Verification of Ionospheric Tomography using MIDAS over Grahamstown, South Africa

A thesis submitted in partial fulfilment of the

requirements for the degree of

MASTER IN SCIENCE

of

RHODES UNIVERSITY

by

Zama Thobeka Katamzi

October 10, 2007

Abstract

Global Positioning System (GPS) satellites and receivers are used to derive total electron content (TEC) from the time delay and phase advance of the radio waves as they travels through the ionosphere. TEC is defined as the integral of the electron density along the satellite-receiver signal path. Electron density profiles can be determined from these TEC values using ionospheric tomographic inversion techniques such as Multi-Instrument Data Analysis System (MIDAS). This thesis reports on a study aimed at evaluating the suitability of ionospheric tomography as a tool to derive one-dimensional electron density profiles, using the MIDAS inversion algorithm over Grahamstown, South Africa (33.30° S, 26.50° E). The evaluation was done by using ionosonde data from the Louisvale (28.50° S, 21.20° E) and Madimbo (22.40° S, 30.90° E) stations to create empirical orthonormal functions (EOFs). These EOFs were used by MIDAS in the inversion process to describe the vertical variation of the electron density. Profiles derived from the MIDAS algorithm were compared with profiles obtained from the international Reference Ionosphere (IRI) 2001 model and with ionosonde profiles from the Grahamstown ionosonde station. The optimised MIDAS profiles show a good agreement with the Grahamstown ionosonde profiles. The South African Bottomside Ionospheric Model (SABIM) was used to set the limits within which MIDAS was producing accurate peak electron density (NmF2) values and to define accuracy in this project, with the understanding that the national model (SABIM) is currently the best model for the Grahamstown region. Analysis show that MIDAS produces accurate results during the winter season, which had the lowest root mean square (rms) error of 0.37×10^{11} $[e/m^3]$ and an approximately 86% chance of producing NmF2 closer to the actual NmF2 value than the national model SABIM. MIDAS was found to also produce accurate NmF2 values at 12h00 UT, where an approximately 88% chance of producing an accurate NmF2 value, which may deviate from the measured value by 0.72×10^{11} $[e/m^3]$, was determined. In conclusion, ionospheric tomographic inversion techniques show promise in the reconstruction of electron density profiles over South Africa, and are worth pursuing further in the future.

Contents

List of Figures

List of Tables

Acknowledgements

I would like to say a big thank you to my supervisors Dr Lee-Anne McKinnell and Dr Pierre Cilliers for their help and guidance throughout this project.

My gratitude to the National Research Foundation (NRF) for their financial support through a DoL Scarce Skills scholarship. My gratitude also to the National Astrophysics and Space Science Programme (NASSP) and the Hermanus Magnetic Observatory (HMO) for giving me the opportunity to join the Space Physics group and conduct research at the HMO facility.

MIDAS is owned and licensed by the University of Bath. I would like to acknowledge the MIDAS group for their help during the MIDAS workshop. A special thank you to Dr Cathryn Mitchell and Dr Paul Spencer for their help in making the necessary adjustments to the software to make it more suitable for the aim of this project. The GPS data used in this project was obtained from the Chief Directorate: Surveys and Mapping (CDSM) and so would like to thank them for making the data available. I'd also like to thank Mr Ben Opperman for helping with the GPS data, and Lee-Anne McKinnell for making the ionosonde data available.

I would also like to thank my fellow students who've been comrades throughout the year. Guys, thank you for your help and I've enjoyed the sometimes heated discussions, mostly on politics, conducted in our office. You've certainly made me aware of the many issues facing our continent.

And lastly I would like to thank my parents, family and friends who have always encouraged me to follow my dreams. I would to acknowledge that their

enthusiasm and encouragements have made it easier for me to continue to pursue the academic field. Special thanks to my dad for proof reading my thesis until the early hours of the winter mornings.

Chapter 1

Introduction

This thesis aims to introduce the application of the concept of ionospheric tomography over the South African region, with the emphasis on evaluating MIDAS, an inversion technique developed at the University of Bath in the United Kingdom (UK), over Grahamstown, South Africa (33.30° S, 26.50° E). In this chapter the ionosphere, with emphasis on the information relevant to this project, will be introduced. This includes the definitions of the ionosphere, its variability and how the ionosphere is studied. The last section (section 1.4) gives an overview of what the remaining chapters contain.

1.1 Project Objectives

This project is a feasibility study for improving the ionospheric electron density profiles obtained from Global Positioning System (GPS) data used by the MIDAS algorithm (Mitchell et al., 1997), by using empirical orthogonal functions derived from the ionosonde data, a Chapman function and the International Reference Ionosphere (IRI) model. The project had two main objectives, namely:

- To derive realistic GPS-based ionospheric electron density profiles over the Grahamstown, South Africa station, using the MIDAS algorithm.
- To verify the procedure using an independent instrument, in this case the Grahamstown ionosonde.

1.2 Introduction to the Ionosphere

The GPS signals used to characterise the ionosphere traverse the ionosphere and therefore it is important to understand its behaviour. The ionosphere is that region of the upper atmosphere, from an altitude of about 50 km to about 1 000 km, where plasma is ionised by extreme ultra violet (EUV) radiation and high energy particles from the sun. When the high energy photons from the sun interact with the atoms and molecules in the ionosphere, they strip away the electrons from the parent atoms and molecules, resulting in a number of free negatively charged electrons and positively charged ions.

The ionosphere is divided into three regions: lower ionosphere, the bottomside ionosphere, and the topside ionosphere. The lower ionosphere is the region between 50 km and approximately 90 km consisting of the D layer and some E layer. The bottomside ionosphere ranges from 90 km to about 400 km and consists of the E and F layers. And the uppermost region is the topside ionosphere at 400 km to 1 000 km.

The ionosphere is very variable, displaying diurnal, seasonal, altitude, geographic or geomagnetic location, magnetic and solar activity variations. During the daytime the ionosphere exhibits three main layers:

D-layer : ranges from about 50 km to about 90 km. The ionisation in this layer is due to the ionisation of NO and at high solar activity ionisation of N_2 and O_2. This layer is predominantly responsible for high frequency radio wave absorption, which is due to the electron collision frequency being high, equivalent to the transmitted radio frequencies, at these heights. This phenomenon is discussed in great detail in Rishbeth and Garriott (1969). The D-layer does not play a significant role in ionospheric tomography, because it exhibits much lower electron densities than the F-layer and, therefore, has a negligible contribution to the total electron content (TEC), within the ionosphere.

E-layer : ranges from about 90 km to about 120 km. EUV ionisation in this layer results in molecular oxygen O_2 being formed as the dominant ion species. This layer can only reflect radio waves with frequencies less than 10 MHz.

The E-layer has a negligible effect on GPS signals with frequencies of 1.4 GHz.

F-layer : ranges from an altitude of about 120 km to about 400 km. In this region ionisation is due to EUV solar radiation ionising atomic oxygen O. During the day this layer splits into two layers called the F1 layer and the F2 layer. The F2 layer has the greatest electron density peak, and is the most important layer for long distance high frequency communications. The peak electron density in the F layer contributes the most to the ionospheric TEC.

The variation of the ionospheric electron density with altitude is due to the fact that different molecules are dominant over the range of altitudes. The neutral gas density decreases with height so that there are fewer neutral atoms to participate in the ionisation process at higher altitudes. In contrast, the radiation intensity increases with altitude. The depth to which the radiation can penetrate depends on its wavelength, and different molecules in the ionosphere are ionised more or less strongly by different wavelengths of the radiation. This variation of the ionospheric electron density with altitude is illustrated in figure 1.1.

The ionosphere displays distinct features at some latitudes. Sometimes at mid and high latitudes, there is a dense layer of ionisation in the E region. This layer, called a sporadic E layer, shows no relation to the daytime E layer and is very irregular in behaviour, in that it occurs at random occasions, its virtual height is independent of frequency, and it is partially transparent to waves reflected from higher altitudes. At high altitudes the ionosphere shows very complex behaviour due to its connection to the outer magnetosphere and the interplanetary medium, through the earth's magnetic field. The equatorial ionosphere displays troughs, regions of suppressed electron density, in the ionisation concentration. For a more detailed discussion see Rishbeth and Garriott (1969), pages 160-186.

The diurnal variation of the ionosphere is a result of the rotation of the earth about its axis. Starting around sunrise and as the sun ascends, solar radiation produces free electrons in the ionosphere and the D and E layers form at low altitudes. The F region decreases in height and splits into the F1 and F2 layers. Recombination and other electron-loss processes dominate as the sun descends towards the horizon and the radiation intensity decreases. The D and E layer regions

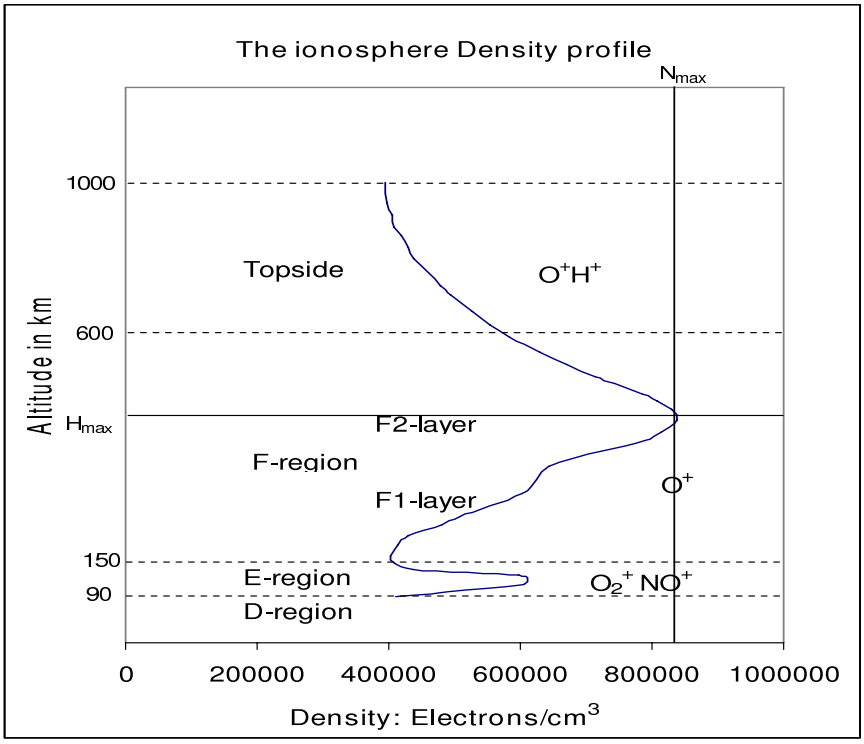

Figure 1.1: A typical electron density profile depicting the different layers of the ionosphere and the dominant ions in each layer. Profile obtained from the Space Plasma Environment And Radio Science (SPEARS) Group website (Honary, 2007)

become insignificant at night and the F1 and F2 layers recombine to form the F layer which rises to higher altitudes while the electron density peak decreases. The F layer remains all night.

The earth revolving around the sun provides an explanation for the seasonal variation of the ionosphere. The D, E and F1 layers experience higher ionisation in summer as the solar angle is highest at that time. The F2 layer, however, does not behave as expected. At mid-latitudes the F2 electron density peak at noon in

winter is higher than electron density peak at the corresponding time in summer. For further reading see Rishbeth and Garriott (1969), page 179.

The solar activity variation is linked to the 11-year solar cycle. When the solar cycle is at its peak, i.e. at sunspot maximum period, the ionisation is greater than at the solar minimum, i.e. at sunspot minimum, because the solar radiation intensity increases thus enhancing the electron concentration in the ionosphere. Magnetic storms are started by solar disturbances, like solar flares, and their recurrence interval is associated with the period of solar rotation. During intense magnetic storms emission of light is observed in the ionosphere at low, mid and high latitudes. At high latitudes the display of emitted light is in the form of an aurora. Further details can be found in Rishbeth and Garriott (1969), pages 252-270.

The GPS and HF radio signals travelling through the ionosphere are subjected to time delay, phase advance, reflection, refraction and rotation of the electric field vector in inverse proportion to the square of their frequency. The ionosphere acts as a refractive medium, to radio signals, with the index of refraction depending on the amount of ionisation. Thus, the electron density profile is very important in determining the state of the ionospheric activities and for the estimation and correction of propagation delays of GPS signals.

1.3 Studying the Ionosphere

It is important to understand the ionosphere as it influences the propagation of radio waves. In particular, imaging the ionospheric electron density distribution is important for estimation and correction of the time delay of radio signals travelling through the ionosphere from earth-orbiting satellites to the receivers on the ground, and for radio communications on the ground. Characterising the ionosphere is also important for ionospheric storm studies, space weather effects and telecommunications. The time-delay experienced by radio waves is proportional to the total electron content (TEC), which is proportional to the plasma refractive index or inversely proportional to the square of the signal frequency. TEC is measured by differencing two frequency signals from satellites, recorded by a dual frequency receiver. The relationship between TEC and the electron density will

be discussed in Chapter 2 of this thesis.

Different instruments are used to observe the ionosphere, but each is restricted to a specific region of the ionosphere. For example, ionosondes can only measure the bottomside region of the ionosphere up to the peak electron density. Global Positioning System (GPS) satellites orbiting at 20 200 km are able to provide information on the entire ionospheric region and plasmasphere. The GPS satellites are used to obtain TEC values with which the electron densities throughout the ionosphere can be determined by a process known as tomographic inversion. There are many ionospheric tomographic procedures. In this project MIDAS (Multi Instrument Data Analysis System) is used to map electron densities throughout the ionosphere over the Grahamstown ionosonde station. In section 4.1 MIDAS will be discussed in greater detail.

1.4 Thesis Overview

In chapter 2 the apparatus and methods used to obtain the TEC values from GPS signals and the electron density values from ionosonde data are described. Included are, a definition of TEC and a brief introduction to the GPS system, as well as the ionosonde.

Chapter 3 presents the theory of tomographic inversion and different methods of solving the tomographic problem. We discuss the Chapman beta layer, the International Reference Ionosphere (IRI) model and the South African Bottomside Ionospheric Model (SABIM) since these models were used to provide the initial guess ionosphere in the inversion procedure.

In chapter 4 an overview of MIDAS is given. It presents the inversion algorithm used and the two methods of solving the inversion problem with emphasis on the Singular Value Decomposition method (SVD). Section 4.2 describes the inversion procedure within MIDAS. This provides the grid vertices used to map the ionosphere. The different methods to create the empirical functions used in the inversion algorithm are also discussed in section 4.2. MIDAS can use data from different instruments. A method of using peak electron density parameters from

the ionosonde data and peak electron density parameters from the SABIM model to supplement the GPS data is discussed.

The results, in the form of one dimensional electron density profiles are presented, in chapter 5. The electron density profiles, presented in section 5.1, were constructed using MIDAS, and the IRI 2001 model, and then compared with the Grahamstown ionosonde profiles. The root mean square error and the co-efficient of correlation are the parameters used to evaluate the performance of MIDAS. The performance of MIDAS is presented and discussed. MIDAS is evaluated for diurnal and seasonal changes, at solar minimum.

Chapter 6 presents an overall discussion and conclusions.

Chapter 2

Instrumentation and Measurements

2.1 Global Positioning System

The Global Positioning System (GPS) is a part of the Global Navigation Satellite System (GNSS) developed by the United States Department of Defence for instantaneous determination of position and velocity. This system is open for use to the general public and has proven to be very useful in navigation, scientific studies and land surveying.

The GPS system consists of three main components: the space segment, the control segment and the user segment. The space segment is actually a constellation of about 27 satellites orbiting at an altitude of approximately 20 200 km above the Earth. Of the 27 satellites, only 24 satellites are operational while the other 3 are spares to replace the operational satellites should they stop functioning. Six groups of 4 operational satellites occupy six evenly spaced circular orbital planes centred on Earth. The orbits are inclined at 55° to the equatorial plane and separated by 60° right ascension of the ascending node. The orbits have a radius of about 26 600 km, resulting in an orbital period of approximately 12 sidereal hours. Thus each satellite makes two complete orbits each sidereal day, and it passes over the same location once each day. This orbital arrangement ensures that at least 4 satellites are always within a line of sight from almost any place on Earth. The GPS satellites broadcast signals continuously on two separate frequencies of 1.58

GHz and 1.23 GHz, known as L1 and L2 respectively. For further reading see Dana (1994).

The satellites are continuously monitored by the ground-based control system, which manages the tracking, telemetry and control functions including maintainance of the station, monitoring the system's well-being, updating the ephemeris and almanacs, and most importantly, maintaining the synchronism of the atomic clocks on board the satellites to GPS system time.

The user segment of the GPS system are the GPS receivers and the user community that receive the GPS satellite broadcasts and compute their precise position, velocity and time. There are two kinds of receivers: the single frequency and the dual frequency receivers. The single frequency receiver can only receive signals on L1 frequency and cannot eliminate the effect of ionospheric delay. The dual frequency receiver can receive both the L1 and L2 frequency signals simultaneously and can eliminate ionospheric effects, thus making it more accurate than the single frequency channel GPS receiver. The military dual frequency receivers are equipped with a classified algorithm that can retrieve the encrypted Precise Positioning Services (PPS), primarily intended for authorised armed services and government agencies, while the civil dual frequency receivers do not have this option available to them and thus use codeless techniques to retrieve the PPS. The single frequency GPS receivers can only retrieve the Standard Positioning Services (SPS), and so are not capable of reaching the precision of dual frequency receivers.

2.1.1 GPS Satellite Signal Process

GPS satellites are equipped with atomic clocks with a clock stability of at least 10^{-13} s. From the base frequency of the atomic clocks (10.23 MHz) all the other frequencies that are required for the GPS satellite are derived. The navigation message, which describes parameters like the GPS satellite orbits and clock corrections, modulates both carrier signals (i.e. L1 and L2) at a data rate of 50 Hz. The resulting signals are further modulated using two different pseudo-random noise (PRN) codes, known as the Coarse Acquisition (C/A) code and the Precise (P) code. The C/A code is transmitted with a frequency of 1.02 MHz, repeating every millisecond. This code, found in all civil GPS receivers, modulates only the

L1 frequency and the SPS is delivered through it. Each satellite has a unique C/A PRN code and thus they are identified by their PRN number. The P-code is a 10.23 MHz PRN code that repeats every week. It modulates the L1 as well as the L2 carrier phases. The principle of this process is illustrated in figure 2.1.

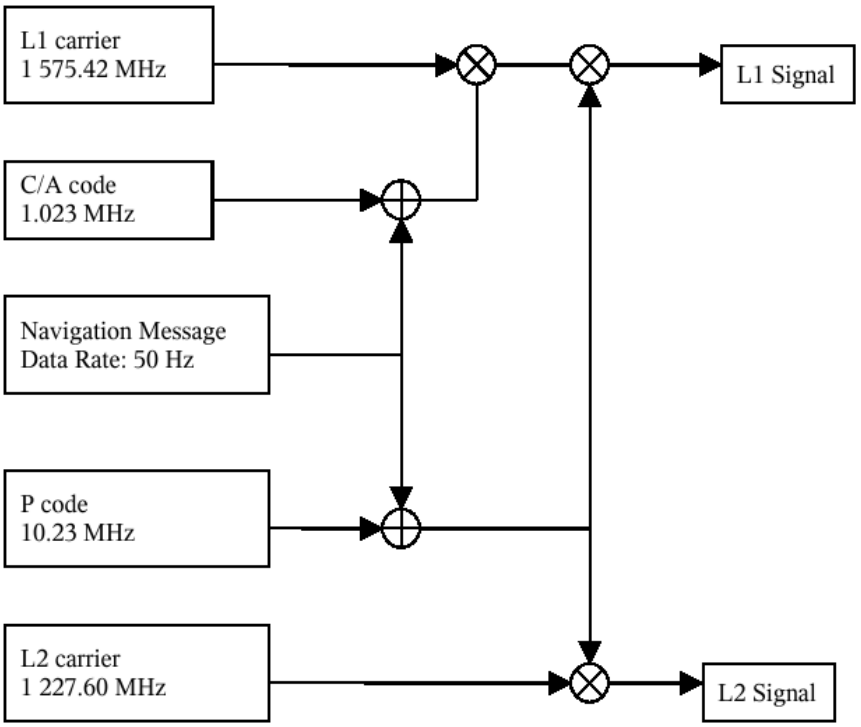

Figure 2.1: The basic principle of signal processing. Reconstructed from on a similar figure in Dana (1994)

2.1.2 GPS TEC Measurements

Total Electron Content (TEC) is defined as the total number of electrons in a column of unit cross-section from the transmitter, in space, to the receiver, on the ground. The mathematical representation of the definition is:

$$I = \int_R^S N_e(r, \theta, \phi, t) ds \tag{2.1}$$

where I stands for TEC, N_e is the electron density, r is the radial distance from the centre of the Earth, θ, ϕ are the longitude, and latitude respectively, t is the

time, and s is the satellite-receiver path length. TEC is usually measured in TEC units (TECU) of 10^{16} electrons per square meter.

Radio waves propagating through the ionosphere experience a group delay and a phase advance, because the ionosphere is a dispersive medium. A change in the refractive index is a source of inaccuracy in GPS signals as shown by the following derivations, from equation (2.2) to equation (2.9). Propagation of the signal can be described by the Appleton-Hartree equation which, when neglecting collisions, is given by:

$$\mu^2 = 1 - \frac{X(1-X)}{(1-X) - \frac{1}{2}Y_T^2 \pm [\frac{1}{4}Y_T^4 + (1-X)^2 Y_L^2]^{\frac{1}{2}}} \tag{2.2}$$

where

$$X = \frac{\omega_{pe}^2}{\omega}$$
$$Y = \frac{\omega_{ce}^2}{\omega}$$
$$Y_L = Y cos\theta$$
$$Y_T = Y sin\theta$$
$$\omega_{pe}^2 = [2\pi f_e]^2$$
$$\omega_{ce}^2 = [2\pi f]^2 = \frac{Be}{m_e}$$

μ is the refractive index; ω_{pe} is the angular electron plasma frequency; ω is the angular frequency of the wave; ω_{ce} is the angular electron gyrofrequency; f_e is the electron plasma frequency; f is the radio frequency; Y is the component of the Earth's magnetic field and θ is the angle between the field and the direction of propagation of the signal. The plus sign refers to the ordinary wave and the minus refers to the extra-ordinary wave.

Assuming that the magnetic field is negligible, i.e. $Y = Y_L = Y_T = 0$, equation (2.2) becomes

$$\mu^2 = 1 - \frac{X(1-X)}{(1-X)} = 1 - X. \tag{2.3}$$

Solving for μ by expansion and ignoring the high order terms we get:

$$\mu = 1 - \frac{1}{2}X \tag{2.4}$$

where

$$X = \frac{\omega_{pe}^2}{\omega^2} = \frac{(2\pi f_e)^2}{(2\pi f)^2} = \frac{40.31}{f^2} n_e$$

and

$$f_e = \frac{n_e e^2}{4\pi^2 m_e \varepsilon_0}$$

where n_e is the electron density; e is the electron charge (1.60×10^{-19} C); m_e is the electron mass (9.11×10^{-31} kg), and ε_0 is the vacuum permittivity (8.85×10^{-12} F/m). Finally,

$$\mu = 1 - \frac{40.31}{f^2} n_e. \tag{2.5}$$

This dispersion relation can now be written in terms of phase and group refractive indices μ_{ph} and μ_{gr} respectively:

$$\mu_{gr} = 1 + \frac{40.31}{f^2} n_e$$
$$\mu_{ph} = 1 - \frac{40.31}{f^2} n_e.$$

The geometric range R from receiver r to satellite s along the line of sight, adopted from Fermat's principle (see Hofmann-Wellenhof et al. (1997) page 102), is given by

$$R = \int_r^s 1 d\ell,$$

and the pseudo-range ρ is given by

$$\rho = \int_r^s \mu d\rho.$$

Then the path length difference is:

$$\Delta\rho = \rho - R = \int_r^s \mu d\rho - \int_r^s 1 d\ell.$$

As the integration is along the geometric path, $d\rho$ can be changed to $d\ell$, and then the group refractive index can be substituted for the group delay:

$$\Delta\rho_{gr} = \frac{40.31}{f^2}\int_r^s n_e d\ell. \tag{2.6}$$

Similarly for the phase advance, the path difference is:

$$\Delta\rho_{ph} = -\frac{40.31}{f^2}\int_r^s n_e d\ell. \tag{2.7}$$

From the definition of TEC in equation (3.2), the above equations can be written as:

$$\Delta\rho_{gr} = \frac{40.31}{f^2} I \tag{2.8}$$

$$\Delta\rho_{ph} = -\frac{40.31}{f^2} I. \tag{2.9}$$

But the geometric range cannot be measured directly, and therefore TEC cannot be calculated from equations (2.8) or (2.9). Thus a method of measuring TEC directly from the differential code delay or carrier phase measurements on both the L1 and L2 frequencies is used. This algorithm is presented in a logical mathematical form in Meggs (2005).

TEC is observed from GPS measurements by a linear combination of the carrier phases of the signal and the pseudo-ranges recorded by a receiver on both carrier frequencies L1 and L2. The carrier phase-based measurements as well as the pseudo-range measurements are as follows:

$$\begin{aligned} L_1 &= \rho + \Delta\rho_{ph,1} + \lambda_1 n_1 + \varepsilon_1^r + \varepsilon_1^s \\ L_2 &= \rho + \Delta\rho_{ph,2} + \lambda_2 n_2 + \varepsilon_2^r + \varepsilon_2^s \\ P_1 &= \rho + \Delta\rho_{gr,1} + \tau_1^r + \tau_1^s \\ P_2 &= \rho + \Delta\rho_{gr,2} + \tau_2^r + \tau_2^s \end{aligned}$$

where ρ is a non-dispersive delay term which lumps together the geometric distance, troposphere delay, clock errors, and non-dispersive delays in the hardware (see Mannucci et al. (1999)); $\lambda_1 n_1$ and $\lambda_2 n_2$ are the unknown integer cycle am-

biguity associated with the carrier phase term. The ε and τ are the dispersive component of the receiver and satellite hardware delays. The errors mentioned here will be discussed in section 2.1.4 on page 16. The P_1 and P_2 are extracted from the P-code.

The phase-based TEC is calculated from the difference between the two observed phase advances:

$$L_I = L_1 - L_2 = 40.31[\frac{1}{f_2^2} + \frac{1}{f_1^2}]I_L + (\lambda_1 n_1 - \lambda_2 n_2) + b'_r + b'_s. \tag{2.10}$$

Similarly the code-based TEC is obtained from:

$$P_I = P_2 - P_1 = 40.31[\frac{1}{f_2^2} + \frac{1}{f_1^2}]I_P + b_r + b_s. \tag{2.11}$$

Having obtained the TEC values it is possible to apply the technique known as ionospheric tomography to derive the electron density values. This technique will be discussed in Chapter 3.

2.1.3 Position Determination

The GPS travel time is the amount of time taken by a radio signal to travel from the satellite to a receiver. So, if the exact time the signal left the satellite is known and the exact time it reached the receiver is known, then the travel time can be calculated. The distance between the receiver and satellite can be determined from the travel time, assuming the signal traveled at the speed of light. The GPS receiver then calculates its position by measuring these path lengths from at least 4 satellites.

Each GPS satellite sends a navigation message to the receiver containing its current position, clock corrections and other system parameters. The GPS receiver identifies each satellite by its unique PRN code and internally produces an identical code to the satellite. By auto-correlating the two sequences, the receiver can determine the time delay of the signal and hence the pseudo-range, i.e. the path length.

Observations from three satellites can be used to determine the receiver's accurate position, provided that the GPS system time, satellite and receiver clocks are precisely and continually synchronised. This is not possible as the receivers use quartz clocks which are not as accurate as the GPS atomic clocks. The satellite clock also drifts from the GPS system time, but this is monitored and corrected, thus effectively synchronising the clock with GPS system time. As the same procedure cannot be applied to the receiver clock, a fourth satellite observation is necessary.

To solve for the receiver's position, the Earth-centred Earth-fixed (ECEF) coordinate system is used to define the satellite and receiver positions (see Meggs (2005)). If $\mathbf{s}$ is the satellite's position vector, and $\mathbf{r}$ is the receiver's position vector, then the geometric range $\mathbf{R}$ is given by

$$\mathbf{R} = \mathbf{s} - \mathbf{r},$$

such that the the pseudo-range is

$$\rho = \| \mathbf{s} - \mathbf{r} \| = \sqrt{(x_s - x_r)^2 + (y_s - y_r)^2 + (z_s - z_r)^2}. \tag{2.12}$$

The pseudo-range is measured by differencing the time the signal left the satellite, t_s, and the time recorded by the receiver when it received the signal, t_r:

$$\rho = c(t_r - t_s).$$

Taking into account the receiver's clock inaccuracy, δt_r, this can be written as:

$$\rho = c(t_r + \delta t_r - t_s) = c(t_r - t_s) + c\delta t_r.$$

Thus, from equation (2.12) and the above equation, the equation for the pseudo-range becomes:

$$\rho = \sqrt{(x_s - x_r)^2 + (y_s - y_r)^2 + (z_s - z_r)^2} + c\delta t_r. \tag{2.13}$$

Since observations from four satellites are required for accuracy, the above equation can be represented by:

$$\rho_i = \sqrt{(x_{si} - x_r)^2 + (y_{si} - y_r)^2 + (z_{si} - z_r)^2} + c\delta t_r \tag{2.14}$$

where i is the index of the satellites. The method of obtaining the solution, known as the navigation solution, is discussed in Hofmann-Wellenhof et al. (1997).

2.1.4 GPS Errors and Accuracy

Although the information obtained from a GPS satellite was designed to be very accurate, it does contain errors from several sources. The position accuracy of the GPS system is measured by two parameters known as the User Equivalent Range Error (UERE), and the Geometric Dilution of Precision (GDoP). The sources of error includes ionospheric effects, ephemeris errors, satellite clock errors, multipath effects, relativistic effects and numerical errors, and are collectively known as UERE. General estimated contribution values for each effect, in metres, is given in Table 2.1, and discussed in detail in NAV (1996).

The major source of inaccuracy is the change in refractive index along the ray

Table 2.1: GPS Range Error Budget.

Source	**P code Error [metres]**	**C/A code Error [metres]**
Ionosphere	4.5	9.8 - 19.6
Troposphere	3.9	3.9
Ephemeris	8.2	8.2
Multipath	2.4	2.4
Others	1.0	1.0

path that changes the speed of the GPS signals as they pass through the ionosphere and troposphere. The atmospheric effect is minimal when the satellite is directly overhead and increases greatly for satellites near the horizon, as the signal has to travel a longer path through the atmosphere.

Radio waves are slowed down at a rate inversely proportional to the square of their frequency when they are passing through the ionosphere. In other words,

ionospheric delay affects the speed of radio waves based on their frequency, a phenomenon known as dispersion. Thus both the L1 and L2 frequencies can be used to reduce this effect by comparing the delays in the frequencies to measure the dispersion, as previously discussed in section 2.1.2.

The troposphere effect causes further variable delay by refraction due to humidity. The error caused by this effect changes faster, thus making it difficult to eliminate, but this effect is very much smaller than the ionosphere effect and hence, for the purposes of this project, the troposphere effect is ignored.

The atomic clocks on board satellites are very accurate, but there are slight variations from satellite to satellite. At GPS orbital speeds the special relativity theory predicts that the atomic clocks will run slower than the stationary ground clocks. On the other hand, the theory of general relativity predicts that the atomic clocks on board the satellite will tick faster than the clocks on the Earth's surface as the satellites' clocks are in a weaker gravitational field. It turns out that the time contraction is greater than the time dilation, thus the GPS atomic clocks run faster and this discrepancy has to be taken into consideration.

The multipath effect is caused by reflection of the GPS signal on objects such as buildings before reaching the receiver, see (Larson et al. (2007)). Other sources of errors are electromagnetic radiation and man-made interferences that can hinder the receiver from receiving signals. All the non-dispersive errors are taken into account by taking the differenced phase and code ranges in equations (2.10) and (2.11). The unknown integer cycle ambiguity, $\lambda_1 n_1 - \lambda_2 n_2$, in equation (2.10) is removed by calibrating the L_I curve to the P_I curve using least square fitting, (see Meggs (2005) page 30).

The alignment or geometry of the satellite constellation from which signals are received is a major factor in position accuracy, because it affects the other errors as well. The quality of the constellation geometry is indicated by the dilution of precision (DoP) values. A good distribution of the satellites across the sky (i.e. large angle between satellites) lowers the DoP, thus providing accurate measurements, while a higher DoP provides poor measurements.

A technique known as Differential GPS (DGPS) allows for improvement in position accuracy by using a second stationary GPS receiver. Since the stationary receiver's position, a previously surveyed benchmark, is known accurately, the correction can be computed by differencing its known and measured positions. The correction factor is then transmitted to other GPS receivers on a separate radio frequency and applied accordingly to the measured data. The disadvantage of this method is that it is restricted to nearby receivers, as the positioning accuracy decreases as the distance between the stationary receiver and other receivers increases.

An enhancement to DGPS used in the United States of America (USA), the Wide Area Augmentation System (WAAS), offers a better accuracy than DGPS. This system consists of a network of ground stations that continuously monitors the GPS satellites, and compares the actual ground position with the position calculated by the GPS satellites. The varying correction factors are then continuously calculated and transmitted to the receivers. Similar systems are under development in Europe (EGNOS - European Geostationary Navigation Overlay Services), and Japan and other Asian countries (MSAS - Multi-functional Satellite Augmentation System). These systems will facilitate a more accurate calculation of the TEC values.

2.2 Ionosonde

An ionosonde is a high frequency, vertically sounding radar used to measure the bottomside ionosphere. It transmits radio waves, with frequencies ranging from 3 to 30 MHz, vertically upwards, which are reflected by the ionosphere. Different frequencies are reflected by different plasma density regions. There are two types of ionosondes: the pulse ionosonde and the chirp ionosonde. The pulse ionosonde transmits short pulses at high peak power, while the chirp ionosonde transmits long pulses at low peak power. Kelso (1964) give a more detailed discussion on ionospheric measurements using ionosondes. The South African ionospheric group uses a digisonde portable sounder (DPS) which is a pulse ionosonde with real-time analysis capabilities, for more information see Galkin (2007). The output of any

ionosonde is an ionogram, a plot of frequency versus virtual height and reveals the heights of reflecting layers and the maximum frequency associated with them, as shown by the example given in figure 2.2.

The ionosonde then measures the time taken for the transmitted signal to re-

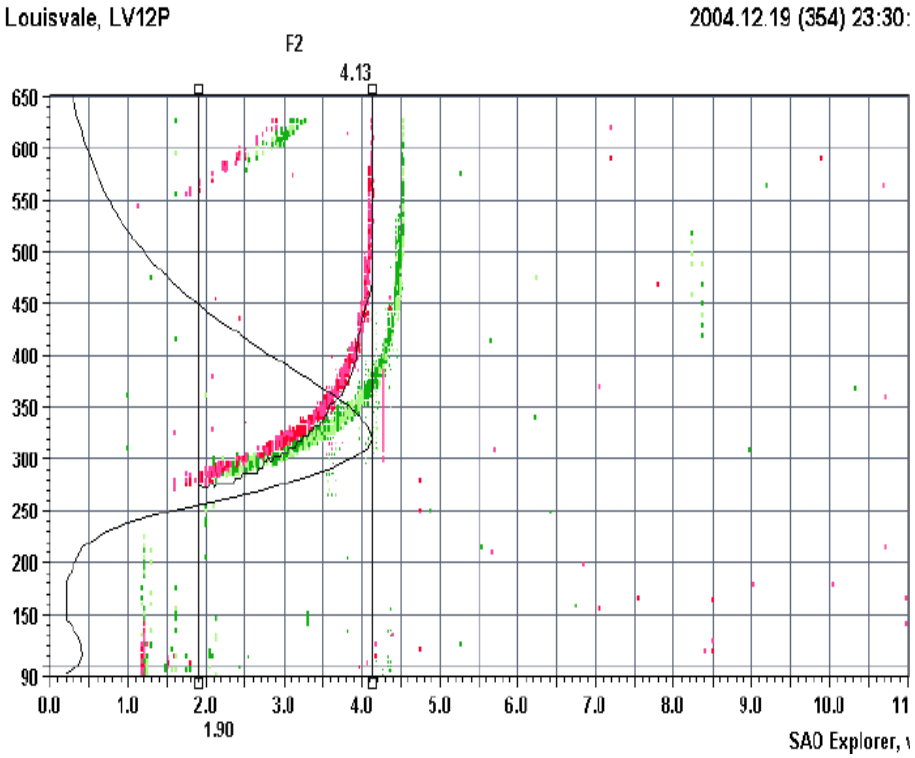

Figure 2.2: An ionogram with a fitted frequency profile was extracted from SAO explorer. The ionogram was obtained from the Louisvale digisonde station (28.50° S, 21.20° E). The vertical axis is the altitude in km, and the horizontal axis is the frequency in MHz.

turn. The height at which the signal would have been reflected had it continued to travel at the speed of light through the ionosphere, known as the virtual height, is obtained from the equation

$$h' = \frac{ct}{2} \tag{2.15}$$

where t is the echo delay and c is the speed of light (see Davies (1990), page 95).

Electron densities at different heights can then be calculated from:

$$n_e = \frac{4\pi^2 m_e \varepsilon_0}{e^2} f^2 = 1.24 \times 10^{10} f^2 \tag{2.16}$$

where f, the radio frequency, is measured in MHz, and n_e in e/m^3. The electron density profile (see figure 1.1), is constructed by plotting the electron density at different actual reflecting heights.

There are two main disadvantages to using the ionosonde to map out the ionosphere: they can only observe the ionosphere up to the height associated with the peak frequency of the F2 region, and they can only observe the local ionosphere directly above their geographic location. However, the ionosonde provides the best 'true' measurement of the ionosphere below the F2 region peak, and therefore, will be used as the comparison tool in this thesis.

2.3 Conclusion

In summary, this chapter introduced the measurements obtained from the GPS system, with emphasis on the TEC value and its use, i.e. determination of the receiver's position. The following chapter will discuss how to use these TEC values to obtain the electron density values, a subject known as ionospheric tomography. Once the electron density values are obtained in a 3-D volume over the site of interest, the electron density profiles can be constructed and these are compared to the ionosonde measured profiles, to evaluate the accuracy of MIDAS.

Chapter 3

Ionospheric Tomography

Ionospheric tomography is the reconstruction of the vertical electron density distribution of the ionosphere from the total electron content (TEC) determined along the ray paths of radio waves travelling through the ionosphere. In other words, ionospheric tomography is the imaging of the ionosphere based on remote sensing or sounding of the ionosphere using the radio waves that are influenced by the ionosphere, (Leitinger, 1999).

The principle of tomography as an imaging procedure has been used in many different disciplines of science and technology, for example, computerised tomography, in medicine, uses X-ray absorption to reconstruct the internal structure of a human body (Donnan et al., 1982), and seismic tomography uses propagation times of seismic waves to reconstruct the earth's interior structure (Bazin et al., 1998).

The aim of this chapter is to present the theory of ionospheric tomography. The different ionospheric methods developed for reconstructing the two-dimensional and three-dimensional structure of the ionosphere, in the form of electron density as a function of three spatial variables, will also be briefly discussed. The discussion in this chapter is based on the work of Leitinger (1999).

3.1 Tomographic Theory

The use of satellite signal measurements in tomographic problems was first suggested by Austen et al. (1988), where the electron density distribution was reconstructed from the total electron content (TEC) along a set of rays. Further research into the use of TEC in different methods of ionospheric tomography has been done by many authors, such as Vasicek and Kronschnabl (1995), Spencer et al. (1998), Hajj and Romans (1998), and Stolle et al. (2003). It is important that as many rays as possible intersect the region of interest, over as wide a range of angles as possible. This can be accomplished by using as many receivers as possible and having a high data rate (Fremouw et al., 1992). For this project 8 to 10 receivers were used, and the data was sampled every 30 seconds. The geometry of the satellite-to-receiver rays is illustrated in figure 3.1.

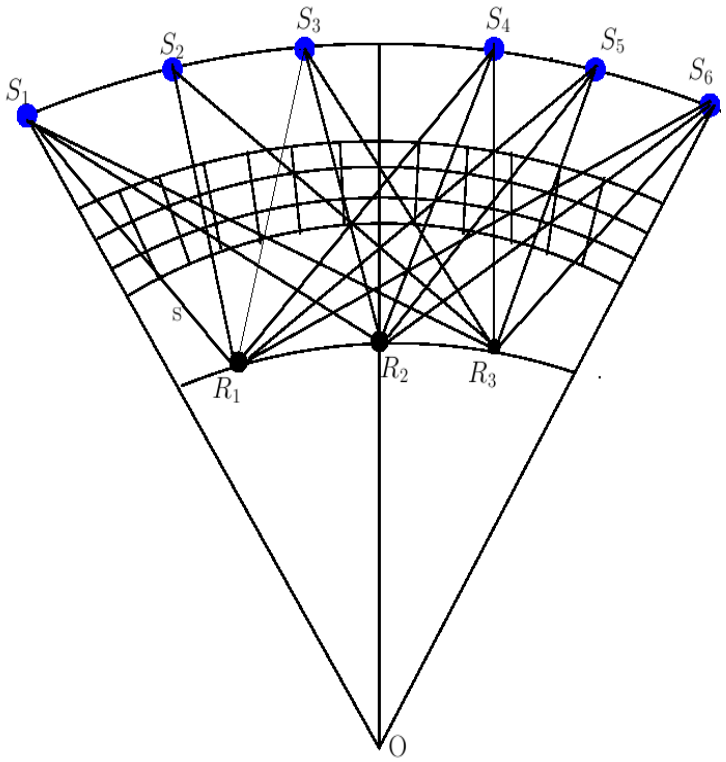

Figure 3.1: The signal rays as they travel from the GPS satellite to the receivers on the ground. This figure also illustrates the division of the ionospheric region into pixels. Adopted from Kunitsyn and Tereshchenko (2003).

The following simple example illustrates the principle of the tomographic problem (Meggs, 2005). For a simple case, consider the grid shown in figure 3.2 consisting of four unknown quantities, each in a pixel with known size. The measured parameter is the sum of the path lengths through each pixel multiplied by the unknown quantity in the pixel. The problem of tomography is to determine the unknown quantities in all the pixels from the values of the measured parameter. So if i is the indices of the projection, and b_i is the ith projection, j is the pixel number and x_j is the unknown quantity in the jth pixel. Then the length of each projection in each pixel is represented by A_{ij}. So for the grid in figure 3.2, the following system of equations can be formed:

$$\begin{aligned} b_1 &= A_{11}x_1 + A_{13}x_3 \\ b_2 &= A_{12}x_2 + A_{12}x_2 + A_{21}x_1 \\ b_3 &= A_{12}x_2 + A_{21}x_1 \\ b_4 &= A_{11}x_1 + A_{12}x_2. \end{aligned}$$

Collectively,

$$\mathbf{b} = \mathbf{A}\mathbf{x} \tag{3.1}$$

which is used to solve for $\mathbf{x}$ using a matrix inversion method.

The influence of the ionosphere on the radio waves properties is represented by:

$$I_{\backslash} = \int_R^S N_e(r, \theta, \phi, t) d\mathrm{s} \tag{3.2}$$

where $I_{\backslash}$ is the measured slant TEC, N_e is the electron density, r is the radial distance from the centre of the Earth, θ, ϕ are the longitude and latitude respectively, t is the time, ds is the satellite-receiver path length element, and the integration is from the receiver R to the satellite S.

For the sake of simplicity, the problem discussed here is two-dimensional, but it can be adapted for a three-dimensional problem by introducing a longitude or latitude dependence of the ionospheric electron density, and dividing the iono-

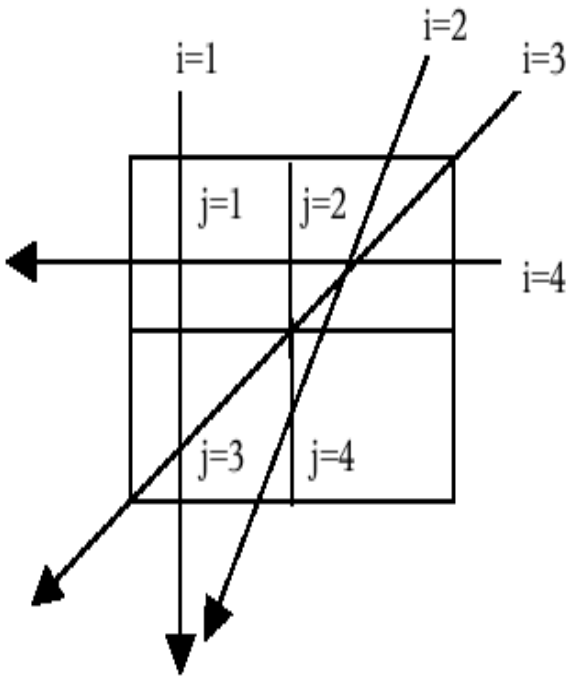

Figure 3.2: A grid of pixels consisting of unknown quantities with rays traversing each pixel. This illustration was taken from Meggs (2005).

spheric region into a grid of three-dimensional volumes. Let the ionospheric region of interest be divided into a grid of two-dimensional boxes bounded in height and angular spacing, as illustrated by figure 3.1. Let i be the index of the rays, I_i represent the total electron content of the i-th ray, j be the index of the pixels and N_j be the unknown electron density of the ionosphere. The length of each ray in each pixel is represented by s_{ij}, and each ray gives a row in a linear-equation system:

$$I_i = \sum_i N_j \mathrm{s}_{ij}. \tag{3.3}$$

The above equation can be in a matrix form of the form of equation (3.1), where I_i will be equivalent to i measurements of $\mathbf{b}$, N_j will be equivalent to j unknown electron densities $\mathbf{x}$, and s_{ij} will be equivalent to the $i \times j$ path elements $\mathbf{A}$.

The body of the Earth is not transparent to radio waves, thus limiting the propagation paths and so the ray paths can only be measured when the satellites are

above the Earth's surface. There are also no receivers perpendicular to the surface of the Earth, and so the ionosphere cannot be scanned in the vertical direction. Therefore equation (3.1), cannot be solved directly as the matrix $\mathbf{A}$ is highly singular. 'Feasible' solutions are gained by using prior information to stabilise the solution to equation (3.1).

There are a number of reconstruction methods, roughly grouped into iterative and non-iterative methods, used to solve for the electron density using the measured slant TEC values. The most famous iterative technique used to solve equation (3.1) is known as ART (algebraic reconstruction technique). There are many derivatives of this technique, for example, MART (multiplicative algebraic reconstruction technique), and SIRT (simultaneous iterative reconstruction technique). Leitinger (1999) states that the ART algorithm uses the current estimate for the solution ($\mathbf{x}^k$ for iteration k) to compute the difference between the measured $\mathbf{b}$ and $\mathbf{b}^k = \boldsymbol{A}\mathbf{x}^k$, and a correction derived from this difference is distributed over $\mathbf{x}^k$ to obtain the next iteration result, $\mathbf{x}^{k+1}$.

Non-iterative techniques include the singular value decomposition (SVD) and modified truncated SVD methods discussed by Raymund et al. (1994) and Mitchell and Spencer (2003). One other technique, described by Fremouw et al. (1992), involves using a multiplication operator to combine a vertical model and a horizontal model. The vertical model is obtained from empirical orthonormal functions (EOFs), which are based on ionospheric models, and the horizontal model is obtained from sine and cosine functions. For further reading, consult Leitinger (1999) and Kunitsyn and Tereshchenko (2003), page 57. The method used for this project is a non-iterative method, adapted from Fremouw et al. (1992) work, which uses the empirical orthonormal functions (EOFs) derived from ionospheric model or ionosonde profiles to represent the vertical structure and the horizontal structure by spherical harmonics.

As there are many methods of solving the tomographic problem it is important to choose the proper method of solving the system of linear equations. The chosen method should be able to describe all the conditions of the ionosphere at all times.

3.2 Modelling the Ionosphere

This section describes modelling the ionosphere for tomographic purposes. Both the iterative and non-iterative algorithms need prior information about the ionosphere to derive the most plausible solution to the tomographic problem expressed by equation (3.1). This prior information comes in the form of ionospheric models. There are many models developed to describe the ionosphere but only the following three will be discussed namely: Chapman profile, the International Reference Ionosphere (IRI) model, and the South African Bottomside Ionospheric Model (SABIM). The Chapman functions and the IRI model are used in the MIDAS algorithm to model the ionosphere. Ionospheric parameters from the model SABIM were used to supplement the GPS data used as inputs in the MIDAS system. These parameters are the peak electron densities and corresponding peak heights of the E, F1 and F2 regions and the propagation factor M(3000)F2.

3.2.1 Chapman Profile

The Chapman profile has two different formulations: the Chapman alpha layer and the Chapman beta layer (Stankov et al., 2003), depending on the assumptions related to the electron recombination theory. The general form of the Chapman profile describing the electron density profile is:

$$N_e = N_{e_0} exp(c[1 - z - sec\chi exp(-z)]) \tag{3.4}$$

where χ is the solar zenith angle, N_e is the electron density and c is the type of coefficient that describes the layer: $c = 1/2$ for Chapman alpha and $c = 1$ for Chapman beta (Stankov et al., 2003). Here z is the reduced height given by,

$$z = \frac{h - h_m}{H}$$

where h is the height, h_m is the peak density height, and H is the scale height. Typical values of the scale height lie in the range 25-50 km according to Stankov et al. (2003).

The Chapman alpha profile is given by the equation:

$$N_e = N_{e_0} exp(\frac{1}{2}[1 - z - sec\chi exp(-z)]) \tag{3.5}$$

and the Chapman beta profile is given by the equation:

$$N_e = N_{e_0} exp[1 - z - sec\chi exp(-z)]. \tag{3.6}$$

Stankov et al. (2003) states that for the Chapman alpha layer, the assumption is that the process dominant in the ionosphere is dissociative recombination, where the electrons recombine directly with positive ions and that no positive ions are present. For the Chapman beta layer, the dominant processes are the charge transfer or atom-ion exchange reactions. For a more detailed explanation see Davies (1990), pages 60 to 65. MIDAS uses the Chapman beta function to create EOFs. The Chapman profile is used to predict the maximum electron densities of the E and F1 layers and does not predict a realistic maximum electron density of the F2 region, (see McNamara (1991) page 24). Examples of Chapman profiles for two solar zenith angles 0° and 80° are given in figure 3.3.

3.2.2 The International Reference Ionosphere (IRI) Model

The IRI model is an international model used for the specification of ionospheric parameters. It was first developed in the 1960's and is continuously being improved by a joint working group of the International Union of Radio Science (URSI) and the Committee on Space Research (COSPAR).

The IRI as an empirical model is based on experimental evidence using existing data records from both the ground and space data resources. It takes longitude, latitude, year, day, hour, and magnetic index as input parameters and predicts the electron density and temperature, ion density and temperature, total electron content and other ionospheric parameters. These are monthly averages for magnetically quiet conditions. For a more detail description of the IRI, see Bilitza (1990) and Bilitza (2001). An electron density profile obtained from IRI is shown in figure 3.4.

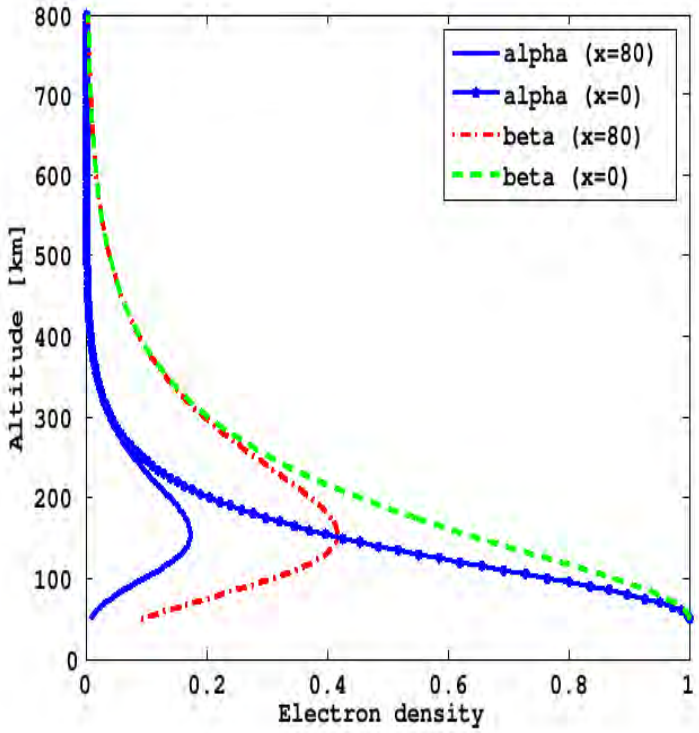

Figure 3.3: Chapman alpha and beta functions for solar zenith angles of 0° and 80°, and using a scale height of 80 km.

There are many versions of the IRI, as the working group meet annually to discuss and improve the model. The version of the model used for this project is the MATLAB (Matrix Laboratory) version of IRI 2001 model, which inputs altitude, latitude, longitude, local or universal time, day, month and year to determine the electron densities. The IRI 2001 model was used because it was the latest model available. The software package of the model is provided by the National Space Science Data Center (NSSDC). It does not give a good description of the ionosphere in the Southern African region as no South African data was used in its data sources.

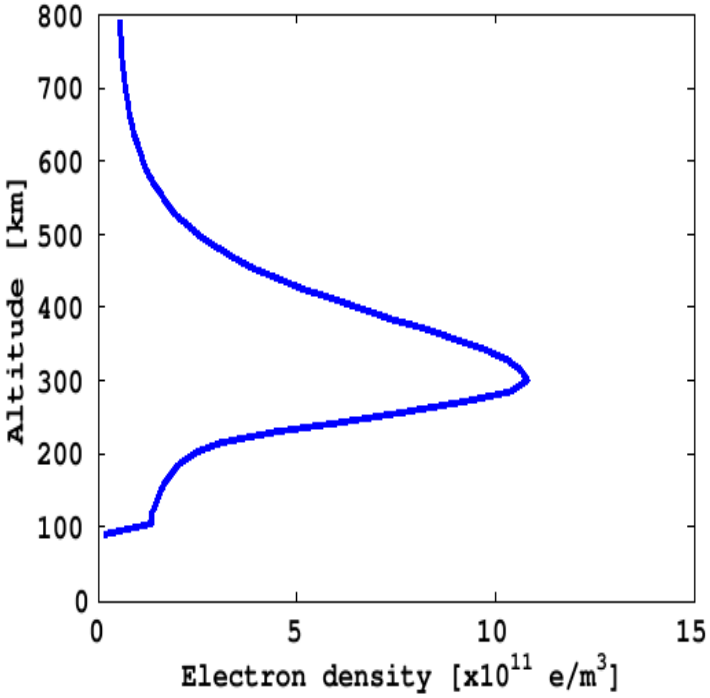

Figure 3.4: The IRI profile over Grahamstown on 12 March 2004 at 12h00 UT.

3.2.3 The South African Bottomside Ionospheric Model (SABIM)

This model was developed by McKinnell (2002), using the method of neural networks to predict the bottomside electron density profiles. An example is shown in figure 3.5. SABIM was first developed using the Grahamstown station data, but now also incorporates the Louisvale and Madimbo data, thus giving the best description of the ionospheric conditions of the bottomside region over South Africa. Required input data for the models are year, day number, longitude and latitude. The parameters predicted by this model and used in this project were the maximum frequencies and corresponding heights of the E, F1 and F2 regions, and the propagation factor M(3000)F2.

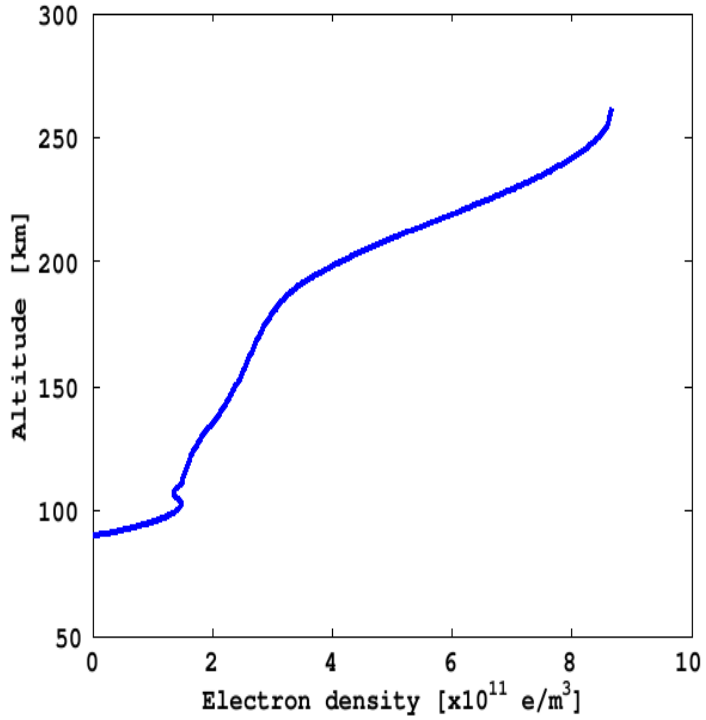

Figure 3.5: The bottomside profile over Grahamstown on 12 March 2004 at 12h00 UT, as predicted by the SABIM model.

3.3 Conclusion

To summarise, the tomographic problem expressed as equation (3.2) is to derive electron density distributions from GPS measurements. This problem, expressed in a system of linear equations as given by equation (3.1), requires prior information from ionospheric models to produce a realistic solution.

In the next chapter a non-iterative method or algorithm of inverting equation (3.1) to solve for the electron densities, which is based on the method introduced by Fremouw et al. (1992), is described. This algorithm, called MIDAS, uses GPS data to derive the electron density distribution in four dimensions, these dimensions being altitude, longitude, latitude and time. MIDAS also gives the option to supplement the GPS data with data from other instruments. For the purposes of this project, only one dimensional electron density profiles were derived from MIDAS, and the dimension used is the altitude.

Chapter 4

MIDAS

Various researchers (Fridman et al. (2006), Jin et al. (2006), Cilliers et al. (2004), Reilly and Singh (2004), Stankov et al. (2003), Mitchell et al. (1997)) have used different inversion techniques to construct the electron density profiles from GPS data networks, but all these investigations have been done in the northern hemisphere with the exception of Cilliers et al. (2004). The Multi-Instrument Data Analysis System (MIDAS) is another such technique that uses data from different instruments, e.g. GPS and ionosonde, to construct three-dimensional and time-dependent electron density maps from TEC measurements. This software was developed by researchers at the University of Bath (UK) and is used with their permission in this project. MIDAS will be applied to the determination of the electron density profile over Grahamstown, South Africa (33.3° S, 26.5° E). In other words, MIDAS will be used to reproduce the electron density profiles obtained from the Grahamstown station ionosonde, to test if the profiles obtained from MIDAS can be used to supplement the ionosonde profiles. A detailed technical report on MIDAS is given by Paul Spencer in Spencer (2000), Spencer (2001), and Spencer (2002). Mitchell and Spencer (2003) explain the algorithm of MIDAS and a brief summary appears in section 4.1.

4.1 The MIDAS Algorithm

The ionospheric region of interest is divided into a grid of three-dimensional volume pixels, also known as voxels, which is set up such that each voxel is bounded in latitude, longitude and altitude, as shown in figure 4.1 obtained from Mitchell

(2005). Assuming that the electron concentration is constant within each voxel and contained in the column vector $\boldsymbol{x}$, the problem of inversion can be expressed as:

$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \tag{4.1}$$

where $\boldsymbol{A}$ is an $i \times j$ matrix of the path length of a satellite-to-receiver signal propagating through each voxel, and $\boldsymbol{b}$ are the i slant TEC (sTEC) measurements. (Mitchell, 2005)

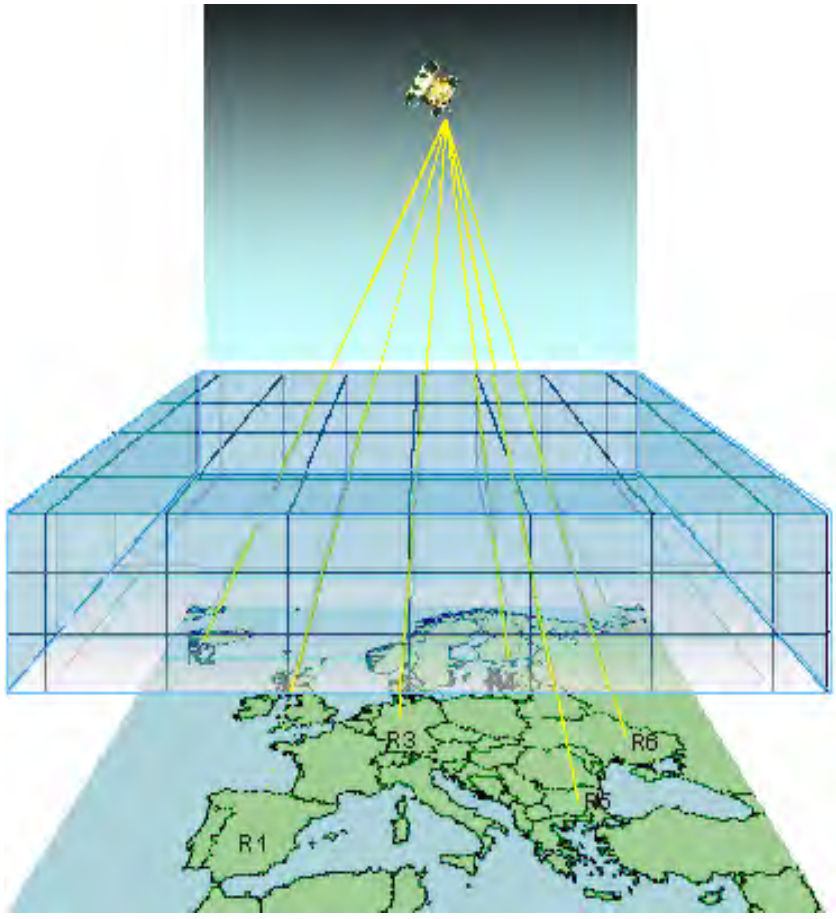

Figure 4.1: Illustration of the concept of dividing an ionospheric region into a three-dimensional grid. The figure also illustrates the ray paths from a satellite to different receivers traversing voxels representing the ionosphere. (Mitchell, 2005)

The matrix $\boldsymbol{A}$ is highly singular and incorporates no prior information as to the likely solution, and thus it is difficult to directly solve for the electron concentration

($\boldsymbol{x}$) in equation 4.1. To overcome this difficulty a mapping matrix $\boldsymbol{X}$ is used to transform the problem into a set of orthonormal basis functions with n unknown coefficients. The problem is then expressed as:

$$\boldsymbol{AXM} = \boldsymbol{b} \tag{4.2}$$

where $\boldsymbol{M}$ are the unknown coefficients representing the relative contribution of the basis functions, such that $\boldsymbol{XM}$ defines a basis set of line integrations of electron concentration. The choice of the orthonormal basis functions and the combination of its coefficients is critical, as they will give the final solution of the electron concentration. MIDAS uses empirical orthonormal functions (EOFs) for the radial variation of the ionospheric electron concentration and the spherical harmonics for the horizontal variation. The n coefficients are obtained from:

$$\boldsymbol{M} = (\boldsymbol{AX})^{-1}\boldsymbol{b} \tag{4.3}$$

where $(\boldsymbol{AX})^{-1}$ is a generalised inverse matrix such that $\boldsymbol{M}$ is the most likely solution.

MIDAS has an option of generating either one-dimensional basis functions or three-dimensional basis functions. In the case of the one-dimensional basis functions, a set of orthonormal functions are generated for the vertical profile and these functions are then modulated by spherical harmonics in latitude and longitude. The choice of one-dimensional basis functions is unstable when there are large variations in the vertical profile within the ray intersection volume, such as those due to the equatorial anomaly. In the latter the option of three-dimensional basis functions can be utilised. For the three-dimensional basis functions, the first order term is essentially the mean model in three-dimensional space, and the higher order terms define the linear departures from this. The basis functions are still modulated by the spherical harmonics in latitude and longitude.

There are two options in MIDAS for solving the inverse matrix $(\boldsymbol{AX})^{-1}$, and those are the singular value decomposition (SVD) method and the LU (Lower and Upper triangular matrix) decomposition method. Applying the SVD method, which is the MIDAS default method, to the matrix $\boldsymbol{AX}$, two orthogonal matrices

$\boldsymbol{U}$ and $\boldsymbol{V}$ and a diagonal matrix of singular values $\boldsymbol{w}$ are returned:

$$(\boldsymbol{AX})^{-1} = \boldsymbol{V}(diag(1/\boldsymbol{w}))\,\boldsymbol{U}^T. \tag{4.4}$$

Therefore, the solution to the inverse problem is given by:

$$\boldsymbol{M} = (\,\boldsymbol{V}(diag(1/\boldsymbol{w})\,\boldsymbol{U}^T))\boldsymbol{b}. \tag{4.5}$$

Finally, the electron densities within each voxel, j, are retrieved using

$$\boldsymbol{x} = \boldsymbol{XM}. \tag{4.6}$$

The problem with using the SVD method is instability when using the ionosonde data to provide information on the vertical profile or when using higher order (3 to 9) EOFs which introduces too much ambiguity in the basis functions. The LU decomposition method provides stability by including a regularisation matrix that defines a default state for the ionosphere in terms of the basis functions themselves. For the one-dimensional EOFs, the default state is provided by a globally constant ionosphere with a profile defined by the first EOF, and for the three-dimensional EOFs, it is the first EOF or the mean model. Spencer (2002) gives more information on this method.

The time-dependent inversion is implemented by assuming that the change of electron concentration within each voxel with time is linear, provided a relatively short time period, about 30 seconds, is used. This is expressed mathematically as:

$$\boldsymbol{Dy} = \boldsymbol{c} \tag{4.7}$$

where the matrix $\boldsymbol{D}$ describes the change in the ray path geometry, $\boldsymbol{y}$ is the unknown change in electron content and $\boldsymbol{c}$ is the change in TEC. The mapping matrix $\boldsymbol{X}$ is used to transform the problem to the one for which the unknowns are the linear changes in coefficients $\boldsymbol{G}$ of a set of n appropriately selected basis

functions

$$\boldsymbol{DXG} = \boldsymbol{c}. \tag{4.8}$$

The change in the unknown contributions of each of these line integrations of electron concentration is :

$$\boldsymbol{G} = (\boldsymbol{DX})^{-1}\mathbf{c}. \tag{4.9}$$

Thus, the time-dependent solution is obtained by

$$\boldsymbol{y} = \boldsymbol{XG}. \tag{4.10}$$

This algorithm is also discussed in Mitchell and Spencer (2003). The default mapping setting of MIDAS is conversion from the voxel-based to the orthonormal representation using the EOFs in the vertical domain and the spherical harmonics in the horizontal domain. The EOFs can be derived from the Chapman model, Epstein model, IRI model, and measured ionosonde profiles.

4.2 The MIDAS Inversion Procedure

Presented in this section is a description of the application of the inversion procedure within MIDAS, and the four different methods used to obtain the best profiles from MIDAS. The GPS data used in all methods was collected every 30 seconds and hourly reconstructions of the ionospheric electron density profiles were done. The reconstruction grid used is a three dimensional space covering the whole earth, with a latitude resolution of 1° and a longitude resolution of 2°. The radial range of the grid was from an altitude of 80 km to 1 181 km above the earth's surface, with a radial resolution of 50 km. The one-dimensional basis functions were used for all reconstructions, and the LU decomposition was used to solve the inversion problem.

METHOD 1: The vertical basis functions, i.e. the EOFs, were first described by the Chapman model, with peak heights ranging from 250 km to 400 km, and with scale heights ranging from 30 km to 120 km. The reconstructed image was

constrained by a latitude range of 35° to 20° south and a longitude range of 10° to 40° east. The latitude and longitude constraint was imposed to coincide with the geographical area covered by the South African region.

METHOD 2: Empirical functions were constructed from the IRI 2001 model profiles for a range of latitudes and longitudes at the reconstruction time. There were 23 profiles obtained from the model and used to construct the EOFs. For 11 of the profiles, the latitude was taken as 25° south, this being the mean latitude in the range of 10° to 40° south, and the longitude was varied in increments of 15° from 75° to the west of Grahamstown to 75° east of Grahamstown. The other 12 profiles were obtained by varying the latitude in 10° increments from 60° north of Grahamstown to 60° south of Grahamstown.

METHOD 3: The inversion procedure utilised the combination of the GPS data and the ionosonde data. The ionosonde data used here was just the peak parameters from the Louisvale and Madimbo stations. The peak parameters chosen were the maximum frequencies of the E, F1 and F2 regions (i.e. foE, foF1, and foF2), the peak heights of the E, F1, and F2 regions (hmE, hmF1, and hmF2) and the propagation factor M(3000)F2. The electron densities obtained from the ionosonde, calculated from the frequencies using equation (3.16), MIDAS treats as absolute data values, meaning that these were taken to be equal to the line integral measurements to supplement the GPS line integral measurements. The Chapman model was used to determine the orthogonal empirical functions, with peak heights in the range of 250 km to 450 km, and with scale heights in the range 30 km to 100 km. The same procedure was followed when the peak parameters from the South African Bottomside Ionospheric Model (SABIM) (McKinnell, 2002), were used to supplement the GPS data, instead of the ionosonde peak parameters.

METHOD 4: The last procedure used the ionosonde profiles from the Louisvale and Madimbo stations to construct the empirical functions. To be able to do this, the ionosonde electron densities have to be interpolated to the radial grid vertices using the MATLAB spline function, before solving the inversion matrix using the LU decomposition method. Consequentially, the radial grid vertices vary for the different reconstructed profiles and thus will be specified for each profile in

the results chapter (chapter 5).

The different procedures mentioned here were used to obtain the empirical functions, and together with the spherical harmonics solve the inversion problem. The results of the performance of MIDAS for each of these procedures are given in the following chapter.

Chapter 5

Results

In this chapter the electron density profiles obtained for eight days in 2005 will be presented and discussed. The year 2005 was chosen at the commencement of the project in August 2006, because the most recent available data was from 2005. The GPS data was obtained from about 8 to 10 South African stations. Their names and locations are given in Table 5.1, and their locations shown in figure 5.1. These stations were chosen due to their proximity to the Grahamstown station at which the algorithm is applied. The average distance between the Grahamstown station and the surrounding stations is 306.73 km.

In this project, the IRI 2001 model is used to produce the lowest limit of the

Table 5.1: GPS receiver locations, and their distance from Grahamstown. The coordinates for the stations were obtained from http://www.trignet.co.za.

Station name	**Station code**	**Latitude °S**	**Longitude °E**	**Distance (km)**
Bloemfontein	BFTN	29.06	26.17	472.70
Beaufort West	BWES	32.35	22.57	411.80
Durban	DRBN	29.57	30.56	564.00
East London	ELDN	33.02	27.49	94.63
George	GEOR	34.00	22.22	406.40
Graaff Reinet	GRNT	32.25	24.53	220.40
Grahamstown	GRHM	33.30	26.53	0.00
Port Elizabeth	PELB	33.59	25.36	113.20
Queenstown	QTWN	31.54	26.55	195.70
Umtata	UMTA	31.32	28.40	281.70

comparison task, i.e. MIDAS is expected to construct a more accurate profile

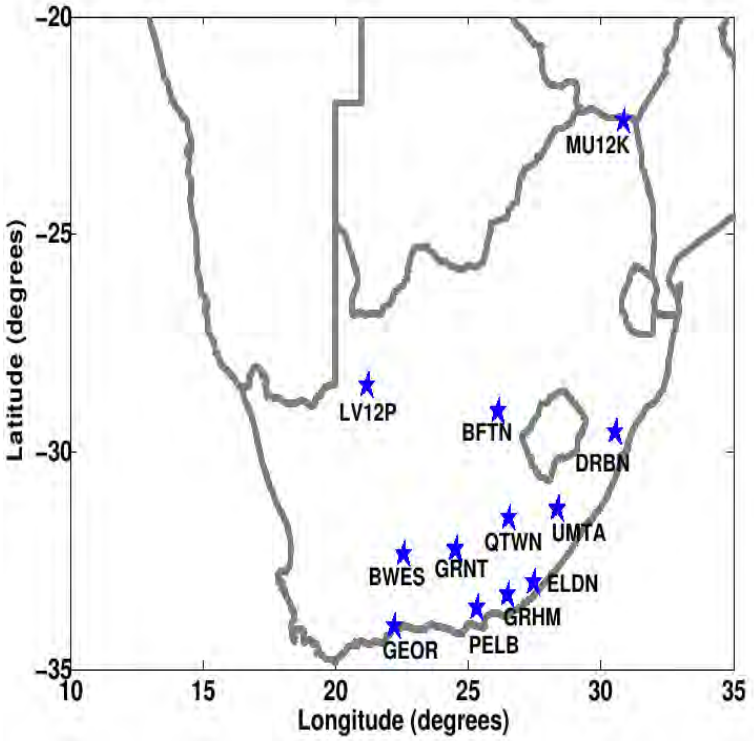

Figure 5.1: The map showing the GPS receivers locations and the ionosonde stations. LV12P is the Louisvale station code and MU12K is the Madimbo station code. The map is only drawn to show the locations of the receivers as well as the ionosonde stations and thus not drawn to scale.

than IRI and produce more accurate NmF2 values, because the IRI model uses mean parameters to produce the electron density profiles. The ionosonde data is used as true measurements. The national model SABIM will be used to gauge the accuracy of the NmF2 value produced by MIDAS.

5.1 Electron Density Profiles

As mentioned in section 4.2, there are different methods of using MIDAS to construct the electron density profiles. The choice of the method to use in constructing the profile is based on how closely the bottomside profile's shape fits that of the ionosonde and how closely the produced NmF2 value is to the measured NmF2 value. The electron density profiles obtained by using the Chapman model to create the empirical functions (method 1 in section 4.2) exhibited negative densities

at low altitudes (less than about 250.00 km) and at altitudes corresponding to the topside ionosphere, when using high order functions (4 to 9 EOFs) in the vertical domain (see figure 5.2(a)). Figures 5.2 (a) to (c) are shown for illustration purposes. The profile in figure 5.2(a) was reconstructed for 21 June 2005 at 10h00 UT using 4 EOFs, 60 latitude and 4 longitude spherical harmonics. Using the IRI model to create the EOFs (method 2 in section 4.2) improved the reconstructed profiles, in that MIDAS was not returning negative densities anymore, but the MIDAS profiles, especially for June and December, were still not accurate as the peak height and peak electron density values were much higher than those of the ionosonde. The profile in figure 5.2(b) was reconstructed using the IRI 2001 model to create EOFs. It was also reconstructed for 21 June 2005 at 10h00 UT, but using 3 EOFs, 60 latitude and 4 longitude spherical harmonics. The profiles obtained by using the ionosonde parameters to supplement the GPS data (method 3 in section 4.2) had an anomalous density peak above the F2 region for EOFs greater than 3, as illustrated by figure 5.2(c). Figure 5.2(c) was reconstructed for 02 April 2005 at 12h00 UT using 4 EOFs that were created from the Chapman model, 40 latitude and 4 longitude spherical harmonics. For EOFs greater than 2 in March and April (and this was also the case when the SABIM model parameters were used) the electron densities went negative at high altitudes, i.e. above 400 km. It was found that the method that optimises MIDAS the best, i.e. produces NmF2 values close to the measured NmF2 and shape of bottomside profile fit that of the ionosonde botttonside profile, is the procedure that uses ionosonde profiles to create EOFs (method 4 in section 4.2), and thus this method will be used to construct the profiles that will be compared to the IRI and ionosonde profiles for evaluation of the performance of MIDAS.

The profiles were obtained for April - October, and December, at the Grahamstown station, South Africa (33.3° S, 26.5° E). One day in each month was chosen, when three random hourly profiles were reconstructed, including a noon profile (except for July, which has only the noon profile because of the sporadic nature of the available GPS data). Hourly reconstructions were chosen such that there was a morning, midday and afternoon/evening profile. The profiles obtained from MIDAS were compared to the profiles obtained from the Grahamstown ionosonde

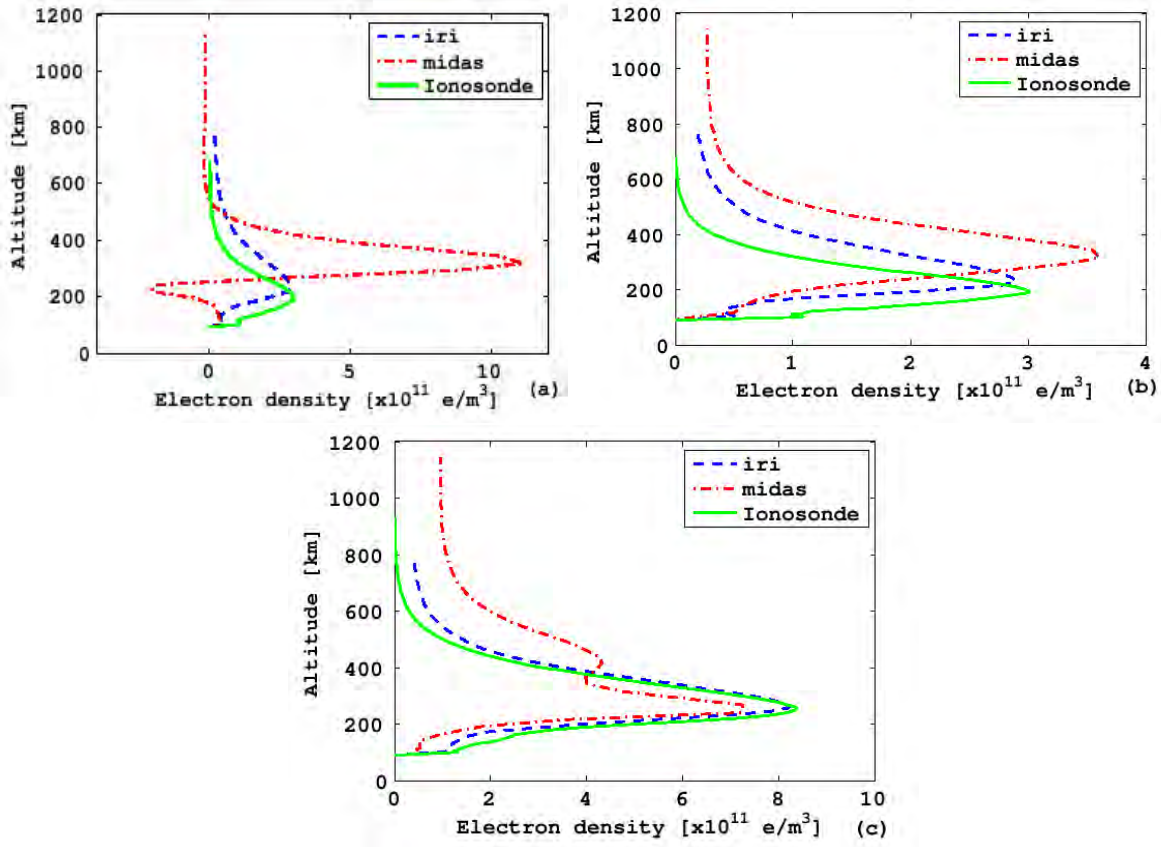

Figure 5.2: Ionosonde, MIDAS and IRI 2001 electron density profiles at Grahamstown: (a) using Chapman Profiles to create EOFs, (b) using the IRI 2001 model to construct the EOFs, and (c) using GPS and ionosonde peak parameter data for the reconstruction. The constructions were done on: (a) 21 June 2005 at 10h00 UT, (b) 21 June 2005 at 10h00 UT, and (c) 02 April 2005 at 12h00 UT. Note: These profile were chosen to illustrate the difficulties in constructing the electron density profiles using methods 1, 2 and 3.

station and to the IRI 2001 model. Profiles from the Louisvale (28.50° S, 21.20° E) and Madimbo (22.40° S, 30.90° E) ionosonde stations were used to create EOFs used in the reconstruction. For a particular day profile the same day profiles from both stations were used to create the EOFs, although sometimes just one station was used depending on the availability of the data. The criteria used to choose these Louisvale and Madimbo profiles, which were used create the EOFs, was that they must lie within two hours of the reconstruction time or that they have an electron density peak around that of the Grahamstown profile produced at the same time as the reconstruction. Figure 5.3(b) shows an example of EOFs generated by

applying the singular value decomposition to 7 Louisvale and Madimbo profiles. These EOF were then used in the reconstruction of the profile in figure 5.3(c). The reconstruction parameters namely number of EOFs, number of latitude and longitude spherical harmonics and power weights of the spherical harmonics were chosen by trial and error. For a first trial the reconstruction parameters chosen were 2 EOFs, 16 latitude harmonics and 4 longitude harmonics. The reconstruction parameters were adjusted, by increasing and/or decreasing the numbers, until the best MIDAS profile that fitted the measured ionosonde profile had been obtained. The power weights of the latitude and longitude spherical harmonics used in all profiles were 2, unless otherwise specified.

The radial grid, which was obtained from the ionosonde data, used for each profile is given in table 5.2. For each day and hour, the ionosonde altitudes were used to extrapolate to the radial grid. Thus the minimum and maximum altitudes of MIDAS were determined by the minimum and maximum altitudes for the ionosonde data used to reconstruct the profile.

5.1.1 02 April 2005

The profile in figure 5.4(a) was obtained using 6 ionosonde profiles from Madimbo and Louisvale stations to create 2 EOFs. The number of latitude harmonics used in the reconstruction is 16, and of longitude harmonics is 3. The midday profile, figure 5.4(b), was reconstructed using 2 EOFs obtained from 8 Louisvale and Madimbo profiles, 16 latitude harmonics and 2 longitude harmonics. For the night profile, figure 5.4(c), 4 profiles were used to create 4 EOFs, and 46 spherical harmonics were used - 40 latitude and 6 longitude. At noon the MIDAS profile produced a good fit to that of the ionosonde, but in the morning and at night MIDAS profile overestimated the peak parameters of the F2 region, NmF2 and hmF2.

5.1.2 12 May 2005

Only the Louisvale profiles were used to create EOFs because of technical problems with the Madimbo station. For the early morning reconstruction 4 profiles were chosen to create 2 EOFs. The horizontal profile was constrained by 4 longitude

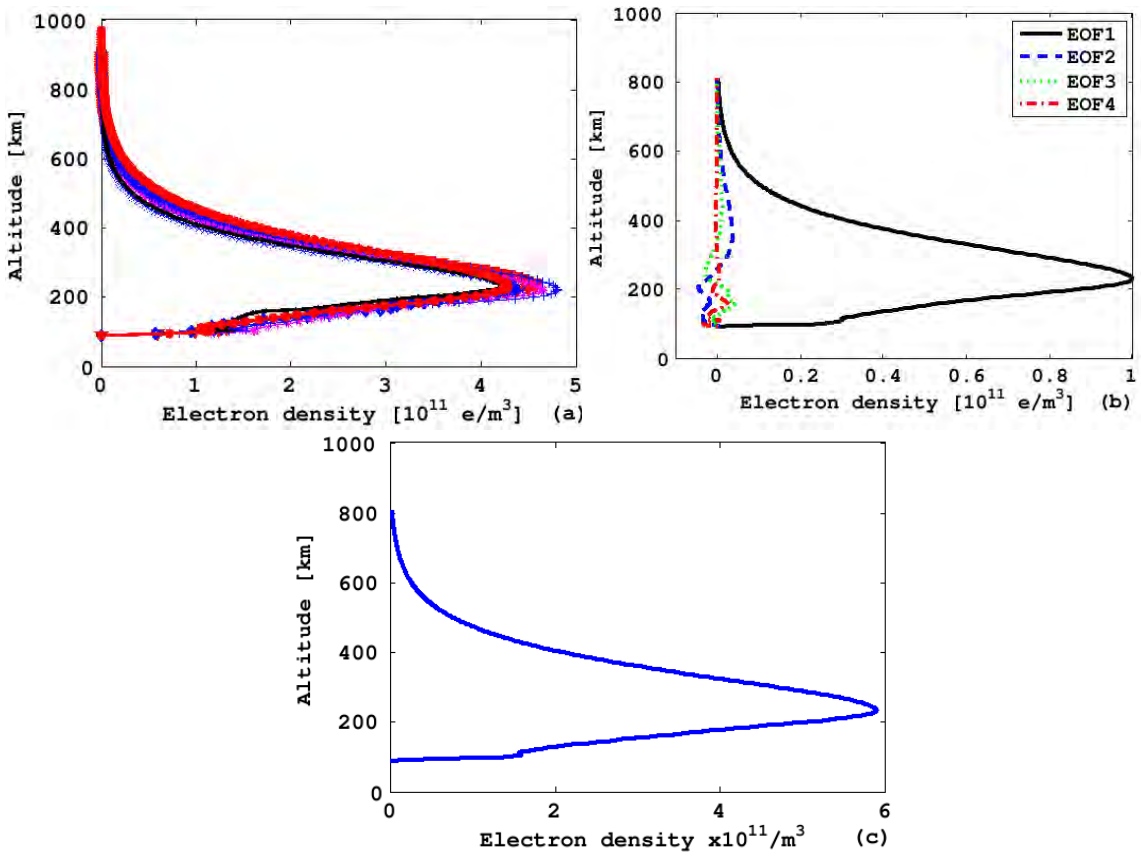

Figure 5.3: Example of how the electron density profile, shown in (a), is reconstructed using EOFs, shown in (b), created from the Madimbo and Louisvale profiles, shown in (c). The reconstruction parameters used were 4 EOFs, 16 latitude spherical harmonics, and 2 longitude spherical harmonics. The reconstruction was done over Grahamstown on 23 August 2005 at 10h00.

harmonics and 4 latitude harmonics. The midday profile was reconstructed using 2 EOFs, 6 latitude harmonics, and 4 longitude harmonics. The EOFs were created from 6 profiles. The profile in figure 5.5(c) was reconstructed by using 4 longitude spherical harmonics, 4 latitude spherical harmonics, and 5 Louisvale profiles were used to create 2 EOFs in this reconstruction. Again at noon the MIDAS profile produced a good fit to that of the ionosonde profile, while in the morning and afternoon the MIDAS profile overestimates the peak values in the F2 region, NmF2 and hmF2.

Table 5.2: Altitude range used to define the radial dimension.

Date	Time	Minimum (km)	Maximum (km)	Increment (km)
05.04.02	06h00	90.00	592.00	2.00
	12h00	90.00	744.00	2.00
	20h00	90.00	566.00	2.00
05.05.12	02h00	90.00	656.00	2.00
	12h00	90.00	760.00	2.00
	14h00	90.00	576.00	2.00
05.06.15	07h00	90.00	710.00	2.00
	12h00	90.00	790.00	2.00
	18h00	90.00	596.00	2.00
05.07.16	12h00	90.00	712.00	2.00
05.08.23	10h00	90.00	812.00	2.00
	12h00	90.00	812.00	2.00
	18h00	90.00	488.00	2.00
05.09.27	10h00	90.00	750.00	2.00
	12h00	90.00	750.00	2.00
	20h00	90.00	550.00	2.00
05.10.20	06h00	90.00	986.00	2.00
	12h00	90.00	986.00	2.00
	18h00	90.00	482.00	2.00
05.12.21	10h00	90.00	800.00	2.00
	12h00	90.00	710.00	2.00
	14h00	90.00	800.00	2.00

5.1.3 15 June 2005

The profile in figure 5.6(a) was reconstructed using 4 EOFs, 8 latitude harmonics and 8 longitude harmonics. The EOFs were created from 4 profiles obtained from the Louisvale and Madimbo stations. The 2 EOFs used in figure 5.6(b) were created from 13 Madimbo and Louisvale profiles. In this reconstruction, 6 latitude harmonics and 4 longitude harmonics were used. The power weights of these harmonics were 4. The profile in figure 5.6(c) was reconstructed using 4 EOFs created from 2 Louisvale and Madimbo profiles. The reconstructed profile used 40 latitude harmonics and 6 longitude harmonics. In the morning and at noon the MIDAS profile produces a good fit to that of the ionosonde profile, but in the afternoon MIDAS overestimates the electron density peak of the F2 region and at noon it slightly underestimates the peak height of the F2 region. The evening profile reconstructed by MIDAS had a greater peak height and electron density

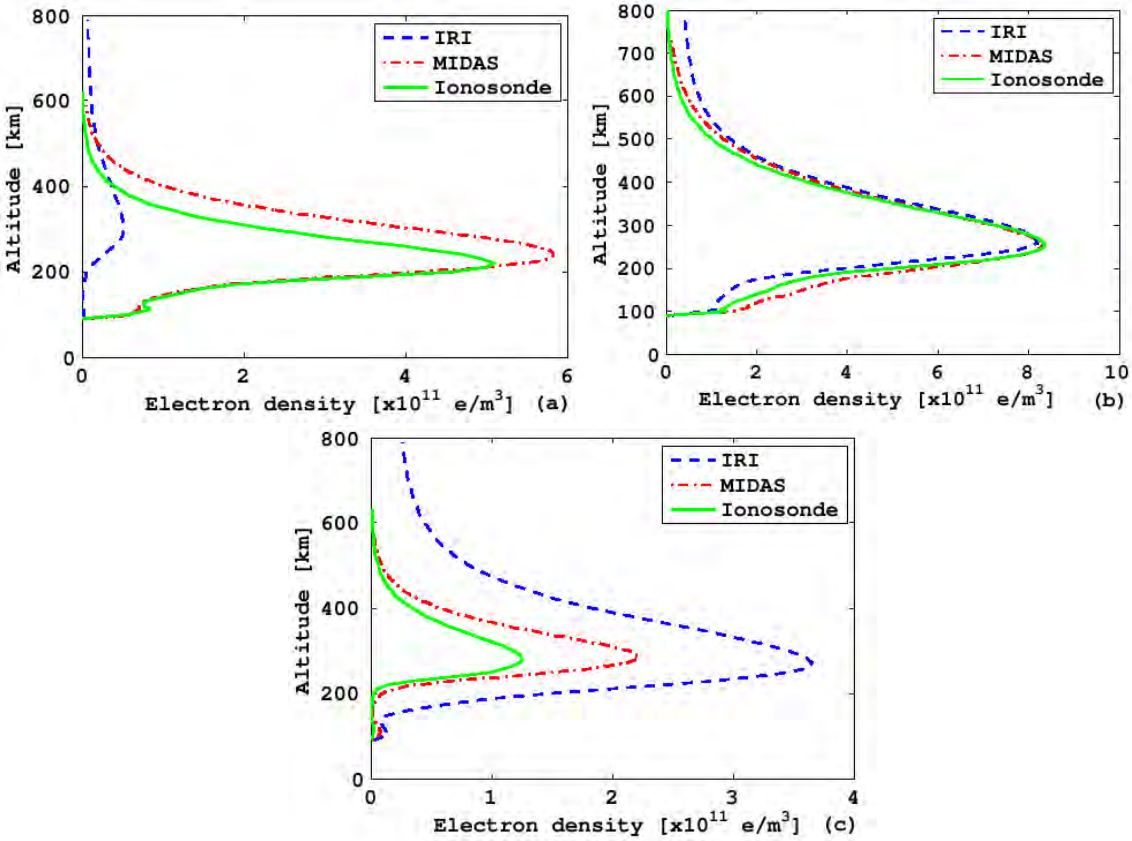

Figure 5.4: Ionosonde, MIDAS and IRI 2001 electron density profiles over Grahamstown obtained on 02 April 2005 at: (a) 06h00 UT, (b) 12h00 UT and (c) 20h00 UT.

peak of the F2 region than those of the ionosonde profile.

5.1.4 16 July 2005

For the reconstruction of the profile presented in figure 5.7, 7 Louisvale and Madimbo profiles were used to constrain the vertical ionosphere using 2 EOFs, while 4 latitude and longitude harmonics were used for the horizontal constraint. There is a good fit of the MIDAS profile to that of the ionosonde, especially when compared to the bottomside ionosphere.

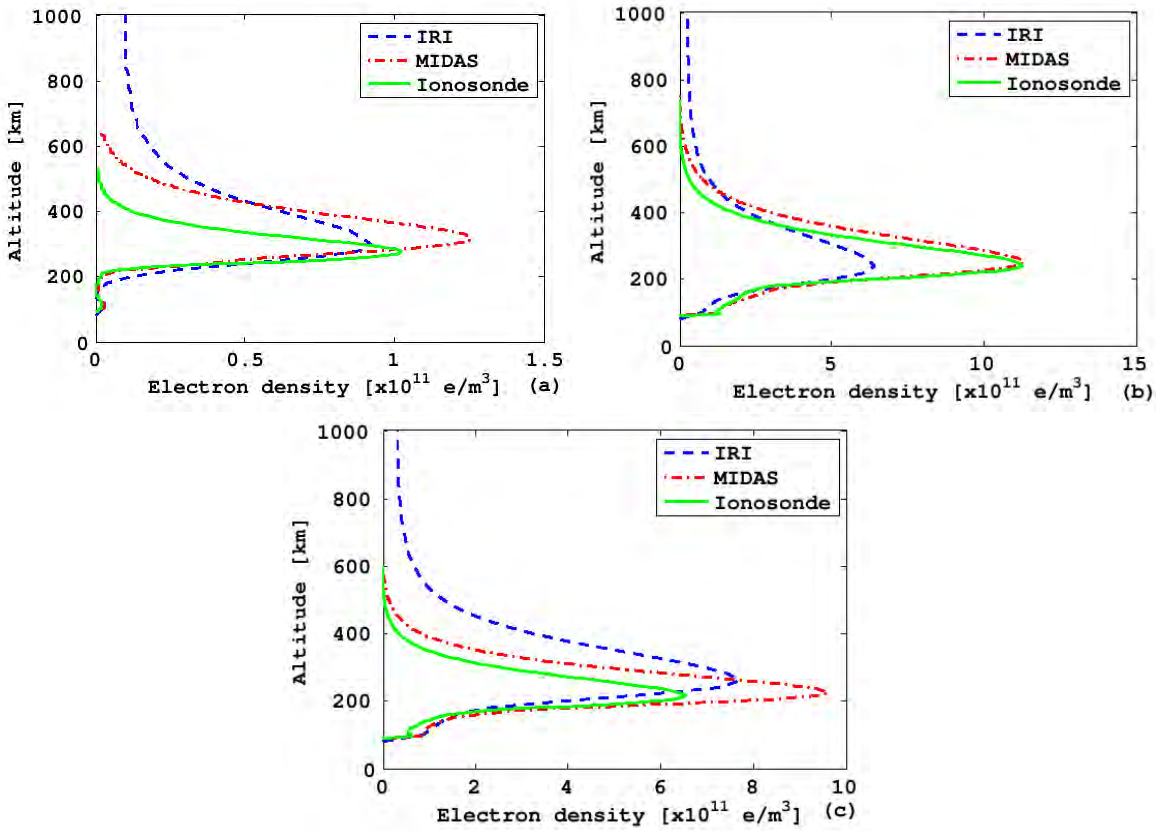

Figure 5.5: Ionosonde, MIDAS and IRI 2001 electron density profiles over Grahamstown obtained on 12 May 2005 at: (a) 02h00 UT, (b) 12h00 UT and (c) 14h00 UT.

5.1.5 23 August 2005

For the 10h00 and 12h00 (UT) profiles, 7 electron density profiles from Madimbo and Louisvale were used to create EOFs. The 10h00 profile used 2 EOFs, while the 12h00 profile used 4 EOFs to constrain the vertical profile. The number of latitude and longitude harmonics used in figure 5.8(a) and figure 5.8(b) is 4. Figure 5.8(c) used 4 EOFs, created from 5 profiles, 4 latitude and longitude harmonics. The MIDAS profiles for this day produced a good fit to that of the ionosonde, but MIDAS overestimated the electron density peak of the F2 region in the evening.

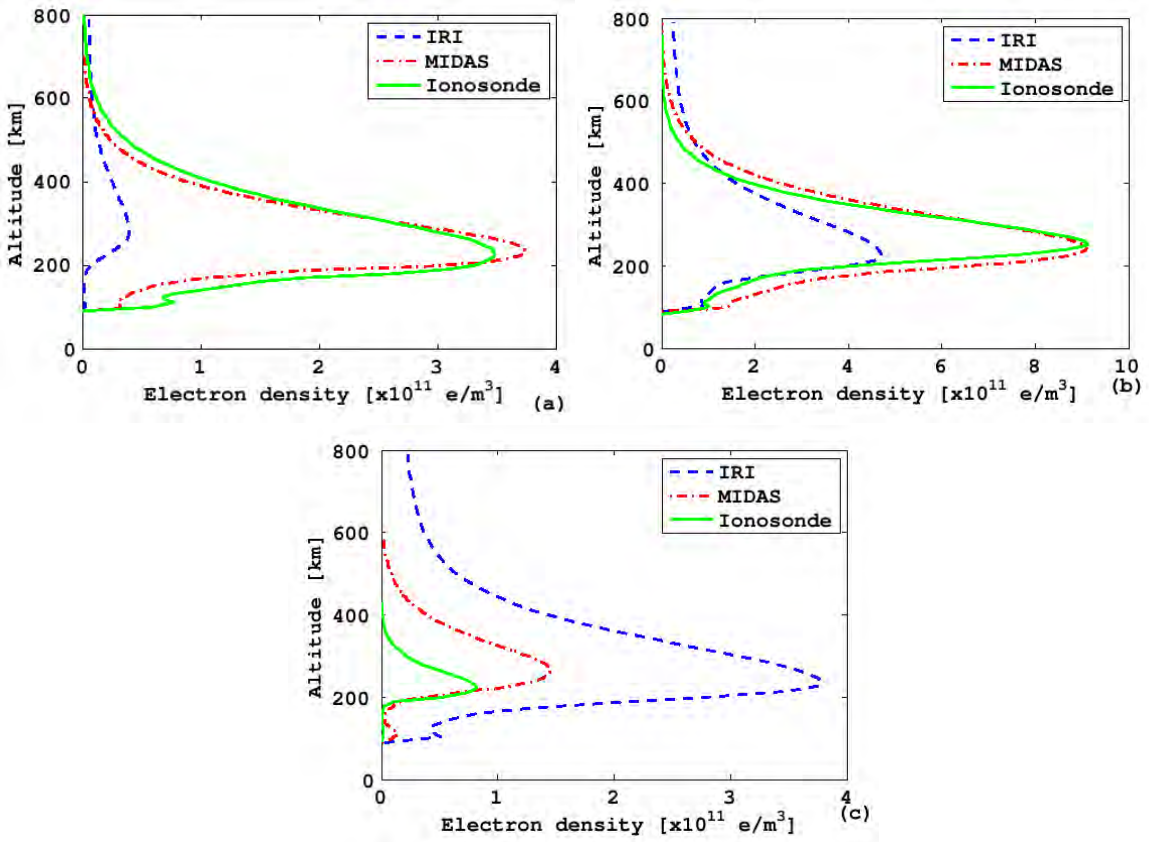

Figure 5.6: Ionosonde, MIDAS and IRI 2001 electron density profiles over Grahamstown obtained on 15 June 2005 at: (a) 07h00 UT, (b) 12h00 UT and (c) 18h00 UT.

5.1.6 27 September 2005

In figure 5.9(a) only 5 profiles were used to create 1 EOF to constrain the vertical ionosphere. There were 4 spherical harmonics, latitudinal and longitudinal, used to construct this profile. The noon profile was reconstructed using only 1 EOF, 2 latitude harmonics, and 3 longitude harmonics. The EOFs were created from 7 profiles obtained from the Louisvale and Madimbo stations. For the 20h00 profile, 5 profiles from only the Louisvale station were used to create the 3 EOFs. There were 40 latitude harmonics and 8 longitude harmonics used for the horizontal constraint of the ionosphere. Figure 5.9(a) and 5.9(b) show that the MIDAS profiles fit to the ionosonde profiles, but also showed a slight deviation of the F2 region peak parameters from the measure F2 region peak parameters. MIDAS overestimated the peak parameters of the F2 region at night.

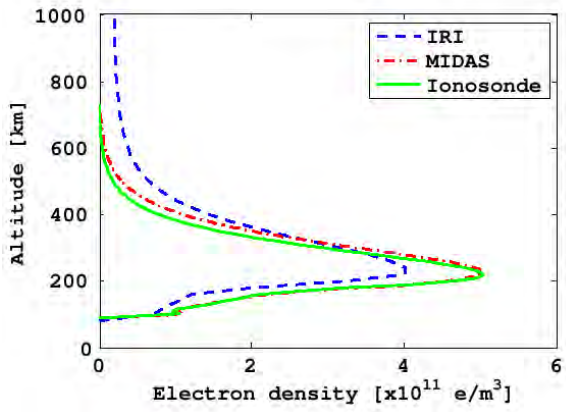

Figure 5.7: Ionosonde, MIDAS and IRI 2001 electron density profiles over Grahamstown obtained on 16 July 2005 at 12h00 UT.

5.1.7 20 October 2005

For the morning profile, i.e. figure 5.10(a), 4 EOFs, 4 longitude and latitude harmonics were used in the reconstruction of the electron density profiles. The EOFs were created from 6 profiles obtained from both ionosonde stations. The noon profile was reconstructed using 4 EOFs, 20 latitude harmonics and 6 longitude harmonics. The 5 profiles used to create the EOFs were obtained from both stations (Louisvale and Madimbo). The profile in figure 5.10(c) was reconstructed by using 4 EOFs, created from 4 profiles, 4 longitude and latitude harmonics. In the morning and in the evening, there was no good correlation between the MIDAS profiles and the ionosonde profiles. MIDAS overestimated the peak height and the electron density peak of the F2 region in the morning and overestimated the electron density peak in the evening profile. The afternoon MIDAS profile produced a good fit to that of the ionosonde, with a slight deviation in the peak height of the F2 region.

5.1.8 21 December 2005

These profiles (see figure 5.11) were reconstructed using 9 profiles from the Louisvale station to create the EOFs. For all profiles 2 EOFs, 4 latitude and longitude harmonics were used to reconstruct the electron density profiles. None of the MIDAS profiles fitted those of the ionosonde and MIDAS overestimated all the peak parameters of the F2 region for all the reconstructions.

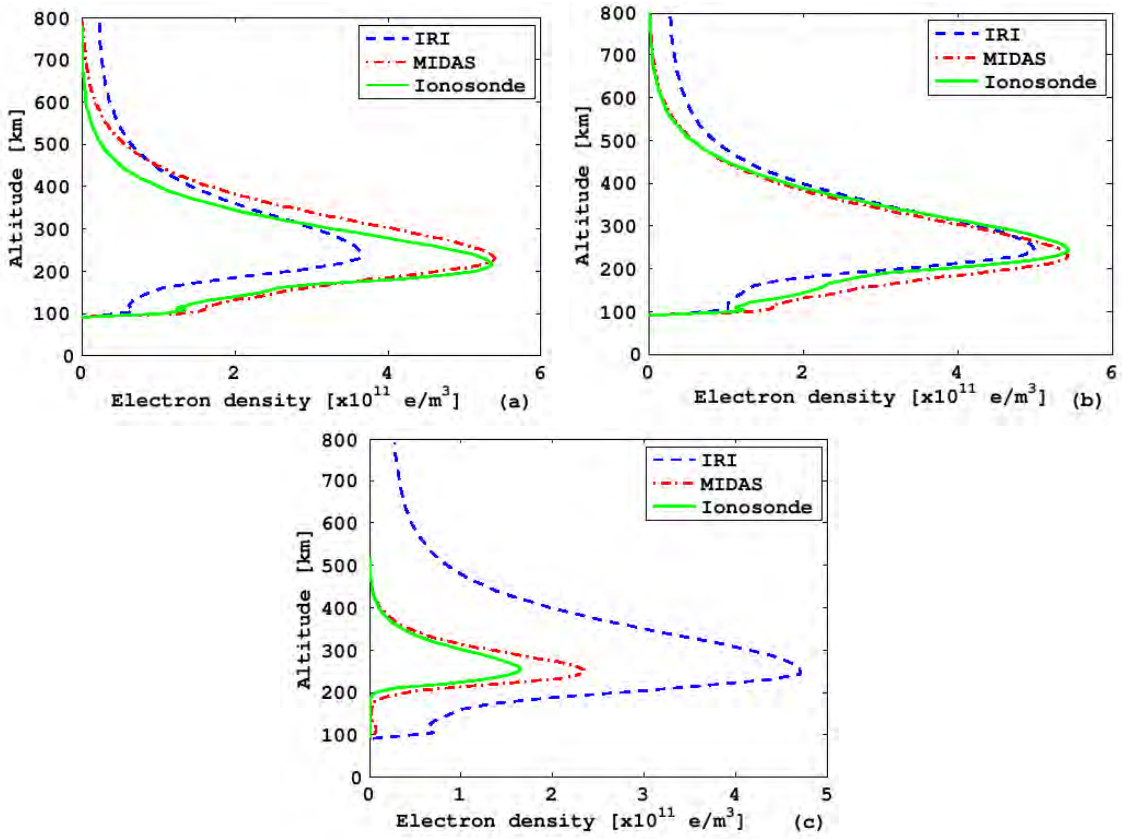

Figure 5.8: Ionosonde, MIDAS and IR 2001 electron density profiles over Grahamstown obtained on 23 August 2005 at: (a) 10h00 UT, (b) 12h00 UT and (c) 18h00 UT.

5.2 Analysis

This section discusses the results presented in the previous section. The important parameter for discussion is the peak electron density of the F2 region, NmF2. The NmF2 values obtained from the MIDAS algorithm are compared to those obtained from the IRI 2001 model and also to those obtained from the ionosonde (see table 5.3 and figure 5.11). The performance of MIDAS in exhibiting diurnal, and seasonal variation is evaluated in figure 5.12 and figure 5.13 respectively.

Figure 5.12(a) illustrates the accuracy of MIDAS and the IRI 2001 model in predicting NmF2 during the course of the year. From this figure it is clear that MIDAS produces more accurate NmF2 values than of the IRI 2001 model. The ionosonde provides the true measured NmF2 value, and it can be seen that MIDAS

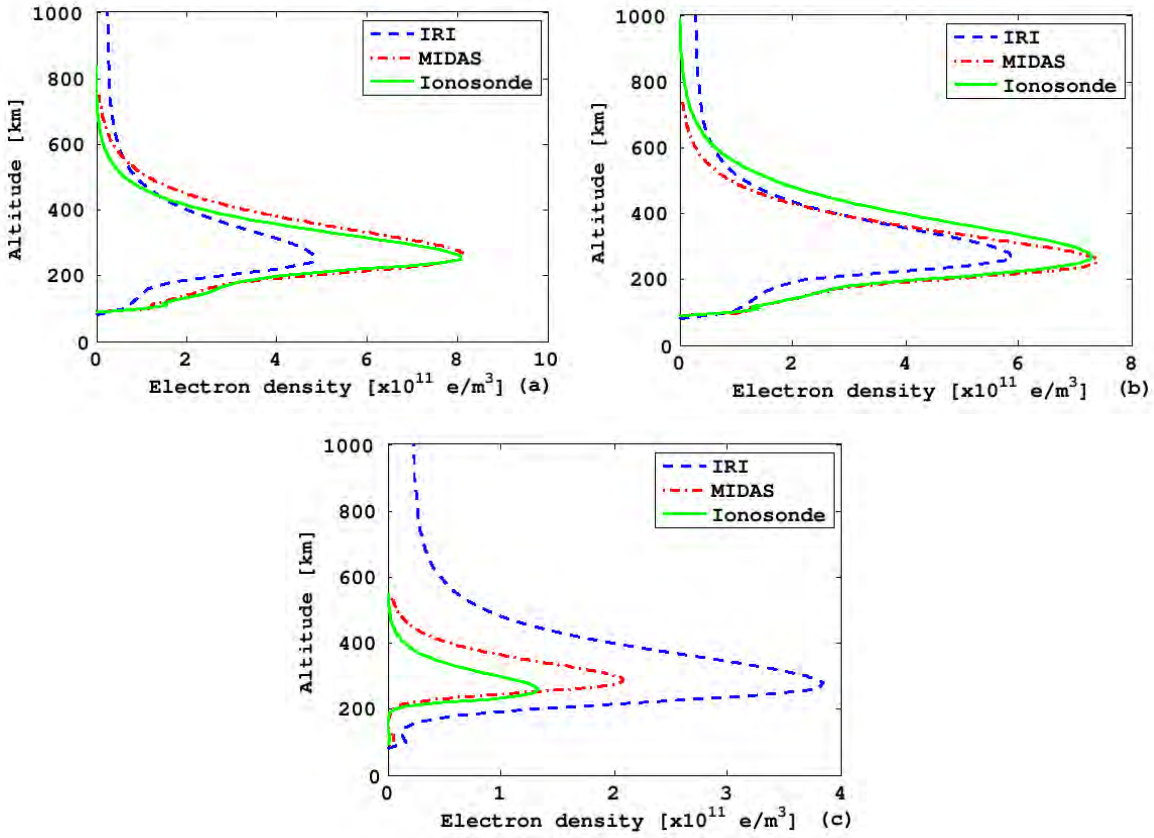

Figure 5.9: Ionosonde, MIDAS and IRI 2001 electron density profiles over Grahamstown obtained on 27 September 2005 at: (a) 10h00 UT, (b) 12h00 UT and (c) 20h00 UT.

produced NmF2 values closer in value to the measured value than the IRI 2001 model.

Figures 5.13(a) to 5.13(c) compare the MIDAS NmF2 values to the ionosonde NmF2 values, and tests for the accuracy in the MIDAS production of the NmF2. Figures 5.13(d) to 5.13(f) compare the IRI 2001 model NmF2 values to those of the ionosonde.

Figures 5.14(a) to 5.14(c) illustrate the accuracy of the MIDAS NmF2 values compared to those of the ionosonde NmF2 values, while figures 5.14(d) to 5.14(f) illustrate a comparison of NmF2 values from IRI 2001 model for three different seasons. The slope, y-intercept, co-efficient of correlation and rms error values for

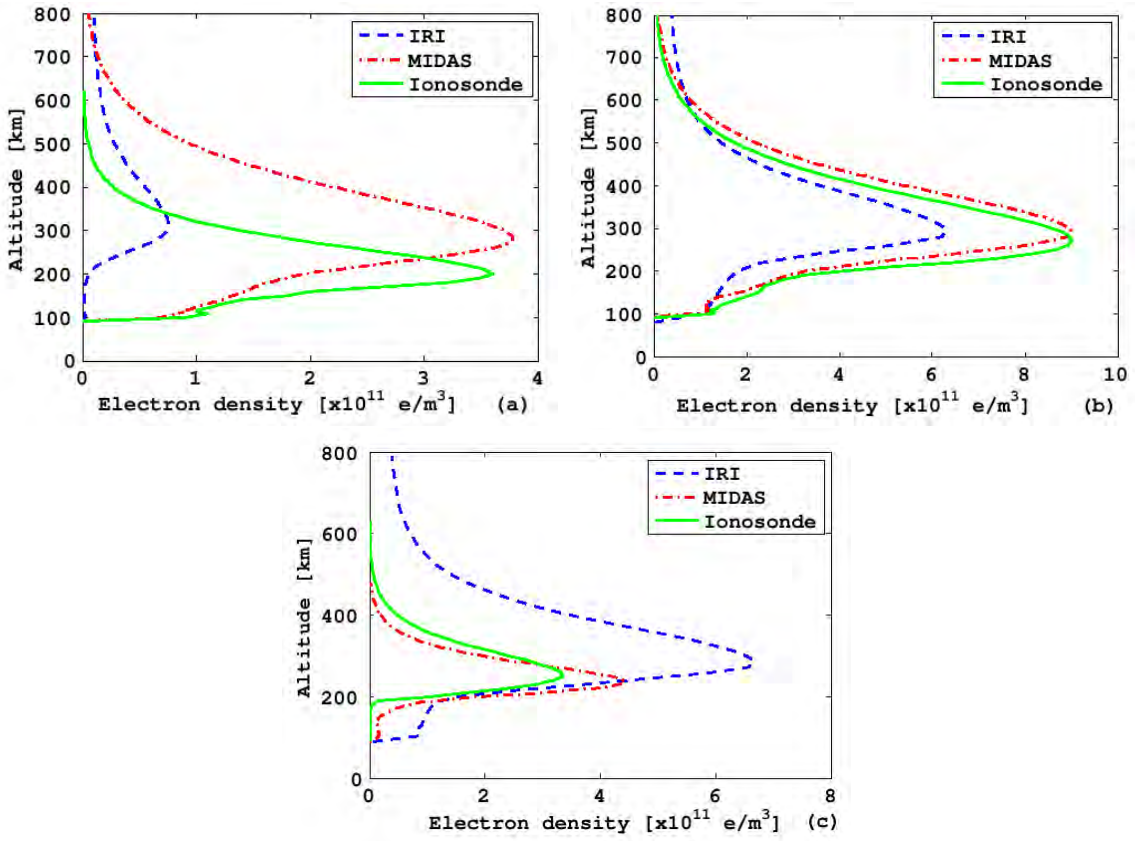

Figure 5.10: Ionosonde, MIDAS and IRI 2001 electron density profiles over Grahamstown obtained on 20 October 2005 at: (a) 06h00 UT, (b) 12h00 UT and (c) 18h00 UT.

figure 5.12, 5.13 and 5.14 are presented in table 5.4 for MIDAS plots and table 5.5 for the IRI 2001 plots.

MIDAS, on average, gets the shape of the profile very close to that of the measured profile (for example see figure 5.6) although it sometimes overestimate the peak parameters (for example see figures 5.4(c) and 5.8 (c)). MIDAS performs the worst in December (see figure 5.11) generally not matching the shape of the measured profile regardless of the hour of the day. This behaviour could be explained by the fact that during summer the ionosphere is very variable to such an extent that the variation of the electron density with time might be nonlinear over the period of one hour. The MIDAS inversion program assumes a linearly variation of the electron density with time during the hour of the reconstruction, and therefore

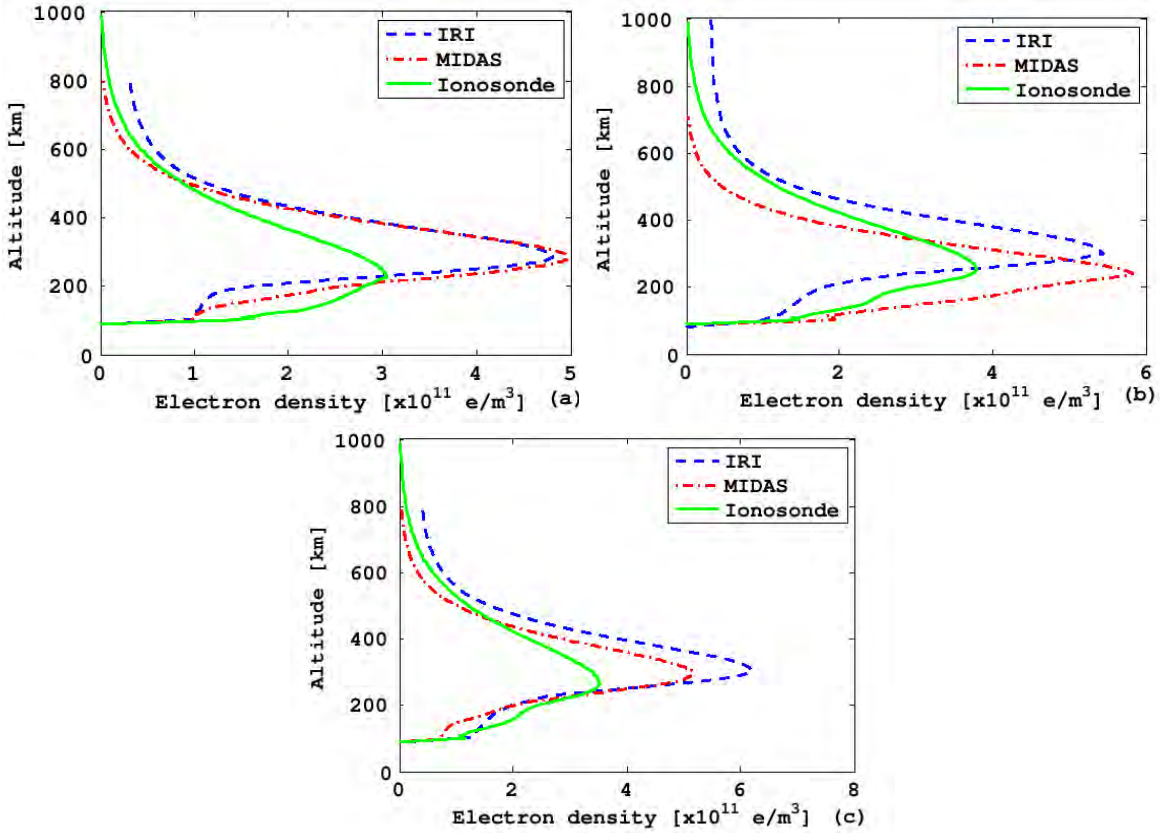

Figure 5.11: Ionosonde, MIDAS and IRI 2001 electron density profiles over Grahamstown obtained on 21 December 2005 at: (a) 10h00 UT, (b) 12h00 UT and (c) 14h00 UT.

this could explain why the shape of the MIDAS profile does not match that of the measured profile in summer.

The root mean square (rms) error determines how much the calculated data deviates from the observed data, in other words, how well the derived or calculated data fit the measured data. Therefore, to determine if MIDAS produces a realistic NmF2 value, its rms error has to be less than that of the IRI 2001 model. Comparing the rms error values in table 5.4 and table 5.5, it is observed that MIDAS generally produces more realistic NmF2 values than the IRI 2001 model because it has lower rms error values. MIDAS performs better at noon than at the other times of the day, as the noon rms error values are generally lower than those of the other times. Also, from the rms error values, it is seen that during the winter

Table 5.3: Peak parameters of the F2 region, NmF2 [$10^{11}e/m^3$] and hmF2 [km]. The date is written in the format yy.mm.dd and the time is universal time (UT).

Date	**Time**	**Index**	**Ionosonde**		**MIDAS**		**IRI 2001**	
			NmF2	**hmF2**	**NmF2**	**hmF2**	**NmF2**	**hmF2**
	06h00	1	5.09	220.00	5.83	241.00	0.51	314.00
05.04.02	12h00	2	8.36	253.22	8.31	257.00	8.20	258.00
	20h00	3	1.25	280.58	2.20	291.00	3.65	272.00
	02h00	4	1.02	280.00	1.25	325.00	0.92	300.00
05.05.12	12h00	5	11.25	241.20	11.29	251.00	6.40	240.00
	14h00	6	6.52	216.70	9.59	225.00	7.69	260.00
	07h00	7	3.48	230.00	3.74	237.00	0.39	286.00
05.06.15	12h00	8	9.12	251.70	9.10	243.00	4.70	230.00
	18h00	9	0.82	224.57	1.46	265.00	3.78	230.00
05.07.16	12h00	10	5.04	220.00	5.02	229.00	4.01	240.00
	10h00	11	5.36	215.80	5.40	231.00	3.66	244.00
05.08.23	12h00	12	5.44	241.96	5.43	233.00	5.01	244.00
	18h00	13	1.65	253.24	2.34	253.00	4.70	244.00
	10h00	14	8.09	253.22	8.14	265.00	4.84	260.00
05.09.27	12h00	15	7.30	270.00	7.39	259.00	5.88	260.00
	20h00	16	1.33	260.00	2.08	289.00	3.85	280.00
	06h00	17	3.60	200.00	3.77	287.00	0.75	314.00
05.10.20	12h00	18	9.01	270.87	9.00	291.00	6.30	300.00
	18h00	19	3.35	252.55	4.43	237.00	6.68	286.00
	10h00	20	3.04	230.00	4.95	291.00	4.82	286.00
05.12.21	12h00	21	3.79	251.73	5.82	241.00	5.45	300.00
	14h00	22	3.52	264.71	5.16	297.00	6.20	300.00

season MIDAS has a greater ability to produce the measured NmF2 values.

The accuracy of the NmF2 values obtained from MIDAS were evaluated by comparing them to the NmF2 values obtained from SABIM. This was done by comparing the absolute error values of SABIM (i.e. |NmF2(SABIM) - NmF2(ionosonde)|) with the absolute error values of MIDAS (i.e. |NmF2 (MIDAS) - NmF2(ionosonde)|). The accuracy of MIDAS will always contain an uncertainty since MIDAS is essentially a modelling algorithm that ingests empirical data. Therefore, to judge the limits within which MIDAS is performing accurately, and to define what is meant by accuracy, the national model SABIM was used, with the understanding that SABIM is currently the best model for the Grahamstown region. So for MIDAS to

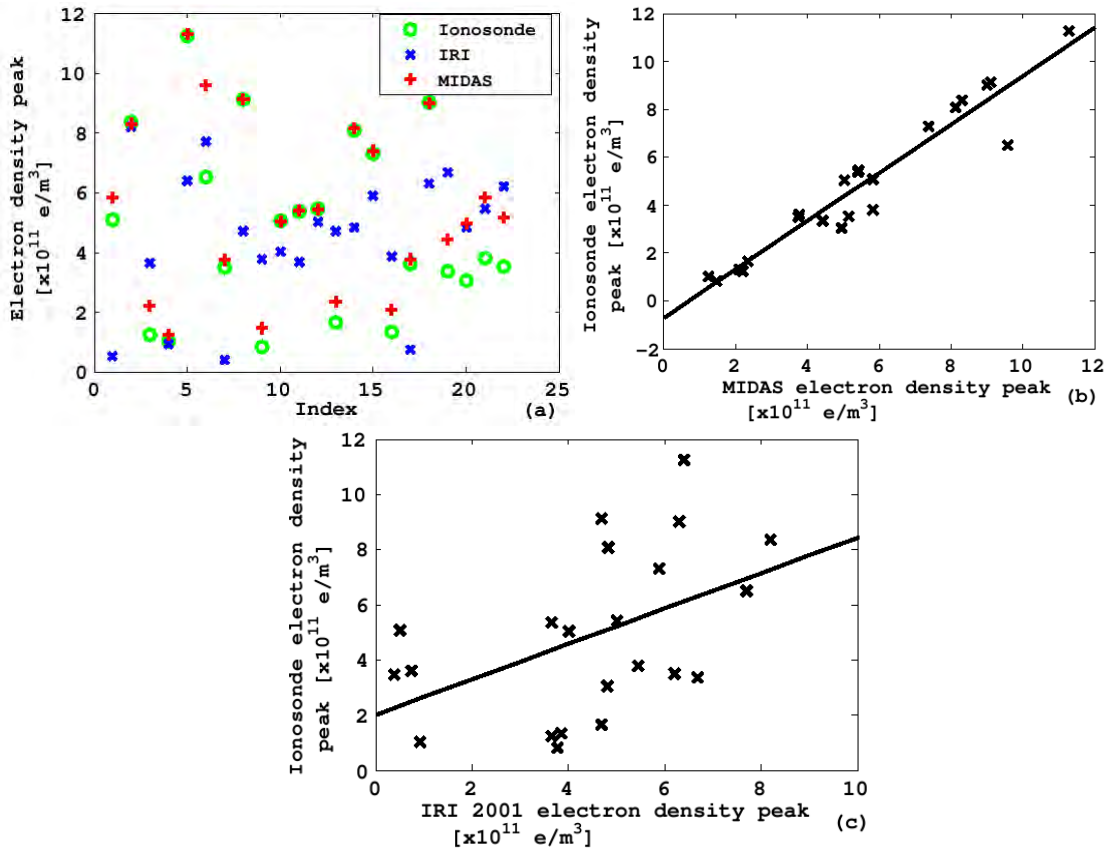

Figure 5.12: Comparison of peak electron density values between: (a) Grahamstown ionosonde, the IRI 2001 model and MIDAS, (b) MIDAS and the Grahamstown ionosonde, and (c) IRI 2001 model and the Grahamstown ionosonde. The index represents the hour and date for which the NmF2 value was obtained.

be good enough to replace this model or even supplement the ionosonde data for places where there are no ionosondes, MIDAS has to produce NmF2 values more accurately than SABIM. For a more accurate prediction, we expect the absolute error of the NmF2 values produced by MIDAS to be less than the absolute error of the NmF2 values predicted by SABIM. For the sample year used in this study, the NmF2 values predicted by SABIM and produced by MIDAS with their respective absolute errors are presented in table 5.6. From this table it is evident that MIDAS produced more accurate NmF2 values 13 out of 22 times. From this statement we can infer that there is an approximately 59% chance that MIDAS will produce an NmF2 value that deviates from the measured value by 1.05×10^{11} $[e/m^3]$.

Table 5.4: The MIDAS-Ionosonde statistical analysis. The rms error is in units of $10^{11} e/m^3$

	slope	**y-intercept**	**co-efficient of correlation**	**rms error**
Figure 5.10b	1.01	-0.70	0.96	1.05
Figure 5.11a	0.99	-0.45	0.95	0.79
Figure 5.11b	1.11	-1.10	0.96	0.72
Figure 5.11c	0.70	-0.10	1.00	1.49
Figure 5.12a	0.95	-0.45	0.96	1.35
Figure 5.12b	1.11	-0.71	1.00	0.37
Figure 5.12c	1.12	-1.02	0.99	0.54

Table 5.5: The IRI 2001-Ionosonde statistical analysis. The rms error is in units of $10^{11} e/m^3$

	slope	**y-intercept**	**co-efficient of correlation**	**rms error**
Figure 5.10c	0.64	2.02	0.48	2.71
Figure 5.11d	0.53	3.04	0.49	2.81
Figure 5.11e	0.92	2.14	0.47	2.66
Figure 5.11f	1.17	-3.48	0.95	2.67
Figure 5.12d	0.81	1.89	0.68	2.93
Figure 5.12e	0.42	2.83	0.24	2.71
Figure 5.12f	0.62	2.54	0.44	2.75

To determine at what time of the day MIDAS performs best, i.e. gives accurate NmF2 values most of the time, table 5.7, 5.8 and 5.9 were drawn up. Tables 5.10 to table 5.12 were drawn up to assess during which season MIDAS produces the most accurate NmF2 values. Table 5.7 gives the NmF2 values during the morning hours, and from this table we can see that MIDAS produces more accurate values, compared to SABIM, 6 out of a total of 7 opportunities. Thus, it is deduced that during the morning hours there is an approximately 86% chance that MIDAS will produce an NmF2 value closer to the measured NmF2 value, within 0.79×10^{11} $[e/m^3]$, than SABIM.

Table 5.8 presents the peak electron density values at noon. From these values it is evident that there is an approximately 88% chance of MIDAS producing NmF2 values closer to the measured value than SABIM, within 0.72×10^{11} $[e/m^3]$.

In the late afternoon/evening hours it is clear that MIDAS produces NmF2

Table 5.6: The NmF2 values from the ionosonde, SABIM and MIDAS, and the corresponding absolute errors of SABIM and MIDAS for the year.

Date	**Time**	**Ionosonde**	**SABIM**	**MIDAS**	$SABIM_{error}$	$MIDAS_{error}$
05.04.02	06h00	5.09	4.15	5.83	0.94	0.74
	12h00	8.36	7.52	8.31	0.84	0.05
	20h00	1.25	1.24	2.20	0.01	0.95
05.05.12	02h00	1.02	1.12	1.25	0.10	0.23
	12h00	11.25	5.94	11.29	5.31	0.04
	14h00	6.52	5.08	9.59	1.44	3.07
05.06.15	07h00	3.48	5.24	3.74	1.76	0.26
	12h00	9.12	4.46	9.10	4.66	0.02
	18h00	0.82	0.94	1.46	0.12	0.64
05.07.16	12h00	5.04	3.38	5.02	1.66	0.02
05.08.23	10h00	5.36	3.96	5.40	1.40	0.04
	12h00	5.44	4.03	5.43	1.41	0.01
	18h00	1.65	0.96	2.34	0.69	0.69
05.09.27	10h00	8.09	5.76	8.14	2.33	0.05
	12h00	7.30	5.71	7.39	1.59	0.09
	20h00	1.33	1.08	2.08	0.25	0.75
05.10.20	06h00	3.60	3.85	3.77	0.25	0.17
	12h00	9.01	7.02	9.00	1.99	0.01
	18h00	3.35	2.55	4.43	0.80	1.08
05.12.21	10h00	3.04	5.37	4.95	2.33	1.91
	12h00	3.79	5.41	5.82	1.62	2.03
	14h00	3.52	4.55	5.16	1.03	1.64

values that are mostly greater than the measured and SABIM values (see table 5.9). Thus, during these hours (from 14h00 to 20h00) there is an approximately 14% chance of MIDAS producing an accurate value for NmF2, which is within 1.49×10^{11} $[e/m^3]$ from the measured NmF2 value. This may be explained by the fact during these hours, the sun is very low and thus the ionosphere is unstable and the algorithm cannot illustrate the instability.

Table 5.10 presents the peak electron density values during the autumn season. From the number of times that the absolute error of the MIDAS NmF2 value is less than that of the absolute error of the SABIM NmF2 value, it is evident that there is an approximately 88% chance of MIDAS producing NmF2 values closer to the measured value ('close' being defined to be within 1.35×10^{11} $[e/m^3]$ from the

Table 5.7: The NmF2 values from the ionosonde, SABIM and MIDAS, and the corresponding absolute errors of SABIM and MIDAS for the morning hours.

Date	**Time**	**Ionosonde**	**SABIM**	**MIDAS**	$SABIM_{error}$	$MIDAS_{error}$
05.04.02	06h00	5.09	4.15	5.83	0.94	0.74
05.05.12	02h00	1.02	1.12	1.25	0.10	0.23
05.06.15	07h00	3.48	5.24	3.74	1.76	0.26
05.08.23	10h00	5.36	3.96	5.40	1.40	0.04
05.09.27	10h00	8.09	5.76	8.14	2.33	0.05
05.10.20	06h00	3.60	3.85	3.77	0.25	0.17
05.12.21	10h00	3.04	5.37	4.95	2.33	1.91

Table 5.8: The NmF2 $[10^{-11}e/m^3]$ values from the ionosonde, SABIM and MIDAS, and the corresponding absolute errors of SABIM and MIDAS at noon (12h00 UT).

Date	**Time**	**Ionosonde**	**SABIM**	**MIDAS**	$SABIM_{error}$	$MIDAS_{error}$
05.04.02	12h00	8.36	7.52	8.31	0.84	0.05
05.05.12	12h00	11.25	5.94	11.29	5.31	0.04
05.06.15	12h00	9.12	4.46	9.10	4.66	0.02
05.07.16	12h00	5.04	3.38	5.02	1.66	0.02
05.08.23	12h00	5.44	4.03	5.43	1.41	0.01
05.09.27	12h00	7.30	5.71	7.39	1.59	0.09
05.10.20	12h00	9.01	7.02	9.00	1.99	0.01
05.12.21	12h00	3.79	5.41	5.82	1.62	2.03

measured value) than SABIM.

Table 5.11 gives the NmF2 values during the winter season, and from this table we can see that MIDAS produces more accurate values (rms error of 0.37×10^{11} $[e/m^3]$) 6 out of a total of 7 opportunities. Thus, it is deduce that during the winter season there is an approximately 86% chance that MIDAS will produce a NmF2 value closer to the measured NmF2 value than SABIM.

From table 5.12, 4 out of 6 NmF2 values produced by MIDAS are closer than SABIM to the measured ionosonde NmF2 values ('close' being defined to be within 0.52×10^{11} $[e/m^3]$). From this we can assume that during the spring season, there is an approximately 67% chance of MIDAS producing NmF2 values that are at least as accurate as our national model.

Table 5.9: The NmF2 $[10^{-11}e/m^3]$ values from the ionosonde, SABIM and MIDAS, and the corresponding absolute errors of SABIM and MIDAS for the afternoon to evening hours.

Date	**Time**	**Ionosonde**	**SABIM**	**MIDAS**	$SABIM_{error}$	$MIDAS_{error}$
05.04.02	20h00	1.25	1.24	2.20	0.01	0.95
05.05.12	14h00	6.52	5.08	9.59	1.44	3.07
05.06.15	18h00	0.82	0.94	1.46	0.12	0.64
05.08.23	18h00	1.65	0.96	2.34	0.69	0.69
05.09.27	20h00	1.33	1.08	2.08	0.25	0.75
05.10.20	18h00	3.35	2.55	4.43	0.80	1.08
05.12.21	14h00	3.52	4.55	5.16	1.03	1.64

Table 5.10: The NmF2 $[10^{-11}e/m^3]$ values from the ionosonde, SABIM and MIDAS, and the corresponding absolute errors of SABIM and MIDAS for the autumn season.

Date	**Time**	**Ionosonde**	**SABIM**	**MIDAS**	$SABIM_{error}$	$MIDAS_{error}$
05.04.02	06h00	5.09	4.15	5.83	0.94	0.74
	12h00	8.36	7.52	8.31	0.84	0.05
	20h00	1.25	1.24	2.20	0.01	0.95
05.05.12	02h00	1.02	1.12	1.25	0.10	0.23
	12h00	11.25	5.94	11.29	5.31	0.04
	14h00	6.52	5.08	9.59	1.44	3.07

It should be noted that the TEC values used to obtain the electron densities were not corrected for any biases and therefore this limits the accuracy of the reconstructed profiles. It should also be noted that the percentages were inferred by assuming that the sample is random and the distribution of the sample is Gaussian. These assumptions are taken at face value and were not tested. In summary, this chapter presented the profiles obtained from MIDAS, and the statistical analyses based on these profiles. In the following chapter, a conclusion will be reached based on the analyses and discussions presented here.

Table 5.11: The NmF2 $[10^{-11}e/m^3]$ values from the ionosonde, SABIM and MIDAS, and the corresponding absolute relative errors of SABIM and MIDAS for the winter season.

Date	**Time**	**Ionosonde**	**SABIM**	**MIDAS**	$SABIM_{error}$	$MIDAS_{error}$
05.06.15	07h00	3.48	5.24	3.74	1.76	0.26
	12h00	9.12	4.46	9.10	4.66	0.02
	18h00	0.82	0.94	1.46	0.12	0.64
05.07.16	12h00	5.04	3.38	5.02	1.66	0.02
05.08.23	10h00	5.36	3.96	5.40	1.40	0.04
	12h00	5.44	4.03	5.43	1.41	0.01
	18h00	1.65	0.96	2.34	0.69	0.69

Table 5.12: The NmF2 $[10^{-11}e/m^3]$ values from the ionosonde, SABIM and MIDAS, and the corresponding absolute relative errors of SABIM and MIDAS for the spring season.

Date	**Time**	**Ionosonde**	**SABIM**	**MIDAS**	$SABIM_{error}$	$MIDAS_{error}$
05.09.27	10h00	8.09	5.76	8.14	2.33	0.05
	12h00	7.30	5.71	7.39	1.59	0.09
	20h00	1.33	1.08	2.08	0.25	0.75
05.10.20	06h00	3.60	3.85	3.77	0.25	0.17
	12h00	9.01	7.02	9.00	1.99	0.01
	18h00	3.35	2.55	4.43	0.80	1.08
05.12.21	10h00	3.04	5.37	4.95	2.33	1.91
	12h00	3.79	5.41	5.82	1.62	2.03
	14h00	3.52	4.55	5.16	1.03	1.64

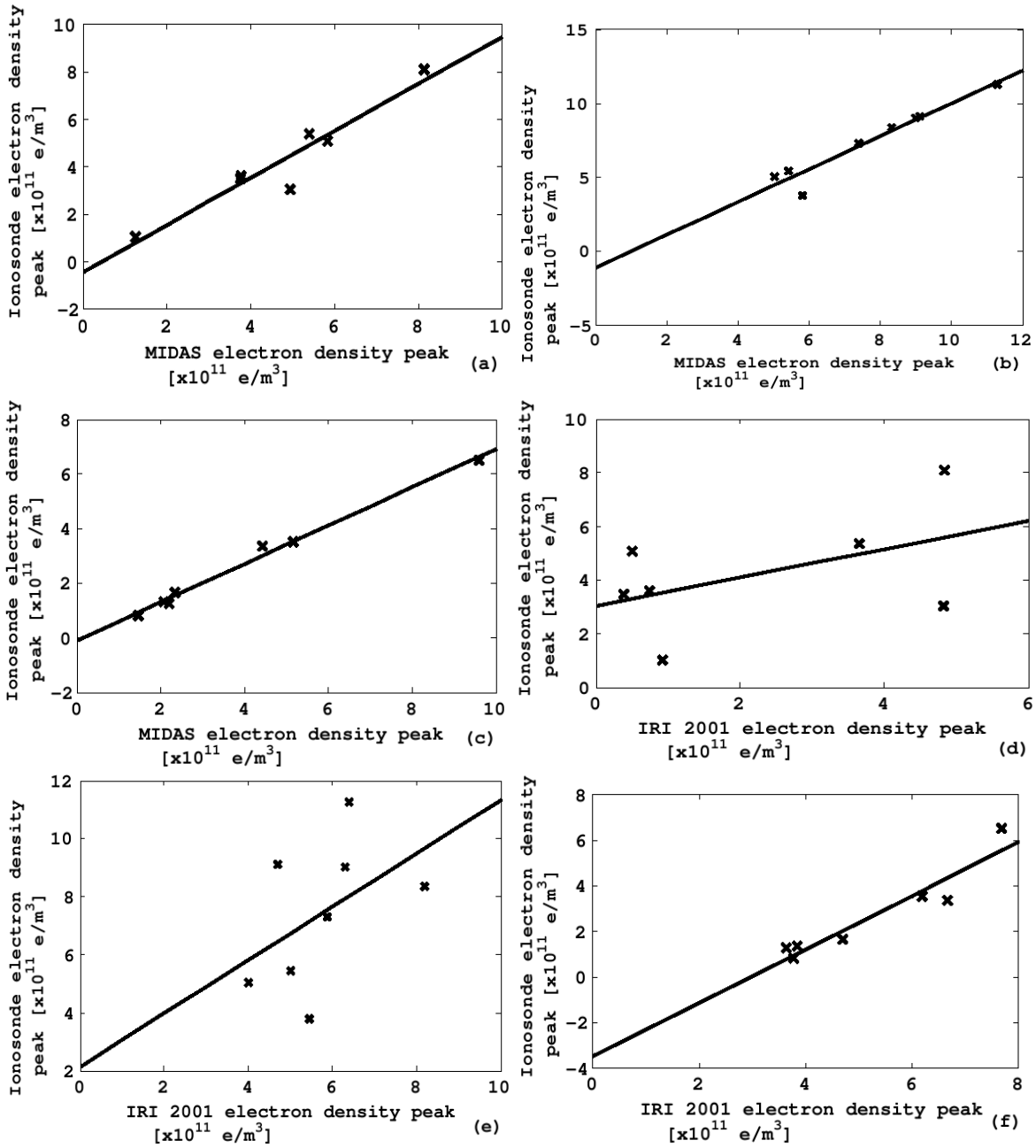

Figure 5.13: The plot of the peak electron density values obtained from MIDAS and the IRI compared with the Grahamstown ionosonde peaks at different times of the day: (a) MIDAS peak values in the morning hours, (b) MIDAS peak values at noon and (c) MIDAS peak values in the afternoon/evening hours, (d) IRI peak values in the morning hours, (e) IRI peak values at noon, and (f) IRI peak values in the afternoon/evening hours. Statistical results are given in table 5.4 and table 5.5

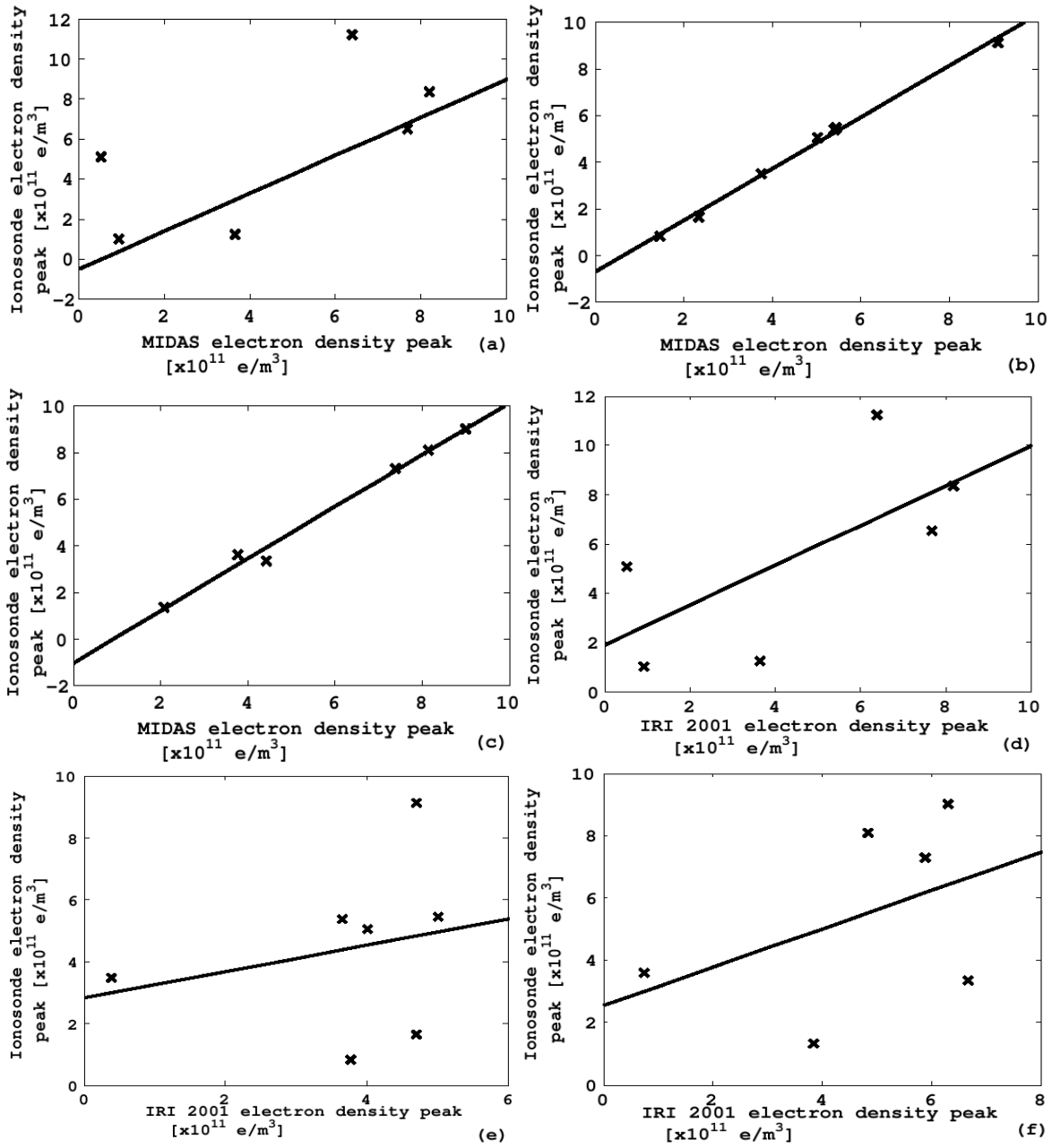

Figure 5.14: The plot of the peak electron density values obtained from MIDAS and the IRI compared with the Grahamstown ionosonde peaks in different seasons: (a) autumn MIDAS peak values, (b) winter MIDAS peak values (c) spring MIDAS peak values, (d) autumn IRI peak values, (e) winter IRI peak values, and (f) spring IRI peak values. Statistical results are given in table 5.4 and table 5.5

Chapter 6

Discussions and Conclusion

The aim of this project was to determine the suitability of ionospheric tomography as a tool for determining electron density profiles from GPS data over South Africa. Therefore, the performance of the MIDAS algorithm in producing the electron density profiles was evaluated by comparing the reconstructed profiles with the ionosonde measured profiles. In order to achieve this, four methods were used to reconstruct the profiles:

- the Chapman profile to create the EOFs,
- the IRI 2001 model to create the EOFS,
- ionosonde peak parameters to supplement GSP data and the Chapman profile to create EOFs
- ionosonde profiles from the Madimbo and/or Louisvale ionosonde stations to create the EOFs.

The method of using ionosonde profiles to measure the vertical electron density produced realistic profiles, and thus this method was used to create all the EOFs that were used to assess MIDAS.

The evaluation of the ability of MIDAS to reconstruct profiles that are close to the measured profile was done by determining how well MIDAS computes the NmF2 values and comparing these MIDAS values with the measured ones from the Grahamstown ionosonde. In order to to test the accuracy of the computed NmF2 value, it was compared to the NmF2 values predicted by the national model (SABIM).

SABIM is currently the best model that describes the bottomside ionosphere over Grahamstown. If the deviation of the MIDAS produced NmF2 value from the measured is less than that of the national model then MIDAS is considered to have computed an accurate NmF2 value.

From the figures and tables presented in section 5.2, it is evident that MIDAS produces accurate results during the winter season, with the lowest rms error of 0.37×10^{11} $[e/m^3]$ and an approximately 86% chance of producing NmF2 closer to the actual NmF2 value than the national model SABIM. At 12h00 UT MIDAS has an approximately 88% chance of producing an accurate NmF2 value which will deviate from the measured from by 0.72×10^{11} $[e/m^3]$. The worst case scenerio is during the afternoon-evening hours, where the chance of MIDAS producing a NmF2 value that is within 1.49×10^{11} $[e/m^3]$ of the measured ionosonde NmF2 value, is about 14%. Over all, for the whole year, MIDAS is expected to produce NmF2 values within 1.05×10^{11} $[e/m^3]$ of the measured values with a success of about 59%. The TEC values were not corrected for the receiver and satellite biases, and these reduced the accuracy of MIDAS. In conclusion, the agreement of the MIDAS computed profiles and NmF2 values to the ionosonde profiles and NmF2 values makes MIDAS a promising inversion technique for use at Grahamstown.

6.1 Future Work

MIDAS was evaluated using a small sample of data, therefore it would be useful to expand on the analysis by using much more data and evaluating the consistency of the accuracy of the MIDAS computations. The dates and hours chosen for the reconstructions in this project coincided with the period of solar minimum and quiet magnetic times. Thus it will be interesting to apply the MIDAS algorithm at solar maximum and during magnetic storms to fully assess its performance under all conditions. The intention is that this study will lead to the eventual implementation of a near real-time ionospheric tomography system within South Africa. The evaluation of MIDAS under all conditions will help in investigating the feasibility of the implementation of MIDAS as a near real-time system.

References

NAVSTAR GPS USER EQUIPMENT INTRODUCTION, URL `http://www.navcen.uscg.gov/pubs/gps/gpsuser/gpsuser.pdf`, 1996.

Austen J.R., Franke S.J., and Liu C.H., "Ionospheric imaging using computerised tomography", *Radio Science*, **23**(3), pp. 299–307, 1988.

Bazin S., van Avendonk H., Harding A.J., Orcutt J.A., Canales J.P., and Detrick R.S., "Crustal structure of the flanks of the East Pacific Rise: Implications for overlapping spreading centers", *Geophysical Research Letters*, **25**(12), pp. 2213–2216, 1998.

Bilitza D., "International Reference Ionosphere 1990", Goddard Space Flight Center, URL `http://modelweb/gsfc.nasa.gov/ionos/iri.html`, 1990.

Bilitza D., "International Reference Ionosphere 2000", *Radio Science*, **36**(2), pp. 261–275, 2001.

Cilliers P.J., Opperman B.D.L., Mitchell C.N., and Spencer P.J., "Electron density profiles determined from tomographic reconstruction of total electron content obtained from GPS dual frequency data: first results from the South African network of dual frequency GPS receiver stations", *Advances in Space Research*, **34**(9), pp. 2049–2055, 2004.

Dana P.H., "Global Positioning System Overview", URL `http://www.colorado.edu/geography/gcraft/notes/gps/gps_f.html`, 1994.

Davies K., *Ionospheric Radio*, volume 31 of *IEE Electromagnetic Waves Series*, Peter Peregrinus, London, 1990, ISBN 086341186X.

Donnan G.A., Tress B.M., and Bladin P.F., "A prospective study of lacunar infarction using computerised tomography", *Neurology*, **32**(1), pp. 49–56, 1982.

Fremouw E.J., Secan J.A., and Howe B.M., "Application of stochastic inverse theory to ionospheric tomography", *Radio Science*, **27**(5), pp. 721–732, 1992.

Fridman S.V., Nickisch L.J., Aiello M., and Hausman M., "Real-time reconstruction of the three-dimensional ionosphere using data from a network of GPS receivers", *Radio Science*, **41**(5), doi:10.1029/2005RS003341, 2006.

Galkin I., "Background to ionospheric sounding", URL `http://ulcar.uml.edu/digisonde_dps.html`, 2007.

Hajj G.A. and Romans L.J., "Ionospheric electron density profiles obtained with the Global Positioning System: Results from the GPS/MET experiment", *Radio Science*, **33**(1), pp. 175–190, 1998.

Hofmann-Wellenhof B., Lichtenegger H., and Collins J., *GPS - Theory and Practice*, Springer-Verlag, Wien, New York, 1997, 4th edition, ISBN 3211828397.

Honary F., "The Ionosphere: an Introduction", URL `http://www.dcs.lancs.ac.uk/iono/ionosphere_intro/`, 2007.

Jin S., Park J.U., Wang J.L., Choi B.K., and Park P.H., "Electron Density Profiles Derived From Ground-Based GPS Observations", *Journal of Navigation*, **59**(3), pp. 395–401, doi:10.1017/S0373463306003821, 2006.

Kelso J.M., *Radio ray propagation in the ionosphere*, McGraw-Hill electronic science series, McGraw-Hill Book Company, 1964.

Kunitsyn V.E. and Tereshchenko E.D., *Ionospheric Tomography*, Physics of Earth and Space environment, Springer-Verlag, Berlin, 2003, ISBN 35400040411.

Larson K., Bilich A., and Axelrad P., "Improving the Precision of High-Rate GPS", *Journal of Geophysical Research*, **112**(B5), doi:10.1029/2006JB004367, 2007.

Leitinger R., *Ionospheric Tomography*, Review of Radio Science 1996 - 1999, Oxford University Press, New York, 1999, chapter 24, pp. 581–623, ISBN 0198565712.

Mannucci A.J., Iijima B.A., Lindqwister U.J., Pi X., Sparks L., and Wilson B.D., *GPS and Ionosphere*, Review of Radio Science 1996 - 1999, Oxford University Press, New York, 1999, chapter 25, pp. 625 – 665, ISBN 0 19 856571 2.

McKinnell L.A., *A Neural Network Based Ionospheric Model for the Bottomside Electron Density Profile over Grahamstown, South Africa*, Ph.D. thesis, Rhodes University, Grahamstown, 2002.

McNamara L.F., *The Ionosphere: Communications, Surveillance, and Direction Finding*, Krieger Publishing Company, 1991, ISBN 0894640402.

Meggs R.W., *Mapping of Ionospheric Total Electron Content Using Global Navigation Satellite Systems*, Ph.D. thesis, University of Bath, Bath, 2005.

Mitchell C.N., "Ionospheric and Imaging Research: Ionospheric Tomography", URL `http://people.bath.ac.uk/eescnm/Tomography.htm`, 2005.

Mitchell C.N., Kersley L., Heaton J.A.T., and Pryse S.E., "Determination of the vertical electron-density profile in ionospheric tomography: experimental results", *Annales Geophysicae*, **15**(6), pp. 747–752, 1997.

Mitchell C.N. and Spencer P.S., "A three-dimensional time-dependent algorithm for ionospheric imaging using GPS", *Annals of Geophysics*, **46**(4), pp. 687–697, 2003.

Raymund T.D., Bresler Y., Anderson D.N., and Daniell R.E., "Model-assisted ionospheric tomography: A new algorithm", *Radio Science*, **29**(6), pp. 1493–1512, 1994.

Reilly M.H. and Singh M., "Electron density height profiles from GPS receiver data", *Radio Science*, **39**(1), doi:10.1029/2002RS002830, 2004.

Rishbeth H. and Garriott O.K., *Introduction to Ionospheric Physics*, volume 14, Academic Press, New York, 1969.

Spencer P., "Multi-instrument Data Analysis Software (MIDAS)", Technical report, Department of Electronic & Electrical Engineering, University of Bath, Bath, 2000.

Spencer P., "MIDAS - Multi-instrument data analysis software", Technical report, Department of Electronic & Electrical Engineering, University of Bath, Bath, 2001.

Spencer P., "MIDAS 1.2", Technical report, Department of Electronic & Electrical Engineering, University of Bath, Bath, 2002.

Spencer P.S.J., Kersley L., and Pryse S.E., "A new solution to the problem of ionospheric tomography using quadratic programming", *Radio Science*, **33**(3), pp. 607–616, 1998.

Stankov S.M., Jakowski N., and Heise S., "A new method for reconstruction of the vertical electron density distribution in the upper ionosphere and plasmasphere", *Journal of Geophysical Research*, **108**(A5), doi:10.1029/2002JA009570, 2003.

Stolle C., Schluter S., Jacobi C., and Jakowski N., "3-Dimensional ionospheric electron density reconstruction based on GPS measurements", *Advances in Space Research*, **31**(8), pp. 1965–1970, 2003.

Vasicek C.J. and Kronschnabl G.R., "Ionospheric tomography: an algorithm enhancement", *Journal of Atmospheric and Terrestrial Physics*, **57**(8), pp. 875–888, 1995.