Content Techniques for Observing the Structure and Dynamics of the High Latitude Ionosphere, Doctoral Dissertation, Department of Physics, University of New Brunswick, Canada.


322
Curriculum Vitae

David R. Themens

Department of Physics
david.themens@unb.ca
University of New Brunswick
Fredericton, NB

EDUCATION

Ph.D., Physics, University of New Brunswick: Sept, 2013 – Present
B.Sc., Physics, University of New Brunswick: Sept, 2007 – May, 2011
  Major: First Class Honours in Physics, Minor: Mathematics

DISSERTATIONS

Ph.D.: Modeling electron density at high latitudes: Development of the Empirical Canadian High Arctic Ionospheric Model (E-CHAIM) - Supervisor: P.T. Jayachandran

M.Sc.: Testing and Evaluation of a Prototype Microwave Radiometer - Supervisor: F. Fabry


RESEARCH WORK EXPERIENCE

Evaluation of International Reference Ionosphere 2007 electron density products at high latitudes.
  Supervisor: P.T. Jayachandran
  Canadian High Arctic Ionospheric Network Summer Research Assistant (May-August, 2011)

Development of a GPS differential receiver bias estimation technique for application at high latitudes.
  Advisor: P.T. Jayachandran
  NSERC Summer Undergraduate Student Research Assistant (May-August, 2010)

Examining the effects of ionospheric electron content on GPS Signal-to-Noise Density Ratio.
  Advisor: P.T. Jayachandran
  NSERC Summer Undergraduate Student Research Assistant (May-August, 2009)
AWARDS AND HONOURS

Prof. Reinhold and Maria Kaiser Memorial Prize in Physics, 2017
URSI 2017 GASS Student Paper Competition Finalist - Honourable Mention
URSI Young Scientist Award, 2015 AT-RASC and 2017 GASS
NSERC Alexander Graham Bell Canadian Graduate Scholarship D (CGS D), 2014 – 2017
NBIF New Brunswick Graduate Scholarship, 2014 – 2017
UNB Board of Governors Award, 2014 – 2017
CAP-DASP Student Presentation Competition – 2nd Place, 2015, 2016
COSPAR Student Grant, 2014, 2016
Don Hornibrook Graduate Student Prize in Physics, 2015
O’Brien Fellowship, 2014
CMOS Student Bursary, 2013
McGill Graduate Excellence Award, 2012 – 2013
Canada Steamship Lines Award, 2011 – 2013
McGill Graduate Excellence Fellowship, 2011 – 2012
Canadian Space Agency Student Travel Grant, 2011
Walter Baker Memorial Prize in Physics, 2011
Tom & Parker Hickey Memorial Scholarship, 2010 – 2011
NSERC Undergraduate Student Research Assistantship, 2009, 2010
Edwin Jacob Special University Scholarship, 2007 – 2011
Dean’s List, University of New Brunswick, 2007 – 2011

PUBLICATIONS


**CONFERENCE PRESENTATIONS, POSTERS, AND CONTRIBUTIONS**


**Themens, D.R.,** and P.T. Jayachandran “Examining the use of the NeQuick bottomside and topside parameterizations at high latitudes” (Apr, 2017). 2017 European Geophysical Union Scientific Assembly, Vienna, Austria. (Oral Presentation)


**McCaffrey, A.,** D.R. Themens, and P.T. Jayachandran. “Comparison of IGS network GPS receiver DCBs provided by CODE and a single station estimation method” (June, 2015). 2015 Canadian Association of Physicists Congress (Poster Presentation)


**Themens, D.R., and F. Fabry.** "Constraining Temperature and Humidity for Convection Forecasting: On the necessity of having scanning ground-based instruments" (Feb, 2014). Presented at the 18th Conference on Integrated Observing and Assimilation Systems for the Atmosphere, Oceans, and Land Surface (IOAS-AOLS) as part of the 94th Annual Meeting of the American Meteorological Society (AMS) held in Atlanta, Georgia. (Oral Presentation)


**Themens, D.R., and F. Fabry.** The Mesoscale Microwave Radiometer: A Quest for Accurate, Three-Dimensional Water Vapour and Temperature Fields (May, 2013). Presented at the 2013 Canadian Meteorological and Oceanographic Society (CMOS) Congress in Saskatoon, Saskatchewan. (Poster Presentation)


**WORKSHOPS AND COMMUNITY INVOLVEMENT**

NCAR Heliophysics Summer School (July, 2014)
SRI/EISCAT 2014 Incoherent Scatter Radar Summer School and Workshop (July, 2014)
2015 International Reference Ionosphere Workshop (Nov, 2015)
International Center for Theoretical Physics (ICTP) Radio Science School (Mar, 2017)

2016 CMOS Congress: Local Organizing Committee - Program Book Lead
41st COSPAR Scientific Assembly: Session Chair (C4.1: Improved description of the ionosphere through data assimilation – Topside Ionosphere) (August 2016 –
Conference Cancelled
32\textsuperscript{nd} URSI General Assembly & Scientific Symposium: Session Co-Convener
   (International Reference Ionosphere – Improvement, Validation, and Usage) (August 2017)
2\textsuperscript{nd} URSI AT-RASC: Co-Convener (Progress in ionospheric modelling and data assimilation) (May, 2018)
2018 Canadian Association of Physicists (CAP) Congress, Division of Atmospheric and Space Physics (DASP) Head Judge, Student Poster and Paper Competition.

Geophysical Research Letters (GRL) – Reviewer (2014 – Present)
Earth, Planets, and Space (EPS) – Reviewer (2016 – Present)
Radio Science – Reviewer (2016 – Present)
National Science Foundation (NSF) – Grant Proposal Reviewer (2016 – Present)
Atmospheric Measurement Techniques (AMT) – Reviewer (2017 – Present)
Advances in Space Research (ASR) – Reviewer (2017 – Present)

Fencing-Escrime NB, Technical Committee (2010 – 2011)
Fencing-Escrime NB Director (2015 – 2018)
Fencing-Escrime NB Tournament Committee (Chair: 2016 – 2018)

**COMPUTER EXPERIENCE**

Languages: FORTRAN90, IDL, MATLAB, HTML
Software: Microsoft Office, Google Suit, SigmaPlot

**PROFESSIONAL MEMBERSHIPS**

Canadian Association of Physicists (Member: 2010 – Present)
Union Radio-Scientifique Internationale (URSI) (Member: 2011 – Present)
American Geophysical Union (AGU) (Member: 2012 – Present)
Ionoosonde Network Advisory Group (Member: 2010 - Present)
GNSS Research and Applications for the Polar Environment (GRAPE) (Member: 2015- Present)
Canadian Meteorological and Oceanographic Society (Member: 2012-2016)