TOMOGRAPHIC IMAGING OF IONOSPHERIC ELECTRON DENSITY OVER ETHIOPIA USING GROUND BASED GPS RECEIVERS

By

Gebreab Kidanu Zewdie

PRESENTED IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE IN PHYSICS
(SPACE PHYSICS)

ADDIS ABABA UNIVERSITY
ADDIS ABABA, ETHIOPIA
JUNE 2012

ADDIS ABABA UNIVERSITY
DEPARTMENT OF
PHYSICS

Supervisor: _____
Dr. Gizaw Mengistu Tsidu

Examiners: _____
Dr. Elias Lewi

_____
Dr. Lemi Demeyu

ADDIS ABABA UNIVERSITY

Date: **June 2012**

Author: **Gebreab Kidanu Zewdie**

Title: **TOMOGRAPHIC IMAGING OF IONOSPHERIC ELECTRON DENSITY OVER ETHIOPIA USING GROUND BASED GPS RECEIVERS**

Department: **Physics**

Degree: **M.Sc.** Convocation: **June** Year: **2012**

Signature of Author

To the people who made significant contributions in developing my knowledge from a villege to the outer space: my parents, my teachers at all levels and few close friends.

Table of Contents

List of Figures

Acronyms

ART	Algebraic Reconstruction Technique
CG	Conjugate Gradient
CGLS	Conjugate Gradient Least Square
CIO	Conventional International Origin
CIT	Computerized Ionospheric Tomography
CODE	Center for Orbit Determination in Europe
CT	Computerized Tomography
CTRS	Conventional Terrestrial Reference System
DLS	Damped Least Square
ECEF	Earth Centered Earth Fixed
EUV	Extreme Ultra-Violet
FTP	File Transfer Protocol
GAMIT	GPS Analysis at MIT
GCV	Generalized Cross Validation
GLONASS	GLObal NAvigation Satellite System
GNSS	Global Navigation Satellite System
GPS	Global Positioning System
GPST	GPS Time
GSVD	Generalized Singular Value Decomposition
IART	Improved Algebraic Reconstruction Technique
IEC	Integrated Electron Content
IED	Ionospheric Electron Density

IERS	International Earth Rotation Service
IFB	Inter Frequency Bias
IGS	International GNSS Service
IGSSA	Institute of Geophysics Space Science and Astronomy
IONEX	IONosphere map EXchange
IRI2007	International Reference Ionosphere
ITRF	International Terrestrial Reference Frame
ITRS	International Terrestrial Reference System
LEO	Low Earth Orbit
MART	Modified Algebraic Reconstruction Technique
MATLAB	MATrix LABoratory
MEM	Maximum Entropy Method
MHD	Magneto Hydro-Dynamics
MIT	Massachusetts Institute of Technology
PRN	Pseudo-Random Noise
RCM	Residual Correction Method
RINEX	Receiver INdependent EXchange
SIRT	Simultaneous Iterative Reconstruction Technique
SOPAC	Scripps Orbit and Permanent Array Center
SP3	Special Products 3
STEC	Slant Total Electron Content
SVD	Singular Value Decomposition
TAI	International Atomic Time
TDB	Barycentric Dynamic Time
TDT	Terrestrial Dynamic Time
TEC	Total Electron Content
TGD	Transmitter Group Delay
TGSVD	Truncated Generalized Singular Decomposition
TSVD	Truncated Singular Value Decomposition
UNAVCO	University NAVSTAR Consortium

UT	Universal Time
UTC	Coordinated Universal Time
UV	Ultra-Violet
VTEC	Vertical Total Electron Content
WGS84	World Geodetic System 1984

Abstract

The Total Electron Content (TEC) from a spatially distributed network of Global Positioning System (GPS) receivers can show the temporal, seasonal, latitudinal and longitudinal variations of the ionospheric electron distribution. However, it does not convey any information about ionospheric electron density structure along the vertical direction. Computerized Ionospheric Tomography (CIT) is a method to investigate vertical ionospheric electron density profile in two or three dimensions. In this work, ground based GPS receivers over Ethiopia have been used to record signals transmitted from GPS satellites and the computed pseudoranges and phase values are used to calculate Total Electron Content (TEC). In two dimensions (2D), the vertical plane along 40.0261^0E longitude is discretized into pixels and the distance traveled in each pixel crossed by GPS signals to four receivers located close to the vertical plane is calculated in a selected reference frame. Using TEC from these stations a 2D, altitude versus latitude, distribution of electron density is inverted with in each pixel. In three dimensions (3D), the volumetric ionosphere over Ethiopia is discretized in to numerous small voxels and the distance traveled by GPS signals in each voxel is calculated from the precise position of the GPS satellites to find the coefficient matrix in an appropriate reference frame. A 3D ionospheric electron density distribution is then obtained using TEC data from GPS stations located all over the country. In both 2D and 3D tomographic reconstructions damped least square inversion algorithm with second order Tikhonov regularization has been employed. We used the L-curve method to determine the regularization parameter. The performance of the inversion algorithm has been numerically validated with a simulated TEC from IRI2007 model. The numerical experiments show that the algorithm produced results with 2-norm relative reconstruction errors 0.04 and 0.02 for the 2D and 3D inversions respectively.

Acknowledgements

I would like to express my heartfelt gratitude to my advisor Dr. Gizaw Mengistu for believing in me and directing to work in the area of ionospheric tomography. I deeply appreciate his wise guidance and comments, proper follow up accompanied by the necessary ignition and valuable suggestions during difficult times. I thank you very much for your support to give me an opportunity to participate and obtain important experiences in international space science workshops and summer schools.

I am very grateful to Prof. Eric Calais and Prof. Thomas Dauterman from Purdue University for providing the TEC data processing software; Prof. Per Christian Hansen from Technical University of Denmark for freely publishing online the MATLAB Regularization Tools. I would like to thank the Center for Orbit Determination in Europe (CODE), University of Bern for making ionospheric exchange data freely available on the web.

I would also thank Dr. Elias Lewi and his student Elias Abebe from the Institute of Geophysics Space Science and Astronomy (IGSSA), Addis Ababa University for providing GPS observation, satellite position data and other Ephemerids files which were difficult to download with our slow internet connection.

Partial funding for this research was provided by the Department of Physics of Addis Ababa University. Additionally, the head of the Department of Physics Dr. Lemi Demeyu and department secretary Ms. Tsilat played a great role by writing supportive official letters which were important to obtain financial support and visa for my participation in space weather related summer schools and workshops. Their help and cooperation is gratefully acknowledged.

Finally, I am happy to thank Space and Atmospheric Physics students of our laboratory for creating a working atmosphere; and a few close friends outside the University for their support and encouragement.

Chapter 1

Introduction

The ionosphere is the region of partially ionized plasma above the Earth's atmosphere formed due to primarily photo-ionization of the neutral atoms and molecules. The increase in the density of the atmosphere as we go further down to the surface of the Earth and the decrease in the intensity of photons as it bombards the neutral atoms and the availability of different atoms and molecules at different heights from the Earth's surface forms a large scale vertical layer of ionization with in the ionosphere. This makes the ionosphere to exist as vertically stratified partially ionized plasma state from at about 60 $km - 1000$ km from the surface of the Earth [1, 2].

The ionosphere affects the propagation of electromagnetic waves passing through it [3]. As radio waves penetrate the ionosphere their refractive index is a function of the waves frequency, the electron density and to a small extent the Earth's magnetic field. This means the ionosphere is a dispersive medium for the propagation of electromagnetic waves.

The space segment of the Global Positioning System (GPS) consists of about 31 satellites in a circular orbit at an approximate height of $20,000$ km above the Earth and their signal must travel through the Earth's ionosphere on their way to GPS receivers on or near the Earth's surface. Hence the ionosphere imposes a dispersive error on GPS signals. The error is that the GPS carrier phases are advanced by the same amount as the

wave group is delayed. In other words the error is negative for carrier phases (i.e measured satellite to receiver range is shorter than geometric range) and positive for the wave group (the measured satellite to receiver range is longer than the geometric range). To the first order approximation the ionospheric error (the group delay and phase advance) is proportional to the integral of the electron density along the GPS signal path and inversely proportional to the square of the carrier phase frequency. We define this integral of the electron density along signal path the Total Electron Content (TEC).

A linear combination of the pseudoranges and carrier phase measurements on the two GPS frequencies can be used to remove the ionospheric effect for positioning on Earth. On the other hand a linear combination of these observables gives an opportunity to directly measure the ionospheric total electron content. The measured TEC provides the potential for investigating the temporal and spatial variations of the ionosphere under normal or geomagnetic storm environments. However the TEC, being the measurement along the line of sight does not convey any information about the ionosphere vertical electron profile. To get the vertical structure of the ionosphere the integral measurement, TEC, should be inverted for the electron density distribution. Computerized Tomography (CT), which is popular in medicine in x-ray imaging of a section of a human body, is a method recently applied to reconstruct the ionosphere electron distribution in full space. The method of applying CT to the ionosphere is said to be Computerized Ionospheric Tomography (CIT) and has become a powerful tool of studying the ionosphere [4].

The aim of this study is to image ionospheric electron density distribution over Ethiopia by the tomography method using TEC measurements from GPS receivers spatially distributed over the country. TEC calculated to the highest accuracy possible for a large number of satellite to receiver pairs has been inverted to reconstruct tomographyically ionospheric electron density distribution. In the inversion process experimental set up has been designed in selected reference frames for algorithm implementation in both 2D and 3D.

This thesis work is structured in seven chapters. A brief introduction of the thesis is given in chapter one. Chapter two is about the Earth's ionosphere with discussions on the formation, layers, geographic regions of the Earth's ionosphere and the effect of the ionospheric plasma on the propagation of electromagnetic waves. Chapter three deals with the principles, and surveying techniques of the Global Positioning System (GPS) with a special emphasis on the applications of GPS for ionospheric modeling. In the fourth chapter detailed discussion of the inverse theory and regularization methods is presented. The fifth chapter briefly explains the tomography method. Chapter six provides TEC data processing and analysis of results of tomographic inversion in two and three dimensions. Finally, conclusion of the thesis is present in the seventh chapter.

Chapter 2

THE IONOSPHERE

2.1 Introduction

Balfour Stewart was the first person to postulate an ionosphere when, in 1882, he attributed currents in the upper atmosphere as the probable origin of the electric currents that produced the solar controlled variation in the Earth's magnetic field measured at the Earth's surface [2]. According to Michael C. Kelley (2009) [1], the discovery of the ionosphere came from radio wave observations and the recognition that only a reflecting layer composed of electrons and positive ions could explain the characteristics of the data. The term ionosphere was first used by Sir Robert Watson-Watt in a letter to the secretary of British Radio Research Board in 1926 [5]. The expression came into wide use during the period 1932-34 when Watson-Wat, Appleton, Radcliff and others used it in papers and books [5]. Today, radio science remains the primary tool for obtaining measurements of the ionosphere. Using short vertically directed radio frequency pulses and inferring the height of electrically reflecting layer from timing the arrival of the reflecting signal is still used in sounding the ionosphere [2].

2.2 The Ionosphere of the Earth

The Earth's ionosphere, that extends at an altitude range of approximately $60 - 2000\ km$, is historically the region of the atmosphere that affects the propagation of radio waves

[6]. One may also define the Earth's ionosphere as the region of the Earth's upper atmosphere partially ionized due to primarily ultraviolet radiation from the Sun. Although the ionosphere becomes sensible at about $50km$ from the the surface of the Earth, its upper boundary is not well defined since it can be interpreted as the electron densities thinning into the plasmasphere (or protonosphere) and subsequently to the interplanetary plasma [5]. But the most important contribution for the ionosphere lies in the $90 - 1000\ km$ region from the surface of the Earth [6]. Most planets in the solar system do have their own ionosphere provided they have their own neutral gas-envelope (atmosphere) [1], but when we say ionosphere in this thesis we mean the ionosphere of the Earth.

2.3 Ionosphere Formation

The ionosphere is a by-product of the interaction between the Sun and the Earth environment. In fact, the ionosphere emerges from the shielding effect exerted by the neutral atmosphere against the penetration of particles coming from outside. These particles can be either photons coming from the Sun, cosmic rays coming from the interplanetary medium, or other particles from the solar wind or the magnetosphere. As the neutral atmosphere shields these particles, it becomes ionized provided that their energies ($h\nu$ in the case of photons, and kinetic energy in the case of other particles) exceeds the ionization potential or binding energy of the neutral atmosphere atomic or molecular electron. Although, in nature atmospheric ionization and hence ionospheric formation is usually the result of these various sources, solar photons in the extreme ultraviolet (EUV) and ultraviolet (UV) wavelength range of approximately 10 nm to 100 nm dominate [2]. In daytime, the extreme ultraviolet (EUV) radiation in the solar spectrum is incident on the neutral atmosphere that is increasing in density exponentially with decreasing altitude. Since the photons are absorbed in the process of photoionization, the beam itself decreases in intensity as it penetrates. Fig. 2.1 shows the different ions formed due to ionization of the neutral atoms or molecules at different heights with in the ionosphere.

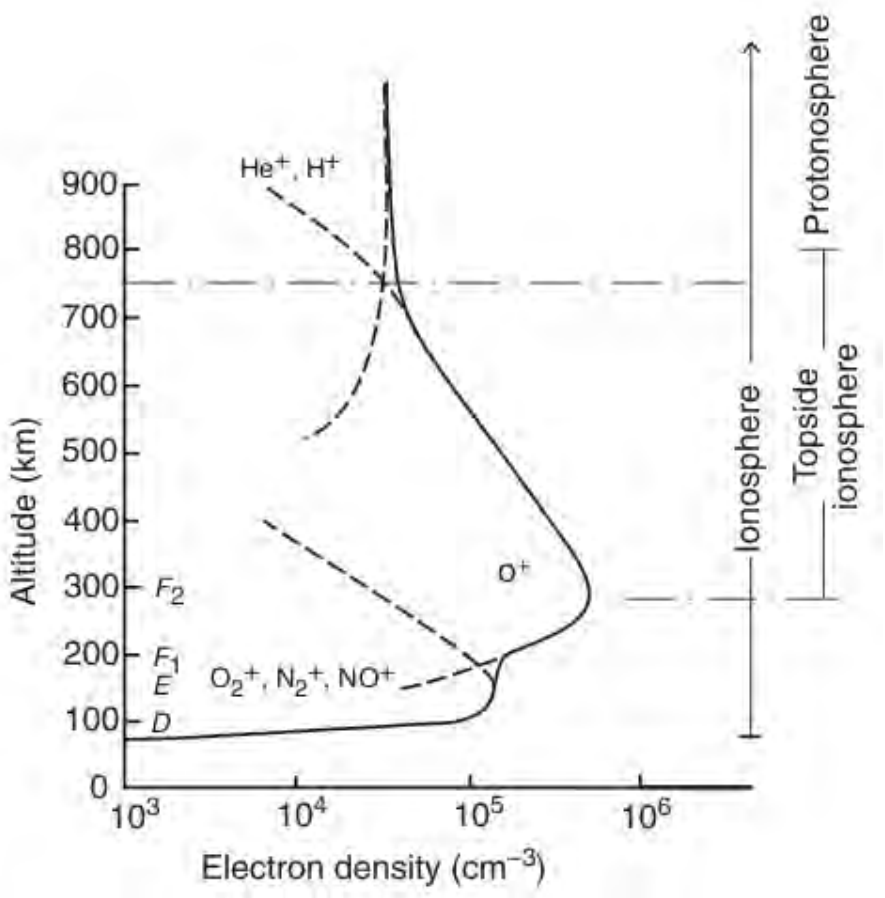

Figure 2.1: Ionospheric ion density profile for the daytime mid-latitude ionosphere showing the layered structure (from [10]).

The combination of decreasing beam intensity and increasing neutral density with the availability of different atoms and molecules at different heights from the Earth's surface provides an explanation for the basic large scale vertical layer of ionization with in the ionosphere as shown in Figs. 2.1 and 2.2. The ionization process, and hence the peak plasma density in each layer is limited by a factor known as recombination rate. Recombination is the process in which ionospheric electrons disappear by combining with ions to form a neutral molecule. Recombination can be radiative ($e + \mathrm{X}^+ \rightarrow \mathrm{X} + h\nu$), dissociative ($e + \mathrm{XY}^+ \rightarrow \mathrm{X} + \mathrm{Y}$), and attachment ($e + \mathrm{X} \rightarrow \mathrm{X}^-$). Radiative and dissociative recombinations are most important with in the bulk of the ionosphere. The local concentration

of ions and electrons with in the ionosphere determines the recombination rate with in each layer.

2.4 Layers of the Ionosphere

As described in Section 2.3, the decrease in light intensity and increase in neutral density with decrease in altitude forms an ionization peak in the ionosphere. However, due to the availability of different atoms and molecules with differing rates of absorption, the ionosphere at all latitudes has a tendency to separate into layers. However, a series of distinct regions or layers of electron density exist in the daytime ionosphere at mid-latitudes [1, 5]. These layers are denoted by the letters D, E, F1, and F2 (Fig. 2.2). Each layer is generally characterized by a density maximum at a certain altitude and a density decrease with altitude on both sides of the maximum. The four layers are collectively known as bottomside ionosphere. Peak electron density of the ionosphere occurs usually at the F2 layer and the region above this layer is said to be topside ionosphere.

There is a minimum frequency of an electromagnetic wave that can penetrate an ionospheric layer. This frequency is called the critical or plasma penetration frequency of the particular layer and it is denoted by foD, foE, foF1, and foF2 according to the designation of the ionospheric layers. The square of a critical frequency is linearly proportional to the maximum electron density of the individual layer and these are denoted by NmD, NmE, NmF1 and NmF2 respectively.

2.4.1 D layer

The D layer is the innermost layer ranging from 60 km to 90 km from the surface of the Earth. The primary source of ionization in this layer is cosmic radiation which is the same by day and by night manifesting itself in a strong solar cycle variation in the D layer electron density. By night, the electrons become attached to atoms and molecules forming negative ions that cause the D layer to diminish in size but some remains due

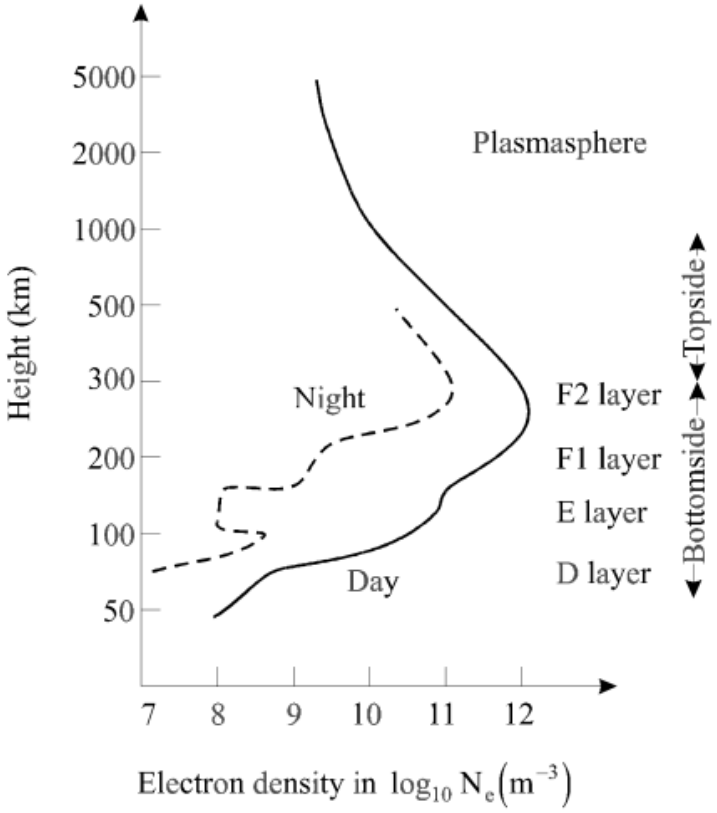

Figure 2.2: Vertical profiles of ionospheric electron density showing the distinct layers (from [5]).

to galactic cosmic rays. During day time due to Sun's radiation, the electrons tend to detach themselves from the ions causing the D layer electrons to re-appear. Hence, the D layer electrons are present by day but diminish highly by night causing a distinct diurnal variation in the electron density. The typical values for the noon time electron densities of the D layer at the mid-latitude region range between 6.1×10^8 to 13.1×10^8 m^{-3}.

2.4.2 E layer

The E layer is the middle layer, 90 km to 120 km above the surface of the Earth and it is the first layer to be detected, followed by the F and D layers. The behavior of the E layer almost entirely depends on the level of solar activity and the zenith angle of the Sun. The primary source of ionization is the Sun's X-ray emissions resulting in electron densities

showing distinct solar-cycle, seasonal and daily variations. The E-layer peaks at 110 km above the Earth with noon peak electron density values of 1.5×10^{11} electrons/m^3. The E layer does not completely vanish at night due to electron and meteor bombardment. It has night time peak electron density of 1×10^{10} electrons/ m^3. However, for practical purposes it is often assumed that its electron density drops to zero at night. The E layer is free of disturbances unlike the D and F layers.

2.4.3 F1 layer

The F1 layer peaks at 200 km above the surface of the Earth with noon peak electron density of 2.5×10^{11} electrons/m^3. The main source of ionization in the F1 layer is the EUV light. The F1 layer is only observed during the day and decays by night since the electron densities are primarily controlled by the zenith angle of the Sun. When it is present, it changes rapidly in a matter of minutes. It is more pronounced during the summer than during the winter months for low solar Sunspot numbers and for periods with ionospheric storms.

2.4.4 F2 layer

The ionosphere's highest electron density values are present in the F2-layer with a value of about 10^{12} at mid latitude at around 300 km. The F2 layer's height and electron density is highly variable; large daily, seasonal and Sunspot cycle variations are combined with general erratic behavior. Ionization is due to extreme ultraviolet (EUV) solar radiation of atomic oxygen (O).

2.5 Geographic Regions of the Ionosphere

As explained earlier, solar radiation dominates in the formation of the ionosphere. Consequently, the plasma parameters (the electron densities in particular) in the Earth's ionosphere display a marked variation with latitude, longitude, altitude, universal time,

season, solar cycle, and magnetic activity. This variation results not only from the coupling, time delays, and feedback mechanisms that operate in the ionosphere-thermosphere system, but also from the ionosphere's coupling to the other regions in the solar-terrestrial system, including the Sun, the interplanetary medium, the magnetosphere, and the mesosphere [1]. With respect to latitude the global ionosphere has three major regions. These are the high-latitude, mid-latitude and equatorial regions as shown in Fig. 2.3. In this section we will see the peculiar characteristics of each region.

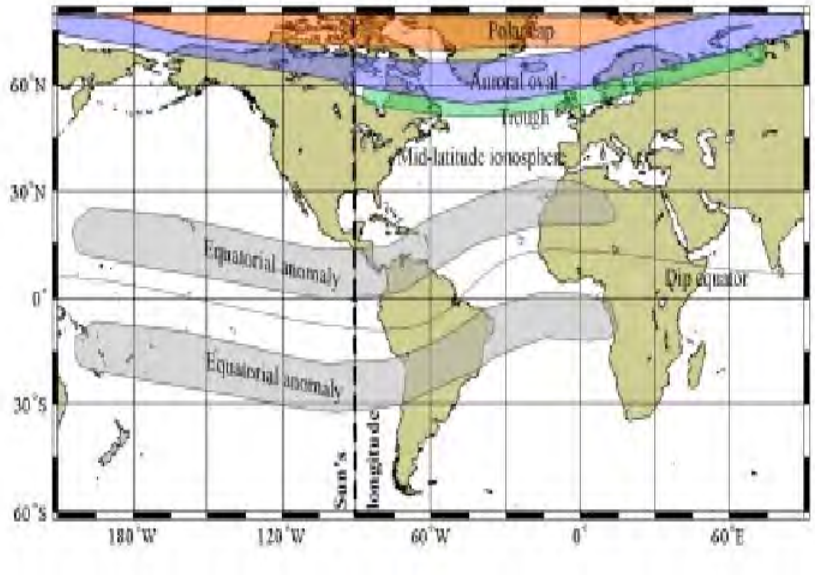

Figure 2.3: Major geographic regions of the ionosphere (from Attila Komjathy, 1997).

2.5.1 Low Latitude Region

This region of the ionosphere is also called the equatorial region and is characterized with the highest values of peak-electron density with the most pronounced amplitude and phase scintillation effects. It usually occurs in the region between 20^0S and 20^0N in geomagnetic latitude. The Equatorial region is strongly influenced by electromagnetic forces that arise because the geomagnetic field runs horizontally over the magnetic equator. Consequently, this region is characterized by effects like equatorial electrojet which is a narrow ribbon current flowing eastward in the daytime region of Earth's ionosphere. Another effect is

fountain effect formed as a consequence of $E \times B$ upward plasma drifts associated with an eastward electric field E and a northward horizontal magnetic field B. The lifted plasma then diffuses downward along the geomagnetic field lines due to the gravitational force and the plasma pressure gradient, and this results in ionization enhancements on both sides of the magnetic equator. The electrons move as far as the geomagnetic latitudes of 10 to 20 degrees on both sides of the geomagnetic equator causing the high concentration of electrons there which are often termed equatorial anomalies, also often called the Appleton anomaly [5, 7].

2.5.2 Middle Latitude Region

The middle-latitude ionosphere is known to be the best understood ionospheric region, largely due to relatively simple physics and reasonably good coverage of ionosphere sensing instruments [7]. It is usually free of the effect imposed by the horizontal geomagnetic field geometry peculiar to the equatorial region.

2.5.3 High Latitude Region

In addition to photoionization, collisional ionization is another source of ionization in the high latitude region. This is because geomagnetic field lines are nearly vertical in high latitude region leading to the charged particles descending to E layer altitudes (about 100 km). These particles can collide with the neutral atmospheric gases causing local enhancements in the electron concentration, a phenomenon which is associated with auroral activity. Aurora is the most spectacular effect of the physical processes involved and is the visual manifestation of the interaction between the hot magnetospheric plasma and Earth's upper atmosphere. The auroral zones are relatively narrow rings situated between the northern and southern geomagnetic latitudes of about 64 and 70 degrees. In general, the intensity and the positions of the auroral ovals are related to geomagnetic disturbances. The ovals expand towards the equator with increasing levels of geomagnetic disturbance [5].

2.6 Ionospheric Disturbances

2.6.1 Ionospheric Storms

At different times the ionosphere suffers major perturbations called storms. They last from a few hours to a few days and tend to occur during times of geophysical disturbance resulting from increases in solar activity communicated via the solar wind. Three phases are identified in ionospheric storm. The first, known as *initial* or *positive phase*, is associated with an increase in electron density from the normal value and usually lasts for a few hours. Then follows the *main* or *negative phase* when the electron density are reduced below normal values. The final phase is the *recovery phase* in which the ionosphere gradually returns to normal over a period of one to several days [8].

2.6.2 Geomagnetic Storms

Geomagnetic storms usually occur in conjunction with ionospheric storms and can be caused by solar flares, high speed solar wind stream (coronal holes) and sudden disappearing filaments. The storms are usually associated with increased electron densities in the lower ionosphere and a simultaneous increase in absorption of radio waves. Like the ionospheric storm, it may last from a few hours to several days and it often exhibits three phases. A geomagnetic storm usually starts with an increase in the Earth's geomagnetic field intensity called the initial phase followed by a large decrease termed the main phase. Because magnetic storms can be monitored without great difficulty using a magnetometer, and long runs of such measurements exist, the magnetic storm has come to be a common reference point in geophysical studies [8].

2.6.3 Ionospheric Scintillation

Small-scale structures in the electron content of the ionosphere can range from a few meters to a few kilometers in extent which can cause both refraction and diffraction effects on the electromagnetic waves propagating through the ionosphere. Consequently,

the wavefront becomes crinkled giving rise to amplitude and phase fluctuations of the signal. These fluctuations lead to fading in signal power caused by small scale ionospheric structures are called *ionospheric scintillations*. Scintillation is more prevalent at very high and very low latitudes.

2.7 Ionospheric Effects on the Propagation of Electromagnetic Waves

The ionosphere has many effects on electromagnetic waves, which apply primarily below 10 GHz and are pronounced below 1 GHz. The effects of the ionosphere correlate with Sun spot activity and the diurnal cycle. Possible effects include: fading (loss in intensity), depolarization or Faraday rotation, time delay, dispersion and absorption [3]. Faraday rotation is an angular rotation of the polarization vector. For circular polarization, the only impact is to change the angle between the two sets of polarization axes. The propagation delay effect is important in satellite navigation such as GPS.

Ionospheric effect on electromagnetic waves, such as GPS signals, cannot be described by a simple dispersion. To describe the behavior of radio waves in the ionosphere we need to realize that the ionosphere is a partially ionized, spherically stratified plasma with irregularities upon which imposed a uniform magnetic field. That is we can treat the ionosphere as a plasma embedded in an electromagnetic field. If the effects of electromagnetic fields can be ignored, the relevant physics equations are equations of hydrodynamics. However because electric and magnetic fields and currents are always important in plasmas, we must introduce the equation of *Magneto Hydro-Dynamics* (MHD) to describe the ionospheric plasma [2].

Applying the MHD equations to ionospheric plasma with the assumption of electrical neutrality to a compressible, collisionless plasma fluid in the absence of gravitational forces the appropriate equations governing its behavior are the continuity equation

$$\frac{\partial n}{\partial t} + \vec{\nabla}.(n\vec{v}) = 0 \tag{2.7.1}$$

the force equation

$$\frac{\partial \vec{v}}{\partial t} + (\vec{v}.\vec{\nabla})\vec{v} = \frac{e}{m}(\vec{E} + \vec{v} \times \vec{B}) - \frac{1}{mn}\vec{\nabla}p \tag{2.7.2}$$

where n is the plasma density, $\vec{v}$ is velocity, and p is pressure taken to be scalar. Details can be found in references [9, 10].

The Maxwell's equations in the plasma fluid can be written as

$$\vec{\nabla}.\vec{E} = \frac{e(n - n_0)}{\epsilon_0} \tag{2.7.3}$$

$$\vec{\nabla}.\vec{B} = 0 \tag{2.7.4}$$

$$\vec{\nabla} \times \vec{E} + \frac{\partial \vec{B}}{\partial t} = 0 \tag{2.7.5}$$

$$\vec{\nabla} \times \vec{B} - \epsilon_0\mu_0\frac{\partial \vec{E}}{\partial t} = \mu_0 en\vec{v} \tag{2.7.6}$$

where n_0 is plasma density in an equilibrium state, $\rho_e = e(n - n_0)$ is charge density and $\vec{J} = en\vec{v}$ is the current density.

In calculating the characteristics electromagnetic waves that can penetrate in the plasma, it is assumed that the plasma is initially uniform and that plasma density n_0, pressure p_0 and magnetic field $\vec{B_0}$ are constant. Also, there are no imposed electric fields ($\vec{E_0} = 0$) the plasma is stationary ($\vec{v_0} = 0$). Then the plasma is perturbed with small amplitude so that:

$$\begin{aligned} n &= n_0 + n_1(\vec{r}, t) \\ \vec{v} &= \vec{v_1}(\vec{r}, t) \\ p &= p_0 + p_1(\vec{r}, t) \\ \vec{E} &= \vec{E_1}(\vec{r}, t) \\ \vec{B} &= \vec{B_0} + \vec{B_1}(\vec{r}, t) \end{aligned} \tag{2.7.7}$$

where the subscript 1 in Eq. (2.7.7) denotes small perturbation. The perturbation can be assumed to be sinusoidal so that

$$n_1, \vec{v_1}, p_1, \vec{E_1}, \vec{B_1} \propto \exp i(\vec{k}.\vec{r} - \omega t) \tag{2.7.8}$$

Substituting Eq. (2.7.8) into Eq. (2.7.7) and Maxwell's equations, Eqs. (2.7.3-6), into the MHD equations, Eq. (2.7.1) and Eq. (2.7.2), we get the equations reduced and linearized in terms of the perturbed amplitudes. After some algebra will arrive at the dispersion equation

$$\omega^2 = \omega_p^2 + c^2k^2 \tag{2.7.9}$$

The details of the calculation can be found in references [2, 5, 9, 10]. Eq. (2.7.9) is a dispersion relation for transverse electromagnetic waves of angular frequency ω in a plasma fluid of frequency $\omega_p^2 = n_0e^2/m\epsilon_0$. At high frequencies the greater inertia of the ions implies that they are unable to follow the rapid fluctuation of the fields. That means only electrons dominate the motion. The ions merely provide a uniform background of positive charge to give electrical neutrality on the average. Moreover, in less dense plasma and high frequency waves, the collision frequency is smaller [9]. Hence we can neglect both the ion frequency and collision frequency. The plasma frequency will be simply the electron frequency.

The wave number k can be written as $k = n\omega/c$, where n is the index of refraction. From Eq. (2.7.9) the phase index of refraction of a plasma is then given by

$$n_{ph}^2 = 1 - \frac{\omega_p^2}{\omega^2} \tag{2.7.10}$$

Eq. (2.7.10) is the Appleton-Hartree equation describing the dispersion encountered by an electromagnetic wave upon transmitted through plasma. Since the ionosphere is a partially ionized plasma, Eq. (2.7.10) is the basic equation in understanding satellite signal transmission through the ionosphere. For high frequency radiation ($\omega > \omega_p$) the index of refraction is real and less than 1, thus the wave travel in the ionosphere faster than vacuum i.e. the phase is advanced.

Note that Eq. (2.7.10) is a phase refractive index. The group refractive index with this equation can be found in references [9, 11], from

$$v_g = v_{ph} - \lambda\frac{dv_{ph}}{d\lambda} \tag{2.7.11}$$

where v_{ph} is the phase velocity and v_g is group velocity of the electromagnetic wave. Using $v_{ph} = c/\sqrt{1-\omega_p^2/\omega^2}$ in Eq. (2.7.11) and the group refractive index $n_g = c/v_g$ with the requirement $\omega > \omega_p$, we get

$$n_g^2 = 1 + \frac{\omega_p^2}{\omega^2} \tag{2.7.12}$$

The fact that n_g is greater than 1 in Eq. (2.7.12) means the wave group travels slower in the ionosphere than its speed in vacuum i.e. the wave group is delayed.

For frequencies lower than the plasma frequency ω_p, n_{ph} is imaginary, consequently such electromagnetic waves incident on plasma (ionosphere) will be reflected from the surface. In the next chapter we will see ionospheric sounding using EM signals from ground receivers satellites which is the theme of this thesis.

Chapter 3

THE GLOBAL POSITIONING SYSTEM

3.1 Introduction

The Global Positioning System (GPS) is a satellite system owned and operated by the United States of America's Departement of Defense and Departement of Transportation. The GPS system provides accurate, continuous, world wide, three dimensional position and velocity information to users with appropriate equipment [11]. According to its organizational structure, the GPS system can be divided into three components: *the user segment*, *the space segment* and *the control segment*. The control segment, currently operated by the US Department of Defense, maintains and controls access to the system. The space segment currently consists of about 31 satellites in circular orbits of $20,200\ km$ above the Earth surface in 6 orbital planes inclined 55 degrees with respect to the equator and with $12h$ periods. Each plane is filled with 4 to 6 satellites spaced to guarantee a minimum visibility of 4 satellite at any time at any location on the globe. To eliminate dispersion and refraction effects due to the Earth's ionosphere, the GPS signals are transmitted on two distinct frequencies $f_1 = 1575.42\ MHz$ and $f_2 = 1227.6\ MHz$ used by anyone for scientific purposes and a third frequency currently in use by the US military only. Each satellite also transmits its own pseudorandom noise (PRN) code which is used to identify the satellites signal in the receiver, hence the satellites are usually identified by

the prefix PRN followed by the satellite number. The user segment consists of the receiver that uses the signal provided by the satellites to calculate positions. GLONASS (GLObal NAvigation Satellite System) is a Russian equivalent to GPS and Galileo is an European Union (EU) satellite system. The combined system of the three satellite systems forms the Global Navigation Satellite System (GNSS) [11, 12].

3.2 GPS Observables

The basic GPS observables are code psedoranges and carrier phases. Doppler measurement of the signal is also another observable less commonly used. The Principles of GPS navigation formulation and any GPS measurements are based on these observables.

3.2.1 Pseudoranges

The pseudorange is the measure of the distance between the satellite and the receivers antenna. The pseudorange is calculated by multiplying the apparent travel time of the GPS signal from a satellite to the receiver with the speed of light. This is not equal to the real satellite-receiver range because of a time difference between receiver and satellite clocks errors and transmitting media delay effects, hence the name pseudorange (false range). The transmitting medium not only delays the signals but also bends their path, consequently the path of the satellite signals differs slightly from the geometric path [11].

3.2.2 Carrier Phases

The Carrier phase is a measure of the phase of the received satellite signal relative to the receiver generated carrier phase at the reception time [11]. The received signal phase is tracked by shifting the receiver generated phase. A full carrier phase is called a cycle, with the frequencies f_1 and f_2 the wavelength of one cycle is $19cm$ and $24cm$ respectively. Range measurement based on carrier phase would simply be the sum of the total number

of full carrier cycles plus fractional cycles at the receiver and the satellite, multiplied by the carrier wavelength . The ranges determined with the carrier phases are far more accurate than those obtained with the code pseudoranges. This is due to the fact that the wavelength (or resolution) of the carrier phase, 19 cm in the case of f_1 frequency, is much smaller than those of the codes [12, 13].

3.2.3 Doppler Measurements

The Doppler Measurement observable is based on counting the frequency shift of the signal caused by relative motion of the GPS satellite and the receiver.

3.3 GPS Coordinate Systems

GPS satellites are orbiting around the Earth with time and most GPS surveys are made on the Earth. These surveys need the formulation of the mathematics of satellite navigation based on a suitable reference coordinate system in which the state of both the satellite and receiver (station location) can be represented. The next four sub-sections describe the coordinate systems used by GPS navigation.

3.3.1 Geocentric Earth-Fixed Coordinate System

This coordinate system is also known as the Earth Centered Earth Fixed (ECEF). It is the convenient coordinate system to describe the location of a GPS station on the Earth's surface [11]. The ECEF coordinate system is a right-handed Cartesian coordinate system $(x,\ y,\ z)$. Its origin and the Earth's center of mass coincide, while the mean rotational axis of the Earth is taken as its z-axis; the x-axis is pointing to the mean Greenwich meridian, while the y-axis is directed to complete a right-handed system, Fig. 3.1. In other words, the z-axis is pointing to a mean pole of the Earth's rotation. Such a mean pole, defined by international convention, is called the Conventional International Origin

(CIO). Then the xy-plane is called mean equatorial plane, and the xz-plane is called mean zero-meridian plane [12].

The ECEF coordinate system is also known as the Conventional Terrestrial Reference System (CTRS). The CTRS must be positioned with respect to the Earth (known as realization) to be of practical use in positioning. This is done by assigning coordinate values to a selected number of well distributed reference stations. One of the most important CTRSs is the International Terrestrial Reference System (ITRS), which is realized as the International Terrestrial Reference Frame (ITRF). The ITRF solution is based on the measurements from globally distributed reference stations using GPS and other space geodetic systems. The ITRF is updated every 1 to 3 years to achieve the highest possible accuracy [13].

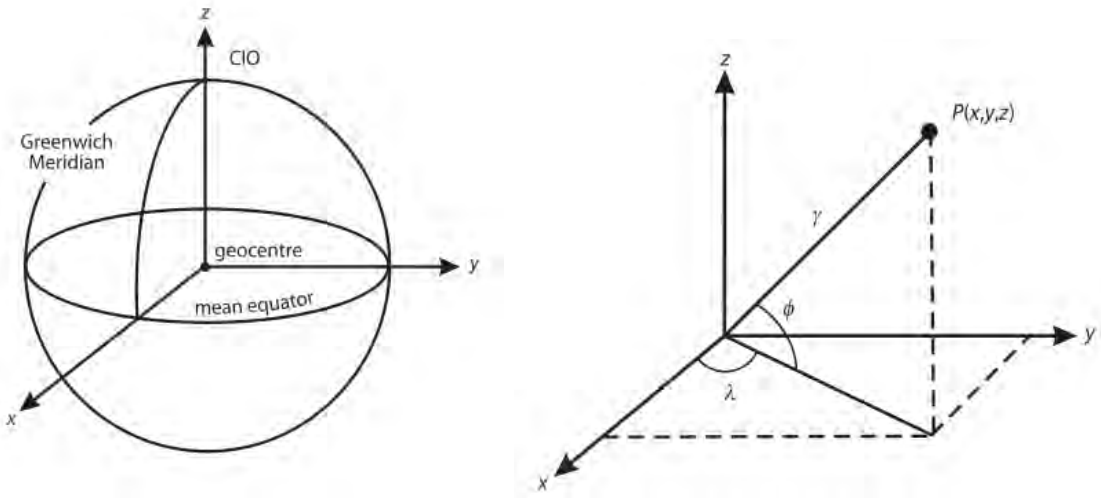

Figure 3.1: Earth Centered Earth Fixed (ECEF) coordinates and its representation in Cartesian and spherical coordinate systems (from [12]).

The ECEF coordinate system can also be represented by a spherical coordinate system (r, ϕ, λ), where r is the radius of the point (x, y, z), ϕ and λ are the geocentric latitude and longitude, respectively (Fig. 3.1). λ is counted eastward from the zero-meridian. The relationship between $(x,\ y,\ z)$ and $(r,\ \phi,\ \lambda)$ is given by Eq. (3.3.1).

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} r\cos\phi\cos\lambda \\ r\cos\phi\sin\lambda \\ r\sin\phi \end{pmatrix} \quad \text{or} \quad \begin{cases} r = \sqrt{x^2+y^2+z^2} \\ \tan\lambda = y/x \\ \tan\phi = z/\sqrt{x^2+y^2} \end{cases} \tag{3.3.1}$$

3.3.2 Geodetic Coordinate System

Geodetic coordinate system is an ellipsoidal coordinate system defined based on the ECEF coordinates. In this system, the coordinates of a point are identified by the geodetic latitude (φ), the geodetic longitude (λ), and the height above the reference surface (h) (Fig. 3.2). Geodetic longitude and geocentric longitude are identical. Two additional parameters are needed to define the shape of the ellipsoid. These could be the semi-major radius (denoted by a) and the semi-minor radius (denoted by b) of the rotating ellipse, or the semi-major radius and the flattening (denoted by f) of the ellipsoid.

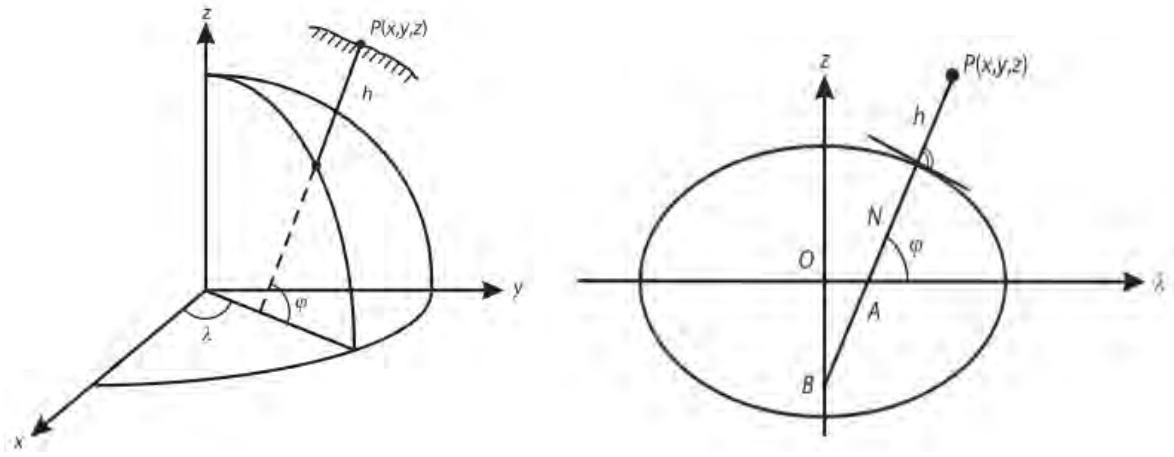

Figure 3.2: Geodetic Coordinate system and the radius of curvature in the prime vertical (from [12]).

The relationship between (x, y, z) and Geodetic coordinates (φ, λ, h) is:

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} (N+h)\cos\varphi\cos\lambda \\ (N+h)\cos\varphi\sin\lambda \\ (N(1-e^2)+h)\sin\varphi \end{pmatrix} \quad \text{Or,} \begin{cases} \tan\varphi = \frac{z}{\sqrt{x^2+y^2}}(1-e^2\frac{N}{N+h})^{-1} \\ \tan\lambda = y/x \\ h = \sqrt{x^2+y^2}/\cos\varphi - N \end{cases} \tag{3.3.2}$$

where

$$N = \frac{a}{\sqrt{1 - e^2 \sin^2 \varphi}} \tag{3.3.3}$$

N is the radius of curvature in the prime vertical as shown in Fig. 3.2, and e is the first eccentricities. In Eq. (3.3.2) φ and h have to be solved by iteration. The flattening and the first eccentricities are defined as:

$$f = \frac{a - b}{a}, \quad \text{and} \quad e = \frac{\sqrt{a^2 - b^2}}{a} \tag{3.3.4}$$

The two geometric parameters used in the World Geodetic System 1984 (WGS-84) are ($a = 6378137m, \quad f = 1/298.2572236$)

Using Eq. (3.3.1) and Eq. (3.3.2), the relation between geocentric and geodetic latitude ϕ and φ is

$$\tan \phi = (1 - e^2 \frac{N}{N + h}) \tan \varphi \tag{3.3.5}$$

3.3.3 Earth Centered Inertial Coordinate System

Earth Centered Inertial (ECI) coordinate system is an inertial coordinate system defined to predict the position of Earth-orbiting satellites. The reason is that the motion of the satellites follows the Newtonian mechanics, and the Newtonian mechanics is valid and expressed in an inertial coordinate system. ECI is fixed in space relative to stars and therefore inertial. ECI is often defined as a Cartesian coordinate system, where the position (coordinates) is given as the distance from the origin, which is the center of mass of the Earth, along the three orthogonal axes [11, 14]. In typical ECI coordinate systems, the xy-plane is taken to coincide with the Earth's equatorial plane, the x-axis is permanently fixed in a particular direction relative to the celestial sphere, the z-axis is taken normal to the xy-plane in the direction of the north pole, and the y-axis is chosen so as to form a right-handed coordinate system. However, the irregularities of

the earth motion that consist of the slow variation in orientation due to nutation and precession make the ECI system not really inertial, because the x-axis is fixed with respect to the earth motion, whereas the z-axis moves following the equatorial plane motion due to the earth motion. Hence, a conventional ECI coordinate system has been defined choosing the orientation of the axes relative to the orientation of the equatorial plane at a particular instant in time of January 1, 2000, at 12:00:00.00 UTC as shown in Fig. 3.3. This coordinate system is known to be the J2000 ECI coordinate system or equatorial coordinates of date. The J2000 ECI coordinate system uses the Earth center as the origin, x-axis the direction of the vernal equinox at January 1, 2000, at 12:00:00.00 UTC. z-axis Earth's rotational axis in the direction of the North Pole at this date and y-axis completes a right-handed, Earth-centered, orthogonal coordinate system. Determination and subsequent prediction of the GPS satellite orbits are carried out in an ECI coordinate system [11].

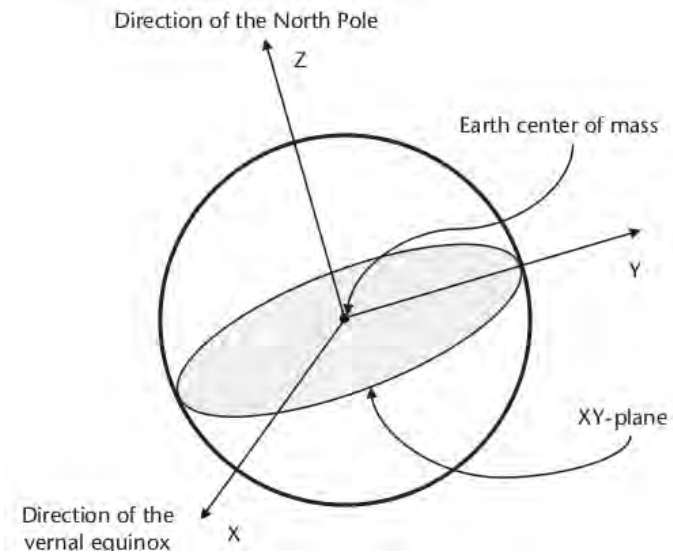

Figure 3.3: Earth Centered Inerital (ECI) coordinate system (from [14]).

3.3.4 Local Coordinate System

The local coordinate system is a left-handed Cartesian coordinate system (x', y', z') defined by placing the origin to a local point on the surface of the Earth in WGS-84 coordinate system. In local coordinate system x'-axis is directed to the north, y' is pointed

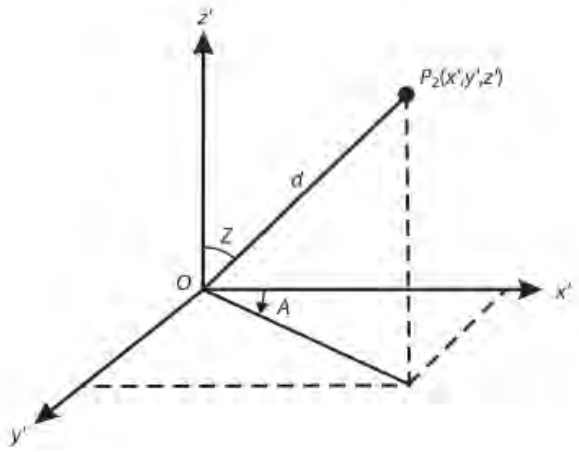

Figure 3.4: GPS local coordinate system (from [12])

to the east and z' axis is pointed to the vertical, Fig. 3.4. The $x'y'$-plane is called the horizontal plane; the vertical is defined perpendicular to the ellipsoid. For any point P_2 as shown in Fig. 3.4, whose coordinates in the global and local coordinate system are (x, y, z) and (x', y', z'), respectively, we have the relations

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = d \begin{pmatrix} \cos A \sin Z \\ \sin A \sin Z \\ \cos Z \end{pmatrix} \quad \text{and,} \quad \begin{cases} d = \sqrt{x'^2 + y'^2 + z'^2} \\ \tan A = y'/x' \\ \cos Z = z'/d \end{cases} \tag{3.3.6}$$

where A is the azimuth, Z is the zenith angle ad d is the distance from the origin to a point P_2; A is measured clockwise from the north,Z is the angle from the vertical to the radius d. We can obtain the local coordinate system by an appropriate rotations of the global coordinate system [12].

3.4 GPS Time Systems

Fixing coordinate system is not enough for GPS positioning, time also plays important role. GPS signal is controlled by accurate timing devices, the atomic satellite clocks. In addition, measuring the ranges (distances) from the receiver to the satellites is based on both the receiver and satellite clocks [13]. GPS disseminates a realization of Coordinated Universal Time (UTC) that provides the capability for time synchronization of users worldwide. Applications range from data time tagging to communications system packet switching synchronization [1]. Three time systems are used in satellite surveying. They are sidereal time, dynamic time and atomic time [12]. Sidereal time is a measure of the Earth's rotation and is defined as the hour angle of the vernal equinox. Universal Time (UT) is a sidereal time measured from Greenwich meridian and orbits uniformly in the equatorial plane. As the angular velocity of the Earth's rotation is not a constant, sidereal time is not a uniformly scaled time. The oscillation of UT is also partly caused by the polar motion of the Earth. The universal time corrected for the polar motion is denoted by UT1 [11]. Dynamical time is a uniformly scaled time used to describe the motion of bodies in a gravitational field. There are two types of dynamic time. Namely, Barycentric Dynamic Time (TDB) and Terrestrial Dynamic Time (TDT). TDB is applied in an inertial coordinate system with its origin being located at the center of mass. Similarly, TDT is used in a quasi inertial coordinate system (such as ECI). TDT is also called Terrestrial Time (TT). Atomic Time is a time system kept by atomic clocks such as International Atomic Time (TAI). It is a uniformly scaled time used in the ECEF coordinate system. GPS Time (GPST) is an atomic time. The relationship among the different time systems is given by [11]:

$$\begin{aligned} \text{TAI} &= \text{GPST} + 19.0sec \\ \text{TAI} &= \text{TDT} - 32.184sec \\ \text{TAI} &= \text{UTC} + nsec \\ \text{UT1} &= \text{UTC} + d\text{UT1} \end{aligned} \tag{3.4.1}$$

where dUT1 can be obtained by International Earth Rotation Service (IERS), n is the number of leap seconds of date and is inserted into UTC on the 1^{st} of January and 1^{st} of July of the years. The actual n can be found in the IERS report.

3.5 GPS Errors

GPS signals in thier way from the satellites to the recieving stations pass through different media such as the reciever and transmitter hardwares, the neutral atmosphere and the electrically charged ionosphere. In this section we will see the effects of these media and other physical influences on GPS signals.

3.5.1 Atmospheric Effect

Contrary to the ionosphere, atmosphere is non-dispersive for GPS signals which means atmospheric effects are indipendent of the frequencies of the signals. The GPS signals are affected by neutral atoms and molecules in the atmosphere. The effects are called atmospheric delay. Historically atmospheric delay is also commonly called tropospheric delay or tropospheric refraction to mean the effects of the atmosphere below the ionosphere. The atmospheric delay in the zenith direction can reach up to 2m [15].

The atmospheric delay is usually of two types. The first component is the mixture of all constituents except water vapor. This component of the atmospheric propagation delay is usually termed the “dry” or “hydrostatic” delay and accounts for nearly all of the atmospheric delay [15]. The other component of atmospheric delay is due to water vapor and is said to be “wet delay” and is zero if there is no water vapor present anywhere along the path of the signal. Wet delay can be useful in modeling atmospheric water content using GPS [15].

3.5.2 Ionospheric Effect

As seen in the last section of chapter 2, the ionosphere has a dispersive effect on the propagation of electromagnetic waves through it. As GPS satellites transmit signals, it passes through the Earth's ionosphere; consequently the ionospheric effect is an error source in GPS navigation. The amount of ionospheric delay on modulated signals or advance on phase of signals can vary from a few meters to more than twenty meters within one day [11]. In general it is difficult to model the ionospheric effect on GPS signals due to complicated physical interactions among geomagnetic field and solar activities. GPS system is designed with several working frequencies, so that ionospheric effect can be measured or corrected using the fact that the ionosphere is dispersive. This thesis is aimed at measuring the ionospheric Total Electron Content (TEC) using GPS satellites and the TEC is inverted to find the density of electrons.

From the last section of the second chapter, using binomial expansion to Eq. (2.7.10) and Eq. (2.7.12), the phase and group refractive indices respectively of an EM waves (such as GPS signals) as they travel through the ionosphere become

$$n_{ph} = 1 - \frac{\omega_p^2}{2\omega^2} \qquad and \qquad n_g = 1 + \frac{\omega_p^2}{2\omega^2} \tag{3.5.1}$$

Substituting the plasma frequency $\omega_p = Ne^2/\epsilon_0 m_e$, where N is the electron density, m_e mass of electron Eq. (3.5.1) becomes

$$n_{ph} = 1 - \frac{A}{f^2}N(\vec{r}) \qquad and \qquad n_g = 1 + \frac{A}{f^2}N(\vec{r}) \tag{3.5.2}$$

where $A = \frac{e^2}{8\pi^2\epsilon_0 m_e} \approx 40.3\frac{m^3}{s^2}$

The travel time of an electromagnetic wave through a medium with refractive index n, at speed v is given as

$$t_{travel} = \int_{raypath} \frac{1}{v} d\vec{r} = \int_{raypath} \frac{n}{c} d\vec{r} \tag{3.5.3}$$

Substituting Eq. (3.5.2) into Eq. (3.5.3) the ionospheric phase delay relative to wave propagation in vacuum, I_p, is

$$I_{ph} = t_{travel}^{medium} - t_{travel}^{vacum} = -\frac{A}{f^2 c}\int_{raypath} N(\vec{r})d\vec{r} \tag{3.5.4}$$

and the wave group delay, I_g, is given by

$$I_g = t_{travel}^{medium} - t_{travel}^{vacum} = \frac{A}{f^2 c}\int_{raypath} N(\vec{r})d\vec{r} \tag{3.5.5}$$

Eq. (3.5.4) and Eq. (3.5.5) means that the group delay is equal to the phase advance.

$$I_{ph} = -I_g \tag{3.5.6}$$

Hence the phase of the GPS carrier wave is advanced by the same amount of time that the information in a wavegroup is delayed. The integral in Eq. (3.5.4) and Eq. (3.5.5) is simply the electron density integrated along the signal raypath from satellite to receiver and is called Integrated Electron Content (IEC) or slant Total Electron Content (sTEC) [16]. The TEC can be defined as the total number of ionospheric electrons in a cylindrical tube of cross-sectional area m^2 along the GPS ray path as shown in Fig. 3.5

$$TEC = \int_{raypath} N(\vec{r})d\vec{r} \tag{3.5.7}$$

This means when satellite radio signals (such as GPS) traverse through the ionosphere, substantial signal delays occur and the amount of delay is proportional to the line integral of Ionospheric Electron Density (IED) along the satellite ray path from the satellite to receiver. In Section 3.6 we will see how to derive the TEC by appropriate linear combination of dual frequency GPS observables.

3.5.3 Clock Errors

The influences of the clock errors on the GPS may be grouped into three types. One is factorised with the speed of light, c. Another is factorized with the speed of satellites. And the third is factorized with the working frequency [12]. The first type of clock error is obvious. For code measurements, one measures the transmitting time of the signal and

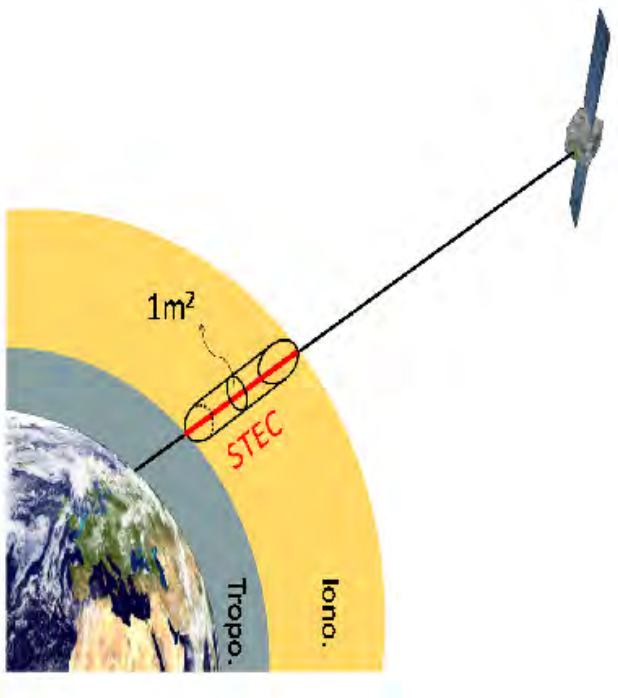

Figure 3.5: Slant Total Electron Content (TEC) along the GPS ray path (from http://www.gps.oma.be).

multiplies the transmitting time with the speed of light c to obtain the transmitting path length. A clock error of δt will cause a path length error of $c\delta t$. Similarly, a clock error of δt will cause a phase error of $c\delta t/\lambda$. Because of the factor c, a small clock error may cause a very large code and phase error. The second type of error comes from the geometric distance between the satellite at the signal emission time and the receiver at the signal reception time. Since position and velocity of satellite are functions of time, clock error δt causes a computing error of the position of the satellite by $\vec{v_s}\delta t$, where $\vec{v_s}$ is the velocity vector of the satellite. Clock errors have also influence on frequency. A phase error $c\delta t/\lambda$ is equivalent to a frequency error of $f\delta t$. This correction has to be taken into account in Doppler data processing.

Synchronization of the clocks on the satellites and receivers is a basic prerequisite of a meaningful GPS measurement. Clock error modeling leads automatically to the synchronization of all clocks [12].

3.5.4 Relativistic Error

Both Einstein's general and special theories of relativity are factors in the pseudorange and carrier-phase measurement process [11]. Special relativity effect arises whenever GPS satellites or GPS receivers are moving with respect to the chosen isotropic light speed frame, which in the GPS system is the ECI frame. General relativity arises as signals sources and receivers are located at different gravitational potentials. The satellite clock is affected by both special and general relativity. To compensate these effects the fundamental frequency (f_0) of GPS system is selected as $10.23MHz$ [12]. All clocks on the GPS satellites and GPS receivers operate based on this frequency to calculate the difference between the setting frequency f' of satellite clock and the fundamental f_0, called the offset in the satellite clock frequency. Such an offset of the relativistic effects has been implemented in the satellite clock settings, and therefore users do not need to consider this effect[12, 11].

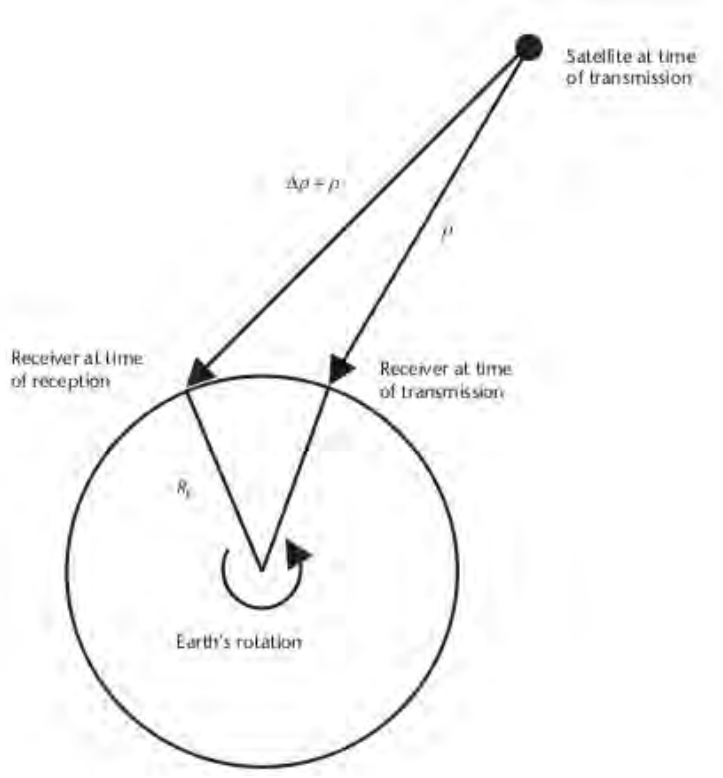

Figure 3.6: Relativistic Effect due to reciever rotation (from [12])

Relativistic effect also arises due to rotation of the Earth when computations for the satellite positions are made in an ECEF coordinate system. Because if the user experiences a net rotation away from the satellite, the propagation time will increase, Fig. 3.6. Relativistic effect due to Earth rotation is known as *Sagnac effect*. And the corrections for the Sagnac effect are often referred to as *Earth rotation corrections* .

Relativistic effects due rotation of the earth corrections, Earth rotation corrections, are given by

$$\Delta\rho = \frac{(\vec{r_r} - \vec{r_s}).\vec{v_r}}{c} \tag{3.5.8}$$

where $\vec{r_s}$ is the geocentric vector of the GPS satellite, $\vec{r_r}$ the geocentric vector of the receiver, $\vec{v_r}$ the velocity vector of the satellite. Details of the calculation can be found in [12].

3.5.5 Multipath Effect

Multipath is the phenomenon whereby a GPS signal arrives at a receivers antenna via more than one different paths, (see Fig. 3.7). Multipath propagation affects both pseudorange and carrier phase measurements. Multipath is a very localized effect, which depends on the local environment surrounding the antenna. Multipath errors vary significantly in magnitude depending on the environment within which the receiver is located, satellite elevation angle, receiver signal processing, antenna gain pattern, and signal characteristics.

Since GPS signals are right handed circularly polarized, receivers designed for such polarized signals can help in rejecting the multipath signals because signals change their polarization upon reflection. Rejecting low elevation angle signals can also reduce multipath effect in GPS.

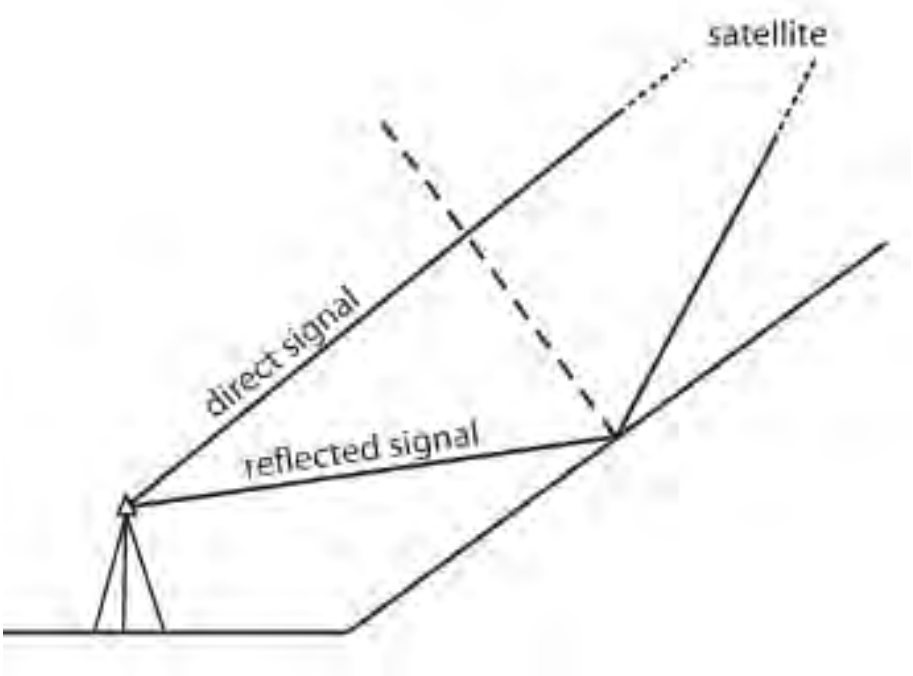

Figure 3.7: Multipath Effect geometry (from [12])

3.5.6 Instrument Biases

Study of ionospheric effects by using GPS observations indicates the existence of the instrumental biases. These error sources are introduced by the different travel time of the GPS signal on the two separate frequencies through the hardwares of satellite and receiver respectively. The relative delay on the satellite is commonly named as transmitter group delay and the delay caused by the receiver is called receiver inter-frequency bias [16].

3.6 Ionospheric Total Electron Content (TEC) Calculation from GPS

In Section 2.7, we have seen the effect of the ionosphere on the propagation of EM waves passing through it. This effect is expressed by the Appleton-Hartree equation, Eqs. (2.7.10) and (2.7.12). Using these equations we have shown in Section 3.5 that the phase of a GPS signal is advanced by the same amount of time that wave group is delayed

(see Eq. (3.5.6)). Based on this, we will see in this section ionospheric TEC calculation based on GPS observables: Pseudoranges and carrier Phase measurements.

One of the fundamental GPS observables is the pseudorange, ρ_i $(i = 1, 2)$ measured on the two frequencies. It is the "false" distance between the GPS satellite and the receiver found by multiplying the apparent travel time of the signal between the satellite and the receiver by the speed of light. This time is derived from the time difference on the clocks of the receiver and the satellite. The pseudoranges can be written as [16]:

$$\rho_1 = \rho + c\tau - cI_1 + cT \tag{3.6.1}$$

$$\rho_2 = \rho + c\tau - cI_2 + cT \tag{3.6.2}$$

where ρ_1 and ρ_2 are pseudoranges corresponding to the two GPS civil frequencies $f_1 = 1575.42MHz$ and $f_2 = 1227.6MHz$ respectively, ρ is satellite to receiver distance, T is tropospheric delay and τ satellite-receiver clock offset. I_1 and I_2 are the frequency dependent ionospheric correction corresponding to f_1 and f_2 . The negative sign for the ionospheric correction is from the fact that the signal is delayed but carrier phase advanced by the ionosphere. The implication is that the pseudorange is increased by the ionospheric effect. Combining Eqs. (3.6.1) and (3.6.2) linearly we get

$$\rho_2 - \rho_1 = -c(I_2 - I_1) \tag{3.6.3}$$

Using Eqs. (3.5.5) and (3.5.7)

$$I_1 = \frac{A}{f_1^2 c} TEC \quad \text{and} \quad I_2 = \frac{A}{f_2^2 c} TEC \tag{3.6.4}$$

Substituting Eq. (3.6.4) into Eq. (3.6.3), we obtain

$$TEC = \underbrace{-\frac{\rho_2 - \rho_1}{\lambda_2}}_{P_G} \frac{f_1^2 f_2}{f_1^2 - f_2^2} \frac{c}{A} \tag{3.6.5}$$

where P_G is the name of the linear combination and we have used $\lambda_2 f_2 = c$. Eq. (3.6.5) provides the measurement for absolute TEC. However, pseudorange measurements are

much noisier and less precise than phase measurements. Hence the linear combination P_G is also less precise than its equivalent derived from phase measurements [16]. In a similar manner, the TEC derived from ionospheric effects on GPS signals can be determined by phase observables. The phase pseudoranges measured in cycles can be written as [16, 12]

$$\Phi_1 = \frac{\rho}{\lambda_1} + f_1\tau + f_1 I_1 + f_1 T + N_1 \tag{3.6.6}$$

$$\Phi_2 = \frac{\rho}{\lambda_2} + f_2\tau + f_2 I_2 + f_2 T + N_2 \tag{3.6.7}$$

where N_1 and N_2 are unknown ambiguities corresponding to f_1 and f_2 respectively. The ambiguities are ambiguous integer number of cycles in the carrier phase caused by the initial lock onto the signal [12, 16]. Forming a linear combination of Eqs. (3.6.6) and (3.6.7) we get

$$\Phi_2 - \frac{f_2}{f_1}\Phi_1 = f_2(I_2 - I_1) + N \tag{3.6.8}$$

Calling this linear combination L_G

$$L_G = \Phi_2 - \frac{f_2}{f_1}\Phi_1 \tag{3.6.9}$$

and using Eq. (3.6.4)

$$f_2(I_2 - I_1) = \left(\frac{f_1^2 - f_2^2}{f_1^2 f_2}\right)\frac{A}{c}TEC \tag{3.6.10}$$

whereby substituting Eqs. (3.6.9) and (3.6.10) into Eq. (3.6.8) gives

$$L_{GU} = \frac{A}{c}\frac{f_1^2 - f_2^2}{f_1^2 f_2}TEC \tag{3.6.11}$$

Here

$$L_{GU} = L_G - N \tag{3.6.12}$$

$$N = N_2 - \frac{f_2}{f_1}N_1 \tag{3.6.13}$$

Solving for TEC from Eq. (3.6.11), we obtain

$$TEC = L_{GU}\frac{f_1^2 f_2}{f_1^2 - f_2^2}\frac{c}{A} \tag{3.6.14}$$

In Eq. (3.6.11), the factor

$$\frac{f_1^2 f_2}{f_1^2 - f_2^2}\frac{c}{A} = 2.3 \times 10^{16} m^{-2} \tag{3.6.15}$$

The value $10^{16} m^{-2}$ is used as a unit of TEC, designated as $1TECU$

$$1TECU = 10^{16} m^{-2} \tag{3.6.16}$$

Since there is no phase ambiguity in the linear combination P_G in Eq. (3.6.5) we can calculate the ambiguity constant N for each satellite pass. From (Eqs. 3.6.5) and (3.6.11)

$$P_G = L_G - N \tag{3.6.17}$$

Using the average difference between P_G and L_G over all available epochs for each arc from a GPS satellite to a receiver gives

$$N = \frac{1}{m}\sum_{i=1}^{m}(L_G - P_G) \tag{3.6.18}$$

3.6.1 TEC Mapping Technique

The TEC is valid for the line of sight from the GPS receiver to the position of the satellite. Generally, the ionosphere with significant electron density distribution begins at an approximate height of 50 km and ends at an approximate height of height 750 km [12]. It is therefore assumed that the ionosphere has an average height of 350 km . The height, h_I, is also the height of main ionospheric concentration. Assuming the free electrons are concentrated in a thin shell at h_I, the slant TEC can be converted to a vertical (zenith) total electron content VTEC.

As shown in Fig. 3.8, let θ be the satellite elevation angle from the receiver and α be elevation angle of satellite from the height of the ionospheric shell, we can have a mapping function $E(\theta)$, so that:

$$VTEC = TEC \times E(\theta) \tag{3.6.19}$$

Denoting zenith angles at receiver position and ionospheric shell by ξ and ξ' respectively as indicated in Fig. 3.8 and using the complementary relation $\xi = \frac{\pi}{2} - \theta$ and the law of

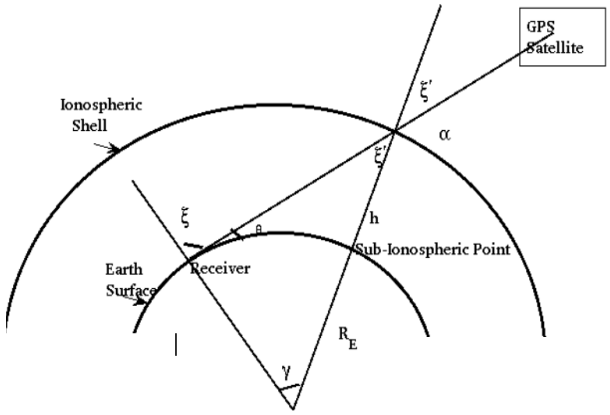

Figure 3.8: Geometry of the ionospheric mapping.

sines gives the mapping function

$$E(\theta) = \sqrt{1 - \left(\frac{\sin(\frac{\pi}{2} - \theta) R_E}{R_E + h_I} \right)^2} \tag{3.6.20}$$

Chapter 4

INVERSE THEORY

4.1 Introduction

Inverse theory is an organized set of mathematical techniques for reducing data to obtain useful information about the physical world on the basis of inferences drawn from observation [17]. The observations collected makeup the data or measurements. The questions we want to answer will be numerical values of specific (but not necessarily directly measurable) properties of the world. These properties will be called *model parameters*. Inverse theory is used in contrast to forward theory, which is a process of predicting the results of measurement (predicting data) on the basis of general principle or model. Inverse theory addresses the reverse problem; starting with data and general principle or model, it determines estimates of model parameters. Inverse problems arise in many branches of the physical sciences. Some of which are medical tomography, image enhancement, earthquake location, determination of earth structure, satellite navigation, etc. Computerized Ionospheric Tomography (CIT) is a recently new inversion technique applied to study the ionosphere electron density distribution using data from satellite signals.

4.2 Formulating Inverse Problems

The basic statement of an inverse problem is that the model parameters m and the data d are in some way related. This relationship is called the *model* [17]. A function, A, may

be specified relating m and d as

$$A(m) = d \tag{4.2.1}$$

In practice the data d may be a function of time and/or space or may be a collection of discrete observables. Actual problems data observations always contain some amount of noise [18]. The forward problem is to find d given m. Computing A might involve solving an ordinary or partial differential equation, evaluating an integral, etc. In inverse problems we focus on finding m given d.

In many cases the model parameters may define a physical entity directly (e.g. density) or may be coefficients in a functional relationship that describes a physical process. In this case we can express the model parameters as a finite sum of n element vector. Similarly, the data can be expressed as m element vector. such problems are called *discrete inverse problems*. In the other case , where the data and model parameters are functions of time and space, the task of estimating the model parameters from the data is called *continuous inverse problem*. Continuous inverse problems can often be well approximated by discrete inverse problems [17, 18].

Mathematical models can be linear systems [18]. Linear systems obey superposition, Eq. (4.2.2), and scaling, Eq. (4.2.3).

$$A(m_1 + m_2) = A(m_1) + A(m_2) \tag{4.2.2}$$

$$A(\alpha m) = \alpha A(m) \tag{4.2.3}$$

In the case of discrete linear inverse problems, Eq. (4.2.1) can be written in the form of linear system of algebraic equations

$$Am = d \tag{4.2.4}$$

4.3 Why Inverse Problems are Hard?

Finding mathematically acceptable answers to inverse problems is not a guarantee for the solution to be true. One reason is that there may be many model parameters that

adequately fit the data. It is essential to characterize just what solution has been obtained, how good it is in terms of physical plausibility and fit to the data, and how consistent it is with other constraints. Jacques Hadmard (1965-1963) introduced, [19], three conditions that a well-posed inverse problem should fulfill. These are *existence*, *uniqueness* and *stability*. However, most inverse problems do not fulfill these conditions and are ill-posed. Usually a solution to an inverse problem is characterized by:

1. Non-existence:- There may be no model parameters that exactly fit the data. This usually can occur as our mathematical model of the systems physics is approximate or the data may contain noise.

2. Non-uniqueness:- If exact solutions do exist they may not be unique. That is, there might be other solutions besides the true solution that exactly satisfy the mathematical model. Non-uniqueness is a characteristic of rank-deficient discrete linear inverse problems [18].

3. Instability:- The process of computing an inverse solution is often extremely unstable. That is, a small change in the measurement (data) can lead to an enormous change in the estimated model parameter.

According to Hadmard, if at least one of the above conditions exist, the inverse problem is *ill-posed*. Hadmard considered that an ill-posed mathematical problem was not physically and/or mathematically meaningful [20]. However it was found that ill-posed inverse problems are physically and mathematically meaningful and can be solved. In fact, the majority of physical problems are ill-posed.

4.4 Solutions to Inverse Problems

In finding a solution to an inverse problem one generally wants to know the numerical values of the model parameters. This solution is commonly known as estimates of model

parameters because it is very rarely that an inverse problem be solved in such a way to yield exact solutions. Estimates are generally the most useful kind of solution to an inverse problem [17].

Inverse problems can be solved in different ways that depends upon weather the problem is linear or non-linear and discrete or continuous. In the following sections we will see the least square and iterative methods of finding a solution to an inverse problem with more emphasis on those methods commonly applied for ionospheric tomography.

4.4.1 Least Square Solution of Linear Inverse Problems

Let us consider a discrete linear inverse problem with an observation data, d and a vector m of model parameters that are going to be solved. The model can be written as a linear system of equations as given in Eq. (4.2.4)

As said earlier, it is the usual case that no solution m satisfies Eq. (4.2.4) exactly. A useful approximate solution may still be found by finding a particular model parameter m that minimizes some measure of the misfit between the actual data d and Am. The misfit is a vector of residuals r and can be expressed as

$$r = d - Am \tag{4.4.1}$$

The common measures of the length of the residuals (misfit) are expressed as $n\text{-}norm$ given by

$$n\text{-}norm: \quad \|r\|_n = \left[\sum_i |r_i|^n\right]^{1/n} \tag{4.4.2}$$

In the case $n = 2$ the length is the $2\text{-}norm$. A model that minimizes the $2\text{-}norm$ is called *a least square solution*. The least square solution is of special interest because it is very amendable to analysis and geometric intuition, and because it is the most likely solution if data errors are normally distributed [18].

The least square solution can be solved by computing the derivative of $r^T r$ with respect

to one of the model parameters, m_q, and setting the result to zero.

$$r^T r = (d - Am)^T (d - Am) = \sum_{i=1}^{m} \left[d_i - \sum_{j=1}^{n} A_{ij} m_j \right] \left[d_i - \sum_{k=1}^{n} A_{ik} m_k \right] \tag{4.4.3}$$

Carrying out the derivative of Eq. (4.4.3) with respect to m_q and setting it to zero, we get

$$\frac{\partial}{\partial m_q}(r^T r) = 2 \sum_{k=1}^{n} m_k \sum_{i=1}^{m} A_{iq} A_{ik} - 2 \sum_{i=1}^{m} A_{iq} d_i = 0 \tag{4.4.4}$$

Expressing Eq. (4.4.4) in matrix notation, yields

$$A^T Am - A^T d = 0 \tag{4.4.5}$$

Presuming, $(A^T A)^{-1}$ exists, the least square solution to the inverse problem is

$$m^{est} = (A^T A)^{-1} A^T d \tag{4.4.6}$$

In the least square solution of an inverse problem, we estimated a solution that best minimizes the 2-*norm* prediction error (residuals). However, least squares fail if the number of solutions that give the same minimum prediction error is greater than one.

Inverse problems are classified as *under determined*, *even determined* and *over determined* on the basis of weather Eq. (4.2.4) provides enough information to specify uniquely the model parameters [17].

Under Determined Problems

An inverse problem is said to be under determined when Eq. (4.2.4) has fewer equations than the variables. That is, when there are more unknowns (model parameters) than data.

Over Determined Problems

When an inverse problem contains too much information in Eq. (4.2.4) it is said to be over determined. Over determined problems have more data than the unknown model parameters.

Even Determined Problems

In this case there is exactly enough information to determine the model parameters.

4.4.2 Damped Least Square Solution

In this case we determine a solution that minimizes some combination $\Phi(m)$ of the prediction error ($r^T r$) and solution length ($m^T m$) for the model parameters [17, 18]:

$$\Phi(m) = r^T r + \alpha^2 m^T m \tag{4.4.7}$$

where α^2 is the weighting factor, that determines the relative importance given to the norm of residuals (prediction error) and solution length. There is no simple method of determining α that will approximately minimize the prediction error while approximately minimizing the length of the under determined part of the solution [18]. Thus the critical question in the solution of inverse problems is the selection of the optimal parameter α. The basic principles of determining α are discussed towards the end of this chapter, when we see regularization of inverse problems.

We can rewrite $\Phi(m)$ in Eq. (4.4.9) as

$$\begin{aligned} \Phi(m) &= [d - Am]^T [d - Am] + \alpha^2 m^T m & (4.4.8) \\ &= \sum_{i=1}^{m} \left[d_i - \sum_{j=1}^{n} A_{ij} m_j \right] \left[d_i - \sum_{k=1}^{n} A_{ik} m_k \right] + \alpha^2 \sum_{j=1}^{n} m_j^2 & (4.4.9) \end{aligned}$$

By minimizing $\Phi(m)$ in Eq. (4.4.9) in a manner described in ordinary least square solution; i.e. differentiating Eq. (4.4.9) with respect to m_q, and setting the result to zero, we obtain

$$(A^T A + \alpha^2 I)m - A^T d = 0 \tag{4.4.10}$$

Hence, the estimate of model parameters is

$$m^{est} = \left[A^T A + \alpha^2 I\right]^{-1} A^T d \tag{4.4.11}$$

This estimate of model parameters is called *damped least square solution*. The under determinacy of the inverse problem is said to have been damped.

So far in our discussion of the damped least square solution, we have minimized a function which is a combination of the prediction error (residuals) and the solution length. In many situations, however, we would prefer to obtain a solution that minimizes some other measure of the model parameters m, such as the first or the second derivative of m. The advantage of using such minimization in the damped least square solution is it incorporates information regarding the smoothness of the solution [21, 22]. This kind of minimization is discussed in detail in Section 4.5.1.

4.4.3 Iterative Methods of Solution of Linear Inverse Problems

The damped least square solution (Tikhonov regularized solutions) to linear inverse problems become impractical when we have large problems in which the matrix A in Eq. (4.2.4) is tens of thousand of rows and columns [18]. These problems require large storage memory even when the majority of the elements of A are zero, i.e A is sparse matrix and we can save memory by only storing the non-zero elements of A.

Iterative methods work by generating a sequence of model parameters m that converge to an optimal solution. Iterative methods commonly involve multiplying the matrix A and its transpose times vectors which can be done without additional storage. We often terminate iterative methods based on some stopping criteria.

Iterative algorithms particularly effective for tomographic applications are the Kaczmarz's algorithm, its Algebraic Reconstruction Technique (ART) and Simultaneous Iterative Reconstruction Technique (SIRT) [18]. Other Iterative methods of inversion include the Conjugate Gradient (CG), the Conjugate Gradient Least Square (CGLS). Details can found in references [18, 23].

The Kaczmarz's ART algorithm is easy to implement for solving a linear system of equations given as Eq. (4.2.4). To understand the algorithm, note that each of the m rows of the system $A_i m = d_i$ defines an n-dimensional hyperplane in R^m. Kaczmarzs algorithm starts with an initial solution $m^{(0)}$, and then moves to a solution $m^{(1)}$ by projecting the

initial solution onto the hyperplane defined by the first row in A. Next, $m^{(1)}$ is similarly projected onto the hyperplane defined by the second row in A, and so forth. The process is repeated until the solution has been projected onto all m hyperplanes defined by the system of equations. At that point, a new cycle of projections begins. These cycles are repeated until the solution has converged sufficiently [18].

To implement the Kaczmarz's ART algorithm, we need to formulate the mathematics to compute the projection of a vector onto the hyperplane defined by equation i. Let A_i be the i^{th} row of A, Consider the hyperplane defined by $A_{i+1}m = d_{i+1}$. Because the vector A_{i+1}^T is perpendicular to this hyperplane, the update to $m^{(i)}$ from the constraint due to row $i+1$ of A will be proportional to A_{i+1}^T

$$m^{(i+1)} = m^{(i)} + \beta A_{i+1}^T \tag{4.4.12}$$

Using the fact that $A_{i+1}m^{(i+1)} = d_{i+1}$ to solve for β, we obtain

$$A_{i+1}m^{(i)} + \beta A_{i+1}A_{i+1}^T = d_{i+1} \tag{4.4.13}$$

Solving for β from Eq. (4.4.13) and substituting back to Eq. (4.4.12) we obtain the update formula to be

$$m^{(i+1)} = m^{(i)} - \frac{A_{i+1}m^{(i)} - d_{i+1}}{A_{i+1}A_{i+1}^T} A_{i+1}^T \tag{4.4.14}$$

As mentioned the ART method of inversion requires some initial values of the quantity to be reconstructed, the a priori value. The initial a priori value in ART represents a gross estimate or guess of what the reconstruction may look like and improves the solution on the a priori value by iterative correction[24]. SIRT and the Modified ART (MART) are other versions of the Kaczmarz's ART. Another commonly used ART method of inversion applied in ionospheric tomography is the Improved Algebraic Reconstruction Technique (IART) given by [25, 26].

$$m^{(i+1)} = m^{(i)} + \lambda_i(d_i - A_i m^{(i)}) \tag{4.4.15}$$

The column vector λ_i is called the relaxation parameter driven by the data and is given as

$$\lambda_i = g^i/(A_i.g^i) \tag{4.4.16}$$

where $g^i = [g^i_1,\ g^i_2,\ g^i_3 ... \ g^i_n,]^T$ and $g^i_m = A_{in}m^i_n$. Debao Wen et.al., [26], has made a comparison between ART and IART and found that IART produces a solution with better accuracy than ART .In a similar manner reference [25] applied IART with an initial guess obtained by damped least square inversion regularized with first order Tikhonov regularization and deduced that the combined algorithm produces a better solution than Tikhonov regularization alone.

4.5 Regularization Methods

At the beginning of this chapter, it is described that the majority of inverse problems are ill-posed, i.e. at least one of the Hadmard's conditions fail. Solutions of ill-posed inverse problems would be extremely unstable and unrealistic [20]. However, ill-posed inverse problems can be treated with certain conditions to become well-posed. These treatments given to ill-posed inverse problems to become well-posed is said to be *regularization methods*. Regularization is frequently essential to producing a usable solution to an otherwise intractable ill-posed inverse problems [18]. Regularization involves stabilizing the solution to inverse problems by imposing additional constraint that bias the solution and/or using a priori value. In this section we will limit ourselves to Tikhonov regularization that is commonly applied to tomographic inverse problems.

4.5.1 Tikhonov Regularization

In Tikhonov regularization, we solve a damped least square problem and select a solution that minimizes the sum of norm of residuals and the norm of the solution, (Eq. 4.5.1).

$$min \left\|Am - d\right\|_2^2 + \alpha^2 \left\|m\right\|_2^2 \tag{4.5.1}$$

Alternatively we might want a solution that minimizes the sum of norm of residuals and norm of some other measure of the solution e.g. the first or the second derivative of the solution, Eq. (4.5.2), as described in Section 4.4.2 [18, 22]:

$$min \left\| Am - d \right\|_2^2 + \alpha^2 \left\| Lm \right\|_2^2 \tag{4.5.2}$$

where L a matrix representing the derivative (see Section 4.5.3). If L is the identity matrix I, Eq. (4.5.2) is reduced to Eq. (4.5.1) . The constant in the preceding two equations, α is called *regularization parameter* and should be greater than zero. The regularization parameter can be used to tune the balance between the residual and solution norms. In the damped least square approach there is no simple method of selecting the regularization parameter. If α is chosen to be too small, the reconstruction will be dominated by large and high frequency components of the data. On the other hand if α is chosen to be too large important information in the data will be suppressed [21]. The indeterminacy of the regularization parameter can be eliminated by the methods of *L-curve* [18, 23] and the *Generalized Cross Validation (GCV)* [19, 21, 23].

4.5.2 Zero-order Tikhonov Regularization

In zero-order Tikhonov regularization we apply matrix a decomposition to find solutions involving Eq. (4.5.1). Any $m \times n$ matrix, such as A, can be written as the product of three matrices [17].

$$A = U \Lambda V^T \tag{4.5.3}$$

where U is an $m \times m$ matrix of eigenvectors that span the data space and V is an $n \times n$ eigenvectors that span the model parameter space. Both U and V are unitary matrices, i.e.

$$U^T U = I, \qquad V^T V = I \tag{4.5.4}$$

The matrix Λ is an $m \times n$ diagonal eigenvalue matrix whose diagonal elements are non-negative, arranged in decreasing order and are called *singular values*, s_i . Some of the

singular values can be zero. Consequently we can partition Λ into a sub matrix Λ_p of p non-zero singular values and several zero matrices.

$$\Lambda = \begin{bmatrix} \Lambda_p & 0 \\ 0 & 0 \end{bmatrix} \tag{4.5.5}$$

where Λ_p= is a $p \times p$ diagonal matrix. Substituting Eq. (4.5.5) into Eq. (4.5.3) and using the property in Eq. (4.5.4) the Singular Value Decomposition (SVD) becomes

$$A = U_p \Lambda_p V_p^T \tag{4.5.6}$$

where U_p and V_p consist of the first p columns of U and V, respectively. The other portions of U and V are canceled by the zeros in Λ. As solved in Section 4.4 the damped least square solution of Eq. (4.5.1) is

$$m_\alpha = \left(A^T A + \alpha^2 I^2\right)^{-1} A^T d \tag{4.5.7}$$

Eq. (4.5.7) is found by minimizing the length of the 2-norm of sum of the residual and the solution (zero-order differential of m). Hence this solution is known to be *zero-order Tikhonov Regularized* solution. Substituting the SVD of A in to Eq. (4.5.7) we obtain the solution to be

$$m_\alpha = \sum_{i=1}^{k} \frac{s_i^2}{s_i^2 + \alpha^2} \frac{U_i^T d}{s_i} V_i \tag{4.5.8}$$

where k is minimum of m and n, so that all singular values are included. If we want to reject zero-valued singular values k would be p. U_i and V_i are the i^{th} columns of U and V, respectively. The quantities

$$f_i = \frac{s_i^2}{s_i^2 + \alpha^2} \tag{4.5.9}$$

are called filter factors. For $s_i >> \alpha$, $f_i \approx 1$, and for $s_i << f_i$, $f_i \approx 0$. For singular values between these two extremes, as the s_i decreases, the f_i decreases monotonically.

4.5.3 Higher-order Tikhonov Regularization

In our damped least square solution of inverse problems we minimized a function $\Phi(m)$ involving a solution norm $\|m\|_2$. This leads to an SVD based zero-order Tikhonov regularization. However, in many situations we would prefer to obtain a solution that minimizes some other measures of m, such as the norm of the first or second derivative. If we want to minimize a solution involving the first derivative we would have an equation of the form

$$\Phi(m) = r^T r + \alpha^2 \frac{\partial m^T}{\partial s} \frac{\partial m}{\partial s} \tag{4.5.10}$$

If we discretized our problem using simple collocation and our model is one-dimensional then, the first derivative becomes

$$\frac{\partial m(s_j)}{\partial s} = \frac{m(s_{j+1}) - m(s_j)}{\triangle s} \tag{4.5.11}$$

This leads to the discrete differentiation matrix

$$L = \frac{1}{\triangle s} \begin{bmatrix} -1 & 1 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots & -1 & 1 \end{bmatrix} \tag{4.5.12}$$

Hence the first derivative of the model parameter is approximated, to a multiplicative constant, by Lm. In Eq. (4.5.11) Lm is a finite difference approximation proportional to the first derivative of m. In a similar manner if we minimize the a solution involving second derivative for our one-dimensional problem,

$$\frac{\partial^2 m(s_j)}{\partial s^2} = \frac{m(s_{j+1}) - 2m(s_j) + m(s_{j-1})}{(\triangle s)^2} \tag{4.5.13}$$

Then, the discrete differentiation matrix becomes

$$L = \frac{1}{(\triangle s)^2}\begin{bmatrix} 1 & -2 & 1 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 1 & -2 & 1 & 0 & 0 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 1 & -2 & 1 \end{bmatrix} \tag{4.5.14}$$

In the above finite difference approximation of the first and second derivative, the model is assumed to be one-dimensional. If our problem is two dimensional or three dimensional then the differentiation matrix would have another form. For a two dimensional problem for which we want to minimize a function involving the second derivative, we employ a fine difference approximation to the Laplace operator. The effect L will have on the solution is that it gives a smooth solution than using the zero-order Tikhonov regularized solution [21, 22].

Damped least square inversion by minimizing a function involving the norm of first derivative of the model parameter is *first order Tikhonov regularization*. Where as, damped least square inversion involving the second derivative of the model parameters is *second-order Tikhonov regularization*. Both regularizations are termed as *higher-order Tikhonov regularization*. Then the Tikhonov regularized solution becomes

$$m_{\alpha,L} = \left(A^T A + \alpha^2 L^T L\right)^{-1} A^T d \tag{4.5.15}$$

For zero-order Tikhonov regularization, where $L = I$, we employed the SVD. In higher order Tikhonov regularization we can employ the generalized version of SVD known as Generalized Singular Value Decomposition (GSVD) [18, 21, 22]. Using the GSVD, the solution to Eq. (4.5.15) is expressed as a sum of filter factors times singular vectors. Assuming that A is m by n and rank of L be p, there exists matrices U, V, Λ, M, and X with the properties: U is m by n with orthonormal columns, V is p by p and orthonormal X is n by n and nonsingular. Λ is p by p diagonal matrix with elements λ_i so that

$$0 \leq \lambda_1 \leq \lambda_2 \leq \ \ldots \ \leq \lambda_p \leq 1 \tag{4.5.16}$$

M is a p by p diagonal matrix with elements μ_i with the property

$$1 \geq \mu_1 \geq \mu_2 \geq \ ... \ \geq \mu_p \geq 0 \tag{4.5.17}$$

The λ_i and μ_i are normalized so that

$$\lambda_i^2 + \mu_i^2 = 1 \qquad\qquad \text{`}i = 1,\ 2,\ ...\ p \tag{4.5.18}$$

The relation among the different matrices is

$$A = U \begin{bmatrix} \Lambda & 0 \\ 0 & I \end{bmatrix} X^{-1} \tag{4.5.19}$$

$$L = V \begin{bmatrix} M & 0 \end{bmatrix} X^{-1} \tag{4.5.20}$$

$$X^T A^T A X = \begin{bmatrix} \Lambda^2 & 0 \\ 0 & I \end{bmatrix} \tag{4.5.21}$$

$$X^T L^T L X = \begin{bmatrix} M^2 & 0 \\ 0 & 0 \end{bmatrix} \tag{4.5.22}$$

When $p < n$, the matrix L will have a nontrivial null space. Using the GSVD, the solution to Eq. (4.5.15) will become

$$m_{\alpha,L} = \sum_{i=1}^{p} \frac{\gamma_i^2}{\gamma_i^2 + \alpha^2} \frac{U_i^T d}{\lambda_i} X_i + \sum_{i=p+1}^{n} (U_i^T d) X_i \tag{4.5.23}$$

where the subscript U_i and X_i for the matrices U and X respectively in Eq.(4.5.23) represent column vectors; $\gamma_i = \lambda_i / \mu_i$ and are the generalized singular values. The ratio

$$f_i = \frac{\gamma_i^2}{\gamma_i^2 + \alpha^2} \tag{4.5.24}$$

are filter factors for the GSVD, analogous to those obtained earlier in the expression for the zeroth-order Tikhonov regularized solution Eq.(4.5.8).

4.5.4 Methods for Selecting the Regularization Parameter

As mentioned earlier, the two most common methods for choosing the regularization parameter, α, in Tikhonov regularization are the L-curve and Generalized Cross Validation (GCV).

The L-curve method is the most convenient graphical tool for selecting α. When plotted on a log-log scale, the curve of optimal values of $\|m\|_2$ versus $\|Am - d\|_2$ often takes on a characteristic L-shape as shown in Fig. 4.1. For higher order Tikhonov regularization

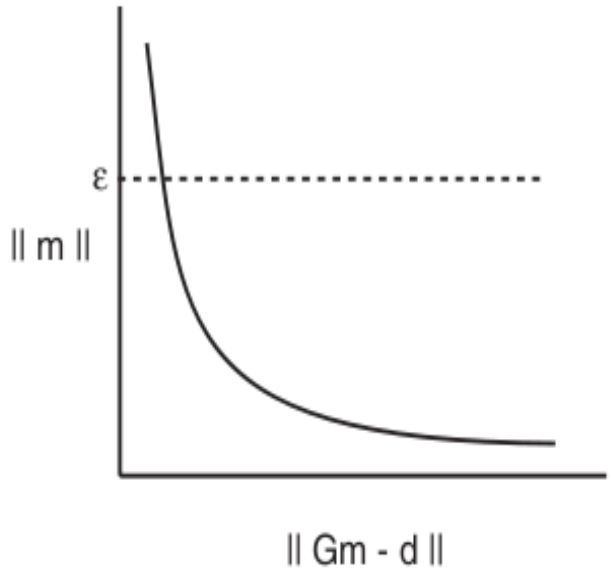

Figure 4.1: A particular model L-curve to determine the regularization parameter. (from [18]).

we plot $\|Lm\|_2$ versus $\|Am - d\|_2$. The L-shape results from the fact that $\|m\|_2$ or $\|Lm\|_2$ in general is strictly decreasing function of α and $\|Am - d\|_2$ is strictly increasing function of α. The sharpness of the "corner" of the L-curve varies from problem to problem, but is frequently well defined [18]. For this reason the curve is called an *L-curve*. In the L-curve method we select the value of α that gives the solution closest to the corner of

the L-curve.

The Generalized Cross Validation (GCV) method of selecting the regularization parameter is based on the philosophy that if an arbitrary data point d_k is left out then the corresponding regularized solution should predict the observation well [19, 22]. If d_k is left out, the regularization problem becomes

$$min \quad \sum_{i \neq k} (Am_i - d_i)^2 + \|Lm\|_2^2 \tag{4.5.25}$$

Calling the solution to this problem $m_{\alpha,L}^{[k]}$, where the superscript indicates that d_k was left out, we select the regularization parameter α so as to minimize the predictive errors for all k.

$$min \quad A(\alpha) = \frac{1}{m} \sum_{k=1}^{m} \left(\left(Am_{\alpha,L}^{[k]} \right)_k - d_k \right)^2 \tag{4.5.26}$$

However, computing $A(\alpha)$ involves solving m problems of the form in Eq. (4.5.26). Generalized Cross validation (GCV) simplifies this computation and leads to the function

$$A(\alpha) = \frac{m \left\| Am_{\alpha,L} - d \right\|_2^2}{Tr\left(I - AA^I\right)^2} \tag{4.5.27}$$

Details of the calculation can be found in references [18, 21, 22]. Where the denominator denotes the trace of the matrix in the parenthesis and A^I is a matrix which produces the regularized solution when multiplied with d i.e. $m_{\alpha,L}^{[k]} = A^I d$. The value of α that minimizes Eq. (4.5.27) can minimize the Tikhonov regularized solution as the size of the data, m, goes to infinity [18].

Chapter 5

TOMOGRAPHY

5.1 Introduction

Tomography is essentially a class of inverse problems in which the term roots from two Greek words *tomos* meaning "to section" and *graphein* "to write" [18]. Meaning it is the process of forming an image of an object sliced into sections from measurements made along rays passing through the sliced object. Tomography is related to recovering a function or parameter from knowledge of line integrals of the function over a collection of lines.

5.2 The Tomography Method

In tomography we image the interior of an object using data from rays passing through it. The Method of tomographic imaging is very popular in medicine. For example, the Computerized Axial Tomography (CAT) Scanner is an x-ray imaging device that has revolutionized brain diagnosis and other medical conditions [17].

In the physical model for tomography, the geometric ray theory in the high frequency limit is valid so that wave energy traveling between source and receiver can be considered to be propagating along narrow paths. We determine a model from path integrated properties of these rays such as attenuation (e.g x-ray, radar, seismic) travel time (e.g electromagnetic waves), or source intensity (e.g. proton emission) [19]. Along the ray

path in the medium to be imaged, let the slowness at a point $\vec{r}$ be $S(\vec{r})$, and the ray path length is known, then the line integral along the ray path is the travel time given by [18]:

$$t = \int_l S(\vec{r}(l))dl \tag{5.2.1}$$

As the ray passes through the medium, it can change direction due to refraction or reflection. In the simplest case where these effects are negligible, ray path can be approximated as straight lines.

5.3 Ionospheric Tomography

As briefly discussed in the second chapter, the Earth's ionosphere is the region above the topmost layer of the earth's atmosphere. It is formed primarily due to photo ionization of the neutral atmosphere. Consequently, depending on the ionization and recombination rates electrons, ions and neutral atoms coexist in the ionosphere.

The ionosphere being an ionized medium has a dispersive effect on electromagnetic waves passing (trying to pass) through it. Due to this dispersive nature of the ionosphere the satellite signal is delayed substantially as it traverses through. On the GPS signals the phase of the carrier wave is advanced and wave group is delayed. As we have seen in section 3.5, the line integral of the electron density along the GPS ray path from the satellite to the receiver is the total electron content TEC given by

$$TEC = \int_{raypath} N(\vec{r})d\vec{r} \tag{5.3.1}$$

where $N(\vec{r})$ is the electron density. Computerized Tomography (CT) that has revolutionized the field of medical diagnostic imaging is now being used to get high quality images of a section of human body [21]. The mathematical foundation of CT was laid as early as 1917 by Radon [21, 22], and is recently applied to study the ionospheric electron density distribution. Austen et al. (1986) [21, 22, 27] first proposed the application of

CT to reconstruct ionospheric structures from TEC data. Application of CT to the ionosphere that follows Austen is said to be Computerized Ionospheric Tomography (CIT). CIT involves reconstructing the electron density N in the inverse problem in Eq. (5.3.1) using TEC data at different viewing angles through the ionosphere. TEC being the total number of electrons in a tube of unit cross-sectional area along the line of sight can show the temporal, latitudinal and longitudinal variation of ionospheric structure . However, it cannot provide any information about the vertical profile of the ionosphere[23]. In CIT we reconstruct the vertical ionospheric electron density profile employing appropriate regularization in the inversion process.

TEC data for inversion can be measured from a satellite to a receiver on the ground like GPS and/or from satellite to satellite occultation observation [28]. In satellite to ground observation data is easily available and has good horizontal resolution. To the contrary, satellite to satellite observation has good vertical resolution and data coverage is global including oceans and poles [28].

The CIT method of imaging ionospheric electron density distribution has strengths than the conventional point measurements. One advantage of which is as instruments are added, the amount of data grows rapidly. Under ideal circumstances, the amount of data grows as the product of the number of satellites and number of receivers. That is many independent data can be obtained with a relatively small number of instruments.

GPS satellite navigation provides the potential for investigating the ionosphere. Ionospheric total electron content derived from GPS has long been an important means of studying the ionospheric disturbance under geomagnetic storms environment. Since Austen et al proposed the CIT technique to study the ionosphere, Kunitsyn (1997) first confirmed the feasibility of applying this technique by using high orbit satellites such as GPS [4].

5.4 Ionospheric Tomography Algorithms

Ionospheric tomography involves two basic steps. One is the calculation of total electron content (TEC) and the other step is model inversion in which a set of model parameters, the electron density, is derived. In the inversion for the electron density, a number of theoretical and experimental CIT algorithms have been used so far in the literature. Some of these algorithms are the Algebraic Reconstruction Technique (ART), the Multiplicative Algebraic Reconstruction Technique (MART) [22, 23], Maximum Entropy Method (MEM) [18, 23], Tikhonov regularized Damped Least Square Method (DLSM) [21, 23, 29], the Truncated Singular Value Decomposition (TSVD) and Truncated Generalized Singular Decomposition (TGSVD) [21, 23]. Raymund (1995) has made a comparative assessment of various CIT algorithms and deduced that no single algorithm can be considered as the best. Some algorithms that reconstruct well in one case can do very poorly in other cases. Some algorithms that do well in some parts of the ionosphere, but not so well at other parts [21, 22]. A combination of two algorithms can produce better results than employing a single algorithm alone [25]. Hence, proper choice of the algorithm will be the prime requisite for meaningful CIT reconstruction.

Chapter 6

GPS DATA PROCESSING AND RESULTS OF TOMOGRAPHIC IMAGING

6.1 Introduction

This chapter deals with TEC calculation from GPS receivers spatially distributed all over Ethiopia and two more receivers from Eritrea close to the Ethiopian border. Then an explanation of the experimental setup made for the tomographic electron density reconstruction in both two and three dimensions (2D and 3D) is given in detail. This will be followed by the tomographic inversion process and analysis of results in both 2D and 3D.

6.2 TEC Calculation from GPS Receivers over Ethiopia

Code pseudorange and carrier phase measurements recorded by GPS receivers all over Ethiopia have been used in TEC calculation. As described in Section 3.6, Eq. (3.6.14) gives the TEC derivation relation for each GPS satellite receiver pair. However, in Eq. (3.6.14) the receiver and satellite hardware biases are not taken into account. These instrument biases implicitly exist in pseudorange measurements and affect the line of sight ionospheric delay measurements derived from differencing dual frequency pseudorange data [16, 30]. Zhizhao Liu and Yang Gao, [30], and references there in show that the receiver bias can be as large as 10 nanoseconds and satellite bias can be as large as a

few nanoseconds. To remove these biases, let us call the relative delay on the satellite as transmitter group delay (t_{TGD}) and the delay caused by the receiver inter frequency bias (t_{IFB}) then

$$\begin{aligned} t_{IFB} &= \delta t_r(f_2) - \delta t_r(f_1) \\ t_{TGD} &= \delta t_s(f_2) - \delta t_s(f_1) \end{aligned} \tag{6.2.1}$$

where t_r denotes the travel time through the receiver and t_s the travel time through the satellite hardware both frequency dependent. Typical values for t_{IFB} are from $-20ns$ to $20ns$ [16]. Modifying Eq. (3.6.14) (see Eqs. (3.6.10 −14) for details),

$$I_2 - I_1 - \frac{N}{f_2} + t_{TGD} + t_{IFB} = \frac{A}{c}\frac{f_1^2 - f_2^2}{f_1^2 f_2^2} TEC \tag{6.2.2}$$

But $I_2 - I_1 - \frac{N}{f_2} = \frac{L_{GU}}{f_2}$, and using the bias free mapped TEC, this equation is reduced to

$$\frac{L_{GU}}{f_2} + t_{TGD} + t_{IFB} = \frac{A}{c}\frac{f_1^2 - f_2^2}{f_1^2 f_2^2}\frac{VTEC}{E(\theta)} \tag{6.2.3}$$

The transmitter group delay, t_{TGD}, is broadcast within the navigation message of the GPS satellite and hence known from ephemerids files. At any epoch a single receiver records 4 to 12 satellites, denoting the individual satellites with an index j, Eq. (6.2.3) becomes

$$L_{GU}^j + f_2 t_{TGD}^j = K\frac{VTEC^j}{E(\theta)^j} - f_2 t_{IFB} \tag{6.2.4}$$

where $K = \frac{A}{c}\frac{f_1^2 - f_2^2}{f_1^2 f_2}$, $VTEC$ and t_{IFB} are still unknowns. Eq. (6.2.4) in its current form cannot be solved. However, when we map the line of sight measurement to the vertical, each measurement from high elevation satellites should yield the approximate same TEC value assigned to the location above the site of the receiver. We assume the same TEC and the system of equations in Eq.(6.2.4) reduced to

$$L = Ax \tag{6.2.5}$$

where

$$A = \begin{pmatrix} \frac{K}{E(\theta)^1} - f_2 \\ \frac{K}{E(\theta)^2} - f_2 \\ \frac{K}{E(\theta)^3} - f_2 \\ \vdots \end{pmatrix} \tag{6.2.6}$$

The data vector L contains the unbiased observables for each satellite L_{GU} corrected for the transmitter group delay and x is the vector of unknowns given by

$$x = \begin{pmatrix} VTEC \\ t_{IFB} \end{pmatrix} \tag{6.2.7}$$

Eq. (6.2.5) is inverted for the unknown as using a simple least square inversion technique [16].

6.2.1 Required Input Data

In this thesis work, four data inputs are used to calculate the ionospheric TEC over Ethiopia. The first data input is GPS data from all receivers shown in Fig. 6.1 with geographic locations given in Table 1. GPS data are usually provided in Receiver INdependent EXchange (RINEX) format and are named *ssssdddf.yyd.Z*, where *ssss* is a 4-character station name, identifier *ddd* the day of the year of first recored, *f* the file sequence number with in day. Files having the value of *f* equal to 0 means the file contains all the existing data of the current day. *yy* denotes the last two digits of the year. A typical GPS RINEX data file consists of a header with information about the site such as receiver type, receiver version, antenna type, station coordinates, etc. Next to the header is the data section containing for each epoch the time of the epoch, satellites with observations, and the observables [31]. RINEX GPS files are compressed with the Hatanaka algorithm and must be decompressed for reading [16].

The second input file is the International GNSS Service (IGS) precise orbit data containing the positions of the GPS satellites for a specific day. IGS provides the positions of GPS satellites in "special products 3" (sp3) format [15, 32]. IGS orbit file is named as *igswwwww.sp3.Z*, with *wwwww* denoting GPS week where the first four *w*'s are GPS week number and the last '*w*' is day of week; 0 stands for Sunday and 1-6 denote the days from Monday-Saturday in their respective order.

Another input file for TEC calculation is the IONosphere map EXchange (IONEX) file. IONEX file is a post processed data from the Center for Orbit Determination in

Europe (CODE) from the University of Bern. It has better estimates for the transmitter group delay, t_{TGD}.

The last optional input file is the broadcast ephemerids file that can be used for extracting t_{TGD}.

6.2.2 Spatial and Temporal Variations of TEC

TEC has been calculated for GPS receivers spatially distributed over Ethiopia and two additional receivers from Eritrea. Fig. 6.1 (or Table 1) shows the location of GPS receivers installed by different institutions over the desired region.

Table 1. Position of GPS receivers over Ethiopia.

Station name	Latitude (^{0}N geographic)	Longitude (^{0}E geographic)
armi	6.0029	37.5609
robe	7.0673	40.0261
nazr	8.5129	39.2906
damy	9.3629	42.0339
bdar	11.5278	37.3597
dasm	11.7056	41.0092
asab*	12.9833	42.6544
damk	13.3991	39.4828
dake	13.3992	39.4827
asma*	15.2492	38.9102

* *GPS receivers located in Eritria.*

A TEC calculating software from Purdue University [16], has been used to calculate the ionospheric TEC for all GPS stations shown in Fig. 6.1 from year 2007 to the end of 2010. Input data have been downloaded from different file transfer protocol (ftp) home pages. GPS RINEX observation files for sites shown in Fig. 6.1 are available at UNAVCO [33], precise IGS satellite position files for each day can be downloaded from SOPAC [32] and IONEX file for extracting the satellite transmitter group delay is obtained from CODE, University of Bern (UB) [34].

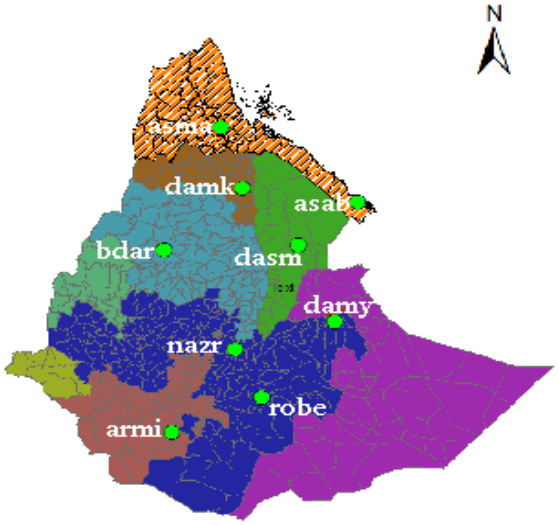

Figure 6.1: Spatial distribution of GPS receivers located over Ethiopia and two more receivers from Eritria close to the Ethio-Eritria border.

TEC has been processed in two ways. First, for each GPS receiver shown in Table 1 (or Fig. 6.1), TEC has been calculated for each GPS satellite available at a time with interval (epoch) of 30 seconds for 24 hours. Eq. (3.6.14) corrected for transmitter group delay (t_{TGD}) is used initially and hence the TEC calculated is corrected for phase ambiguity and satellite bias. The TEC obtained is then followed by correction for receiver hardware delays using satellites at high elevation, Eq. (6.2.5). TEC corrected for phase ambiguity, satellite and receiver hardware delays is then obtained for all GPS satellites under observation at a GPS receiver at an epoch of 30 seconds. This TEC obtained for a GPS satellite to receiver pair is the measurement that we should use as observation input data for tomographic inversion to retrieve ionospheric electron density. Fig. 6.2 illustrates the slant TEC calculated for each GPS satellite available for 24 hours on July 11, 2008 for GPS sites designated as ARMI, NAZR, ROBE and DAMY; and located at Arbaminch, Nazret, Bale Robe and Alemaya respectively.

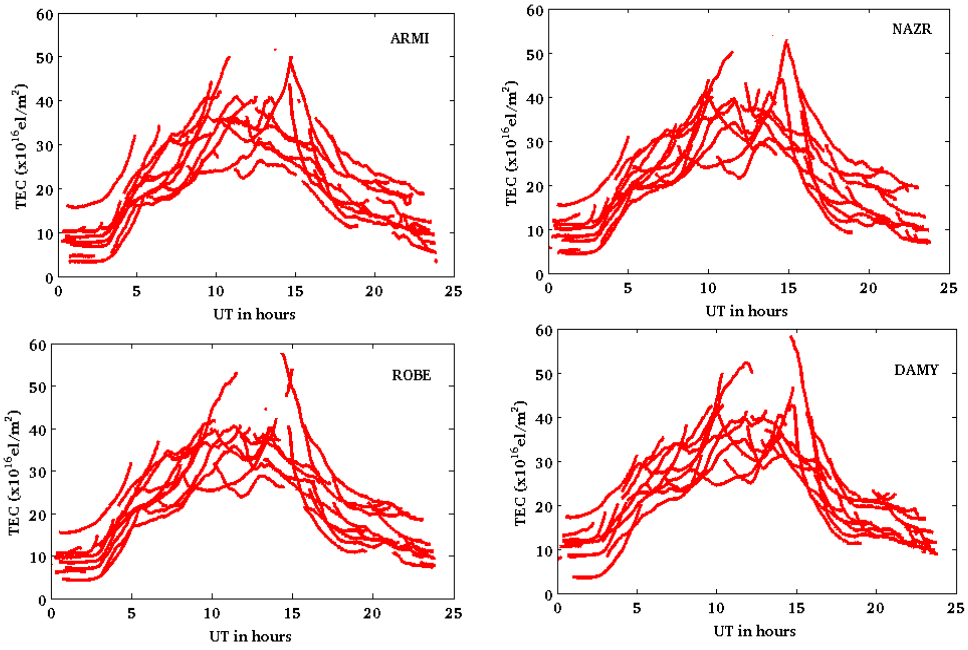

Figure 6.2: Slant TEC calculated for each GPS satellite available at any time per GPS receiver on July 11, 2008.

The slant TEC mapped to the vertical (Vertical Total Electron Content (VTEC)) is calculated for each GPS receiver using Eq. (6.2.4) and Eq. (6.2.5). Fig. 6.3 shows the slant TEC mapped to the vertical over the GPS receiver for a single day. Monthly VTEC plots for some of the GPS receivers is shown in Fig. 6.4. In the TEC calculation satellites below an elevation of 15^0 have been rejected to safeguard the effect of multipath and the ionospheric shell height is taken to be 350 km. Elevation cut of 15^0 is the widely used cutoff for GPS data [30].

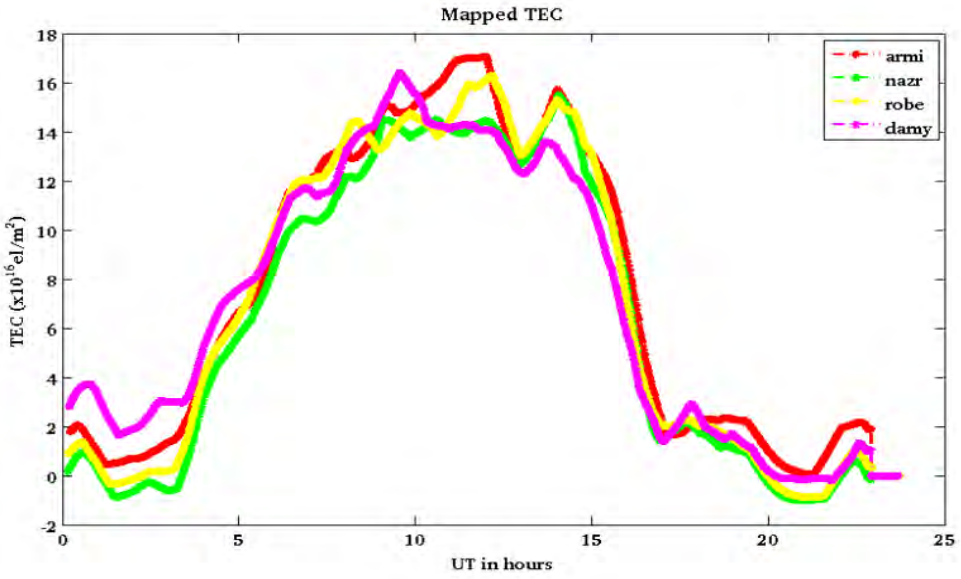

Figure 6.3: Mapped TEC to the vertical (VTEC) for four GPS receivers on July 11, 2008.

Ionospheric TEC exhibits generally two types of variation. The first is the temporal variation that describes the variation of TEC with time. The temporal variation can be further classified as diurnal variation, day to night variation of TEC due to Earths rotation and seasonal variation that describes TEC variation due to the position of the sun.

Figs. 6.2 - 6.4 depict the diurnal variation of TEC for four GPS receivers over Ethiopia. Night time electrons are much lower than day time values because of higher recombination

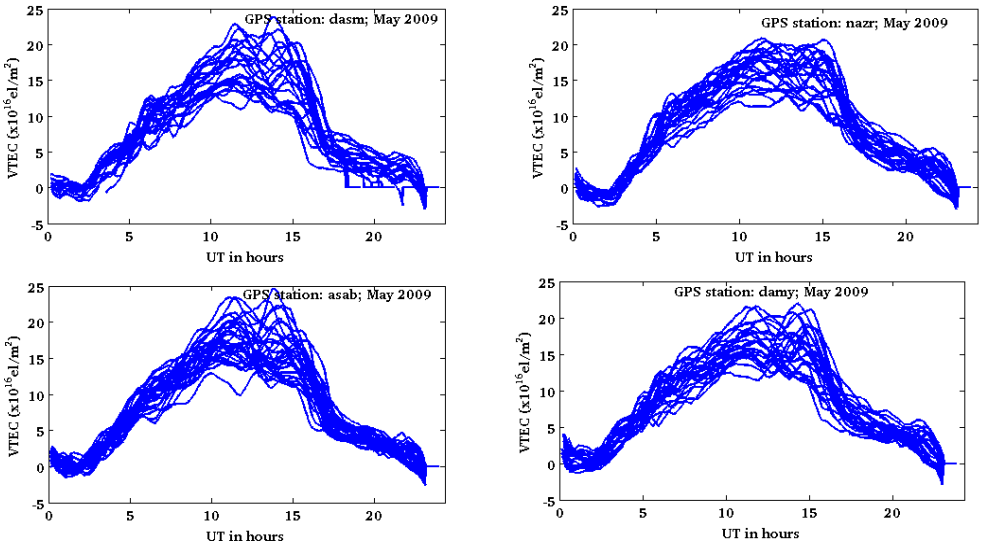

Figure 6.4: Monthly mass plot of VTEC for four GPS receivers for May 2009. The VTEC for each receiver clearly exhibit the diurnal variation

rates in the absence of radiation source. For all GPS sites day time electrons typically reach maximum value at about 11UT which is the Ethiopian local ionospheric noon.

Seasonal variation of ionospheric TEC is from the fact that the Sun is vertically above different geographic locations at different times of the year. At equinox (March 21, September 23) the sun is vertically above GPS receivers at the equator and on solstice dates (December 21, 21 June) the sun is at the tropics of Capricorn and tropics of Cancer respectively. Fig. 6.11 is a plot of the daily mean vertical TEC versus time in GPS days for GPS sites NAZR and ASMA for two years. It clearly illustrates the seasonal variation of VTEC. For both stations the VTEC gets its peak starting at about GPS day of year 300 which at this dates the sun is coming to the equator (September 23 equinox) on its way to the southern hemisphere. Another peak VTEC is observed at around March 21 equinox. In both cases the peak VTEC is due to strong vertical illumination from the Sun over the Ethiopian ionosphere that results higher ionization rates and hence higher VTEC for GPS receivers NAZR and ASMA. Fig. 6.11 shows a VTEC data missed from

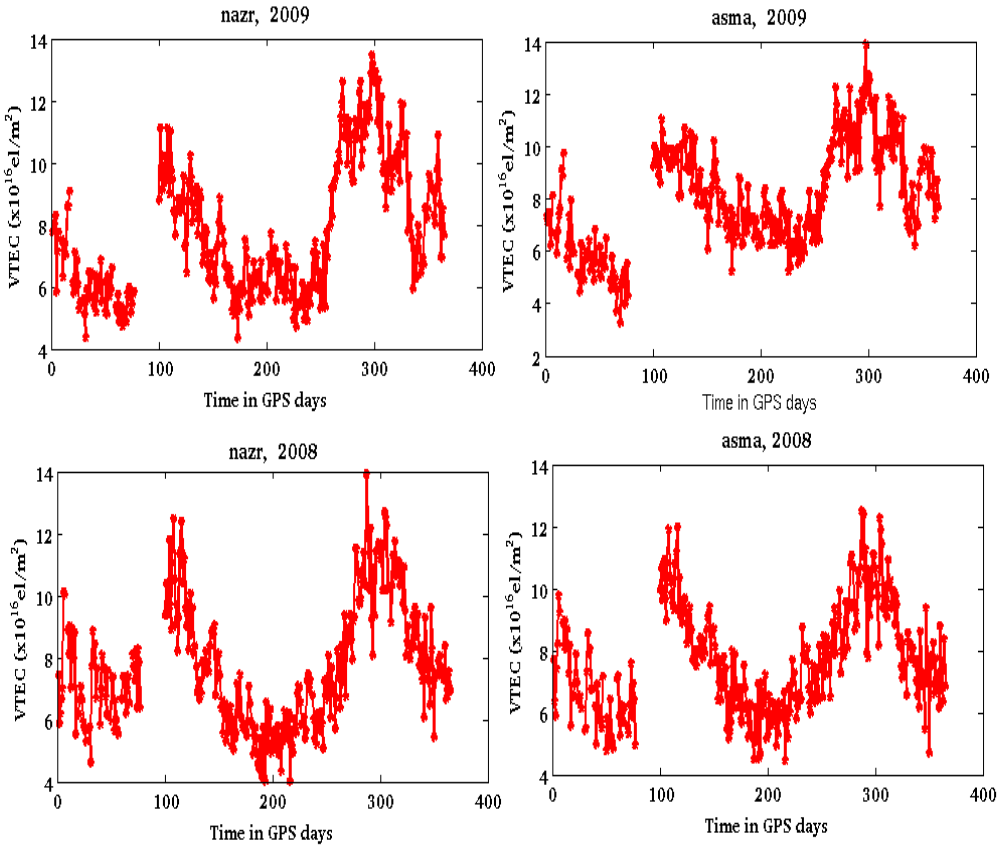

Figure 6.5: Daily mean value of VTEC plotted versus GPS days for 2008 and 2009 for GPS receivers NAZR and ASMA. Peak VTEC occurs at around equinox dates

the 90^{th} GPS day to the 120^{th} GPS day.

In addition to the temporal variation, ionospheric TEC also show spatial variations: latitudinal and longitudinal variations. Latitudinal variation of ionospheric electron densities is due to the sun's position relative to the Earth's atmosphere. The solar zenith angle, χ, the angle measured from an observer local vertical to the sun, determines the intensity of ionization where locations with small zenith angles are exposed to higher radiation rates. Latitudinal and longitudinal variations of TEC are better observed if we have receivers located at large range in latitude and longitude respectively and depends on the position of the Sun with respect to the Earth.

6.3 Ionospheric Tomography Experiment over Ethiopia

Ionospheric Total Electron Content from a network of GPS receivers can show the temporal, seasonal and spatial variations of ionospheric electron distribution. But it does not convey any information about electron density profile along the vertical direction. The aim of this thesis work is not simply deriving ionospheric TEC over Ethiopia and monitor the spatio-temporal variations. Indeed, the main objective of this thesis work is to setup a tomography experiment and invert the ionospheric electron density profile first in 2D followed by in 3D full dimensions of space. Consequently two types of experiment have been done over the Ethiopian ionosphere. The first is the 2D ionospheric tomography experiment to invert the ionospheric vertical electron density profile on a plane along a relatively fixed longitude. The second experiment is the 3D tomography experiment to invert the ionospheric electron density over Ethiopia in full dimensions of space.

6.3.1 2D Experimental Setup

In two dimensions (2D), an imaginary vertical plane at fixed longitude of 40.03^0 is selected over the Ethiopian sky. The altitude range of the plane is from 100 km to 900 km and, therefore, it is essentially with in the Earths ionosphere. The latitude range is limited

with in the range $3^0N - 15^0N$. Next, this plane is discretized into small rectangular planes called pixels of 83.49 km by 100.00 km. The number of ionospheric vertical discretization depends on the resolution requirement and density of ionospheric GPS measurements [22]. Consequently, near altitudes of high ionospheric concentration (200 km − 400 km) pixels are made to have smaller size of 83.49 km by 50.00 km for better resolution. Table 2 summarizes the geometry used for 2D ionospheric discretization process.

Table 2. Geometry used for 2D Ionospheric Discretization

Altitude range	100 km − 900 km
Latitude range	$3^0N - 15^0N$
Vertical dimension of a pixel	100 km, 50 km
Horizontal dimension of a pixel	83.486 km ($\approx 0.75^0$ in latitude)
Number of receivers	4 (See Fig. 6.9)
Receivers	armi, robe, nazr, bdar
Receiver Position (^0N)	6.0229, 7.0673, 8.5129, 11.5278
Total number of pixels	165

Eq. (5.3.1) shows that the model equation is continuous and non-linear. Subject to this discretization, for a GPS satellite to receiver ray path, Eq. (5.3.1) is reduced to a linear discrete problem given by

$$y_i = \sum_{j=1}^{n} A_{ij}x_j + e_i \tag{6.3.1}$$

For TEC measurements along many ray paths Eq. (5.3.1) becomes

$$y_{m\times 1} = A_{m\times n}x_{n\times 1} + e_{m\times 1} \tag{6.3.2}$$

where y is column matrix of m TEC measurements, x the electron density in n pixels. A is the coefficient matrix with A_{ij} distance traveled by i^{th} GPS satellite in the j^{th} pixel.

6.3.2 3D Experimental Setup

The geographic location of Ethiopia is with in the latitude and longitude ranges of 3^0N to 15^0N and 33^0E to 48^0E, respectively. For 3D ionospheric tomography experiment the volumetric region of space with latitude $3^0N - 15^0N$, longitude $33^0E - 48^0E$ and vertical height 100 km − 1100 km is divided into numerous small voxels. That is the 3D ionosphere over Ethiopia is discretized into voxels. Similar to the 2D ionospheric discretization, voxels with in ionospheric altitude range of 200 km to 400 km are made to have a vertical dimension of 50 km for better resolution and the top two slices are made to have vertical dimension of 200 km. Table 3 summarizes the geometry used for 3D ionospheric tomography over Ethiopia.

Table 3. Geometry used for 3D Ionospheric Discretization

Altitude range	100 km $-$ 900 km
Latitude range	$3^0N - 15^0N$
Longitude range	$33^0E - 48^0E$
Vertical dimension of a voxel	100 km, 50 km, 200 km
Eastward dimension of a voxel	128.26 km ($\approx 1^0$ in longitude)
Northward dimension of a voxel	102.75 km ($\approx 0.92^0$ in latitude)
Number of receivers	8 (See Fig. 6.13)
Total number of voxels	1560

Similarly, in this 3D discretization the continuous non linear problem, Eq. (5.3.1), is reduced to

$$y_{m\times 1} = A_{m\times n}x_{n\times 1} + e_{m\times 1} \tag{6.3.3}$$

where x_n is the electron density in the n^{th} voxel and A is the coefficient matrix constructed by calculating the distance traveled by the GPS signal in each of the voxels. $e_{m\times 1}$ is an error that arises due to measurement and/or discretization error.

Hence solving Eq. (5.3.1) for the electron density $N(\vec{r})$ is reduced to solving a set of linear equations in the ray tomography problem.

6.4 Analysis of Results and Discussion

Various algorithms have been employed in the literature for ionospheric electron density inversion using TEC data from satellites in orbit around the Earth. For instance, Bruce M.Howe et.al, [27], and Zhizhao Liu and Yang Gao, [30], have used spherical harmonic expansion and Kalman filter to assimilate the simulated data in to model. An algorithm based on Residual Correction Method (RCM) has been used by Eric Sutton and Helen Na, [35], and it is found to be fast, efficient, and numerically stable ionospheric tomography algorithm. GSVD based Tikhonov regularization have been employed to solve the inverse problem in the damped least square method [21, 22], and demonstrated that Tikhonov

regularized damped least square inversion is effective in that the solution is least dependent on pixel size and number of satellite ray paths. Iterative inversion algorithms such ART, IART, MART, etc. have been used for various ionospheric tomography problems with a priori value [4, 24, 35]. Others used a combination of two algorithms in which solution of the first algorithm is used as a background ionosphere for the second algorithm [25, 35, 36]. K.Bhuyan et.al, [22], and references therein describe that Raymund (1995) compared different algorithms ad deduced no single algorithm can be considered best. Hence, selecting a proper inversion algorithm is prerequisite in ionospheric tomography. In this work we made a comparative assessment among ART, Tikhonov regularized damped least square inversion and Maximum Entropy Method (MEM). The next two subsections give a comparison and validation between the ART and Tikhonov regularized damped least square inversion algorithms using numerical TEC in 2D and 3D. GSVD is used to decompose the Matrices in Tikhonov regularized damped least square inversion and the algorithm with better reconstruction efficiency is applied to realistically invert ionospheric electron density over Ethiopia in both 2D and 3D.

6.4.1 2D Algorithm Validation

In 2D, a set of ionospheric electron density distribution for the region described in Section 6.3.1, with the exception that the size of each pixel is taken to be equal 77.06 km by 100.00 km with the latitude range limited between $3^0N - 12^0N$, have been obtained at the center of each pixel from the International Reference Ionosphere, IRI2007 model on July 11, 2008 at 3UT. Then the actual position of the GPS satellites have been used to calculate the distance traveled by the signals in each pixel crossed in a selected reference frame for four ground receivers (see Fig. 6.10). This makes the coefficient matrix, A. Then a simulated TEC data is produced by multiplying the electron density derived from IRI2007 model by the coefficient matrix A. The simulated TEC is then inverted by using the Kaczmarz's ART algorithm. As discussed in Section 4.4.3, the Kaczmarz's ART

algorithm is given by Eq. (4.5.3). The ART algorithm usually fail to give a plausible result if an a priori value (background ionosphere) is not introduced [24, 25, 26]. In this case 80% of the IRI2007 model ionospheric electron density is used as a priori value.

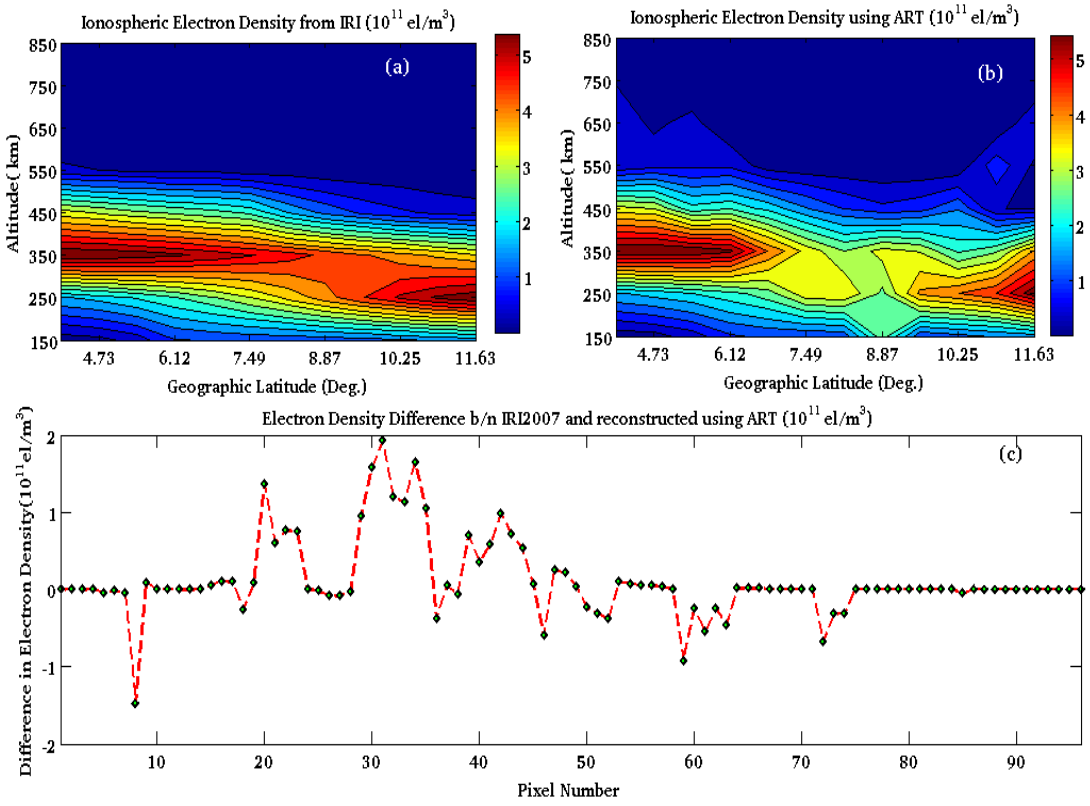

Figure 6.6: Validating the ART algorithm for 2D ionospheric tomography. (a). Ionospheric electron density obtained from IRI2007, (b). Reconstructed electron density using the ART algorithm (c). Electron density difference between IRI2007 and ART reconstructed electron density.

Fig. 6.6a shows the ionospheric electron density obtained from IRI2007 model and the reconstructed electron density using ART is shown in Fig. 6.6b. Finally the electron density difference between IRI2007 model and the reconstructed electron density using ART is given in Fig. 6.6c. From Fig. 6.6c we observe that the norm of the difference in electron density for some pixels can reach up to $2 \times 10^{11} el/m^3$.

The Simulated TEC obtained from IRI2007 model and inverted by ART is used as

measurement for damped least square inversion with higher order Tikhonov regularization. As discussed in Chapter 4, the damped least square inversion approach with higher order Tikhonov regularization is to find the electron density x that minimizes Eq. (4.5.2) and the solution to this problem given by Eq. (4.5.15) Applying GSVD to A and L given by Eqs. (4.5.19) and (4.5.20), Eq. (4.5.15) become

$$x_{\alpha,L} = \sum_{i=1}^{p} \frac{\gamma_i^2}{\gamma_i^2 + \alpha^2} \frac{U_i^T d}{\lambda_i} X_i + \sum_{i=p+1}^{n} (U_i^T d) X_i \tag{6.4.1}$$

That is, the solution x is a function of the regularization parameter α and the matrix L. The solution varies with the regularization parameter quite strongly as compared to L. If α is chosen to be too small, the reconstruction will be dominated by large, high frequency components. If α is chosen to be too large, important information in the data will be suppressed [21]. L will have different forms for different orders of Tikhonov regularization, Section 4.5.3. K. Bhuyan et. al (2004), [21], and K. Bhuyan et. al (2002), [22], have used the GCV method to determine the regularization parameter α for first order Tikhonov regularization where L will be given by Eq. (4.5.12). In the second order Tikhonov regularization where L is given as Eq. (4.5.14) large gradients of ionospheric electron density are heavily penalized and has disadvantage in using it for a disturbed ionosphere. For a disturbed ionosphere it is recommended to use first order Tikhonov regularization [21]. In this research we have used L-curve method to determine α and second order Tikhonov regularization with L given by Eq. (4.5.14).

Applying the GSVD to Eq. (6.4.1) is based on the method of regularization of the problem that eliminates the need for truncation of the singular values. GSVD based CIT algorithm in Tikhonov regularized solutions has advantages in that it creates electron density distribution satisfactorily and has the ability to reconstruct the electron density with out the help of any initial guess regarding the state of the ionosphere [21]. The Simulated TEC from IRI2007 model in 2D is then inverted using the second order Tikhonov regularization with α determined by the L-curve method.

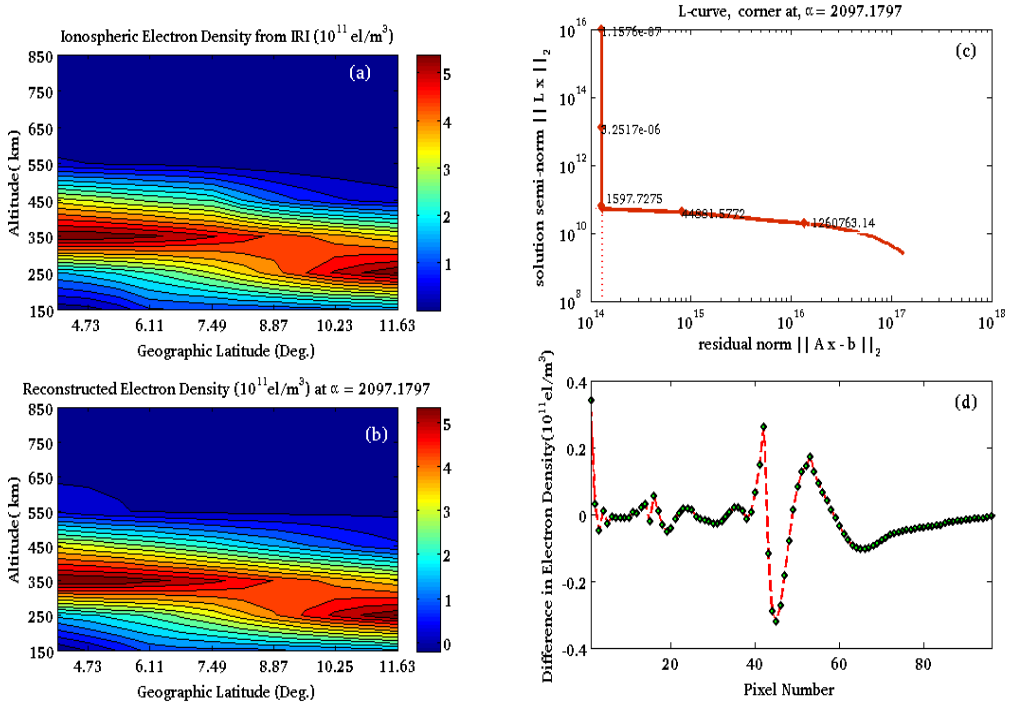

Figure 6.7: Validating the damped least square inversion for 3D ionospheric tomography. (a). Ionospheric electron density obtained from IRI2007, (b). Electron density inverted using damped least square method with second order Tikhonov regularization, (c). L-curve to find α, (d). Electron density difference between the IRI2007 and the inverted electron density.

Fig. 6.7a shows contour plot of ionospheric electron density distribution on a vertical plane at longitude 40.03^0E over Ethiopia on July 11, 2008 at 3UT obtained from IRI2007. Fig. 6.7b is contour plot of the reconstructed electron density using damped least square inversion algorithm with second order Tikhonov regularization where the simulated TEC is used as measurement input. Comparing these two figures it can be seen that the algorithm reconstructed the ionospheric structure very well. The L-curve used to locate the regularization parameter is depicted in Fig. 6.7c. The regularization parameter is found to be 2097.18 and falls at the corner of a perfect L-curve. Finally, Fig. 6.7d illustrates the electron density difference between the IRI2007 and the inverted electron density at $\alpha = 2097.1797$. It can be seen that the difference between electron densities is minimum falling close to zero except some points between the 40^{th} and 60^{th} pixels where higher gradients of electron density exist. This shows that the damped least square inversion algorithm with second order Tikhonov regularization reconstructed the 2D ionospheric electron density with sufficiently high accuracy than the ART method of inversion (see Fig. 6.6 and Table 4).

Table 4. Norm of difference (misfit) between IRI2007 electron density and reconstructed electron density in 2D.

Altitude (in km)	Norm of misfit (Tikhonov reconstruction) $\times 10^{11} el/m^3$	Norm of misfit (ART reconstruction) $\times 10^{11} el/m^3$
150	0.0432	0.1414
250	0.0261	0.3395
350	0.0173	0.8373
450	0.1479	0.4247
550	0.0766	0.2048
650	0.0867	0.1665
750	0.0429	0.0553
850	0.0160	0.0075

Table 4 shows the norm of the difference (misfit) between the numerical electron density obtained from IRI2007 model and the inverted electron density by damped least square method with second order Tikhonov regularization and ART method of reconstruction averaged for the pixels at equal altitudes in 2D ionospheric tomography algorithm validation. From this table it is generally observed that the second order Tikhonov regularization in the damped least square inversion produces solution with smaller misfit than ART reconstruction. Maximum contribution for the misfit comes from rows of pixels at 450 km altitude for damped least square inversion with second order Tikhonov regularization and at rows of pixels at altitude 350 km for the ART method.

The relative reconstruction error can be defined as [21, 36, 37],

$$\sigma = \left(\frac{\|x_{simu} - x_{recons.}\|^n}{\|x_{simu}\|^n} \right)^{1/n} \tag{6.4.2}$$

where x_{simu} denote the simulated ionospheric electron density, and $x_{recons.}$ reconstructed ionospheric electron density. n represents the various norms. In our 2D validation x_{simu} is the electron density obtained from IRI2007 and $x_{recons.}$ is reconstructed or inverted ionospheric electron density. Using the 2-norm ($n = 2$), in the damped least square inversion with second order Tikhonov regularization σ is found to be 0.04 and $\sigma = 0.22$ for the ART method of reconstruction. Hence, we conclude that the damped least square inversion with second Tikhonov regularization produces results with better accuracy than ART.

6.4.2 3D Algorithm Validation

In a similar manner, an equivalent numerical experiment has been performed to validate the damped least square inversion with second order Tikhonov regularization algorithms for 3D ionospheric tomography. However, in this case numerical ionospheric electron density for the region discussed in Section 6.3.2 is obtained by inverting TEC with damped least square inversion with second order Tikhonov regularization algorithm itself along 1437 number of GPS ray paths on the same date but for 20 minute observation at an epoch

of 30sec close to 10UT (from 9.83UT to 10.16UT). A simulated TEC is then obtained by multiplying the the electron density inverted by the coefficient matrix calculated from the actual position of GPS satellites at this time. Eight ground receivers all from Ethiopia as shown in Fig. 6.13 have been employed.

In validating the damped least square inversion with second order Tikhonov Regularization for the 3D ionospheric tomography, the simulated TEC obtained by multiplying the numerical electron density by the coefficient matrix is inverted using Eq. (6.4.1) with α determined by the L-curve method. Different amount of maximum random TEC noise has been added to the simulated TEC data before inversion.

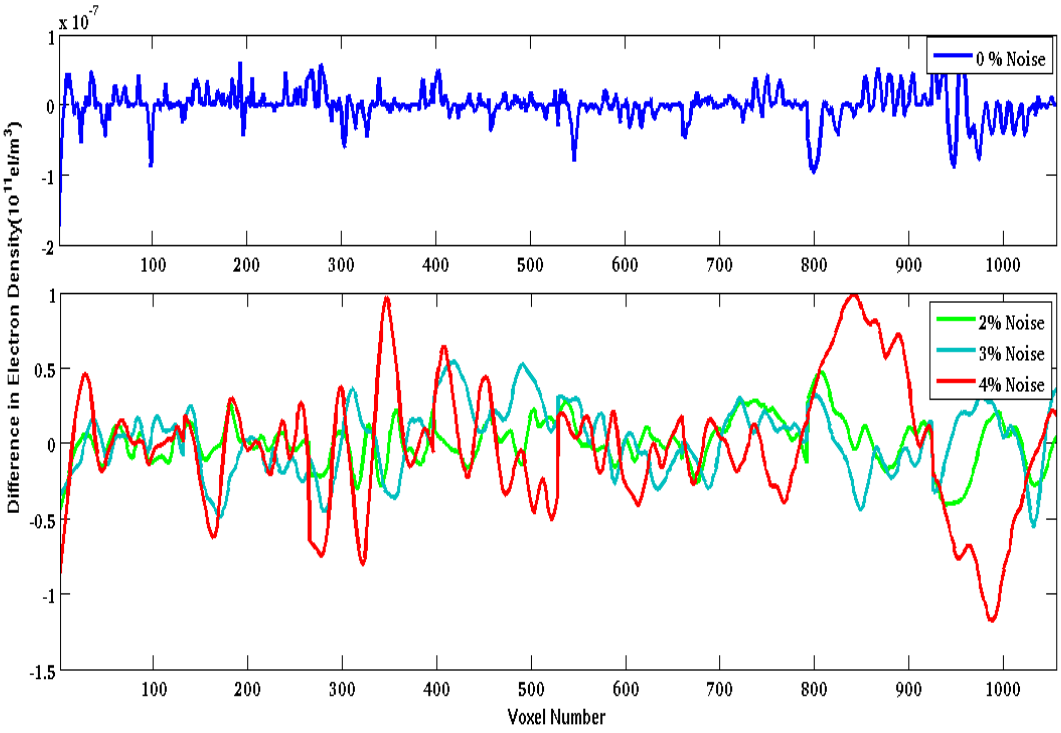

Figure 6.8: Validating the damped least square inversion for 3D ionospheric tomography. Difference between ionospheric electron density used for simulated TEC calculation in the forward model and the inverted electron density with damped least square inversion with second order Tikhonov Regularization.

Fig. 6.8 illustrates the difference between the electron densities used for the simulated TEC calculation and the reconstructed electron density with different maximum percentage TEC noise added to the simulated TEC. The top subplot of Fig. 6.8 is the difference

in the electron densities with no noise added to the simulated TEC. We clearly see that the difference in the electron densities for the voxels is close to zero with a value of order -7 in IED units. That is the inversion algorithm recovers the electron density with high accuracy when the measurement slant TEC is free from noise. The bottom subplot in Fig. 6.8 is the electron density difference with a maximum of 2%, 3% and 4% random TEC noise added to the simulated TEC data before we invert it. It can be observed from Fig. 6.8 that as the percentage noise level increases the norm of the difference in the electron density deviates increasingly from zero. These results demonstrate that the Tikhonov regularized damped least square inversion is highly feasible for ionospheric electron density reconstruction.

Table 5. Norm of Difference (misfit) between numerical electron density and reconstructed electron density in 3D .

Altitude (in km)	Norm of misfit for Tikhonov reconstruction ($\times 10^{11}$ $electrons/m^3$)
150	0.0945
225	0.0590
275	0.1309
325	0.0929
375	0.0738
450	0.1558
550	0.1590
650	0.2040
800	0.1318

Table 5 shows the norm of the difference (misfit) between the numerical simulated and the inverted electron density by damped least square method with second order Tikhonov regularization averaged for the voxels at a fixed altitudes in the 3D ionospheric tomography algorithm validation. In this inversion a maximum of 2% random TEC noise has been added to the simulated numerical TEC. The maximum contribution to the misfit

comes at 275 km and high altitudes (450 km-800 km) voxels. The reconstruction error is found using Eq. (6.4.2) with $n = 2$. In this 3D case the the damped least square inversion with second order Tikhonov regularization reconstructs the simulated electron density with reconstruction error 0.02. We observe that the damped least square inversion with second order Tikhonov regularization has inverted the simulated TEC with good accuracy even with 2% noise added.

6.4.3 2D Ionospheric Tomography Results and Discussion

In Section 6.4.1 we have verified that the damped least square inversion with second order Tikhonov regularization method has effectively recovered the IRI2007 model electron density with small reconstruction error than ART. The damped least square inversion with second order Tikhonov regularization minimizes the norm of the solution residuals plus the norm of the second derivative of the electron density and is applied to realistically invert the 2D ionospheric electron density over the Ethiopian ionosphere on a vertical plane at a fixed longitude of 40.03^0E as described in Section 6.3.1. In this process slant TEC of four GPS receivers close to the vertical plane at 40.03 have been taken as shown in Fig. 6.9. As we see these four receivers are not located at the same longitude which is the ideal geometry for 2D ionospheric tomography and as a result there is a time difference between the actual position of receivers and the plane at 40.03^0E. Fortunately, the ionosphere varies much more rapidly with latitude than it varies with longitude and consequently the time difference can be ignored. However, to eliminate this time difference we have used the relation

$$\Delta T = \frac{L_{40} - L_{GPS}}{15} \tag{6.4.3}$$

where ΔT is the time difference in UT, L_{40}, the longitude 40.03^0E, L_{GPS} the longitude value of the coordinate position of the receiver. In the application of GPS for CIT, horizontal rays that are important for vertical resolution are absent [28, 38]. GPS signals at low elevation can contribute for vertical resolution in the absence of horizontal rays. However, GPS signals at low elevation are prone to multipath effect and should be removed. Considering the need of horizontal rays and the effect of multipath an elevation cut of 20^0 is adopted for all slant TEC reading at all times for all GPS to receiver pairs. Though, the TEC is calculated at 30 second epoch, for our 2D tomography, TEC at an epoch interval of 60 seconds for not more than 10 epochs have been taken and the distance

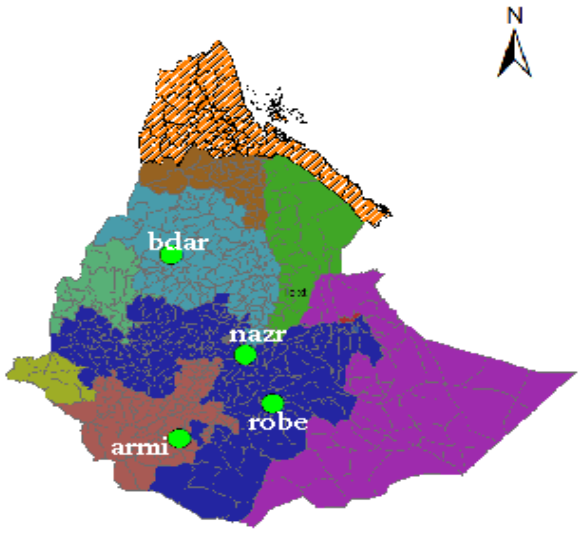

Figure 6.9: Location of GPS receivers used for 2D tomography.

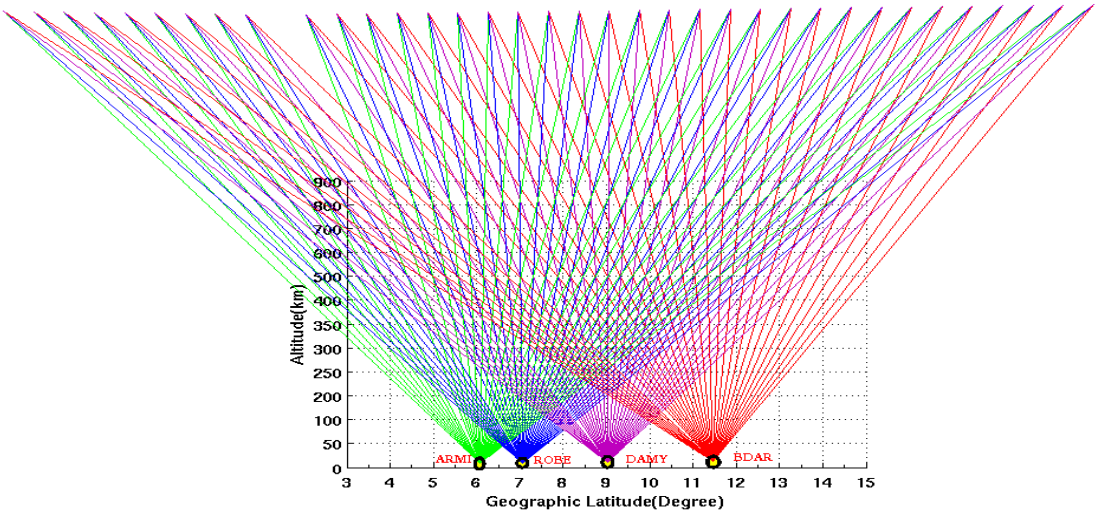

Figure 6.10: Schematic diagram of GPS signal paths to receivers on the ground used for 2D tomography.

traversed by each GPS signals through all pixels (see Fig. 6.10) are calculated at this time intervals, with a zero value for pixels not crossed by the signal. We have made an important assumption that the ionosphere does not vary for short period of time. Considering the electron density with in each pixel to be constant and concentrated at the center, the ionospheric electron density in two dimensions for pixels described in 6.3.1 have been inverted with second order Tikhonov regularization in the damped least sense. L-curve method is employed to find the regularization parameter.

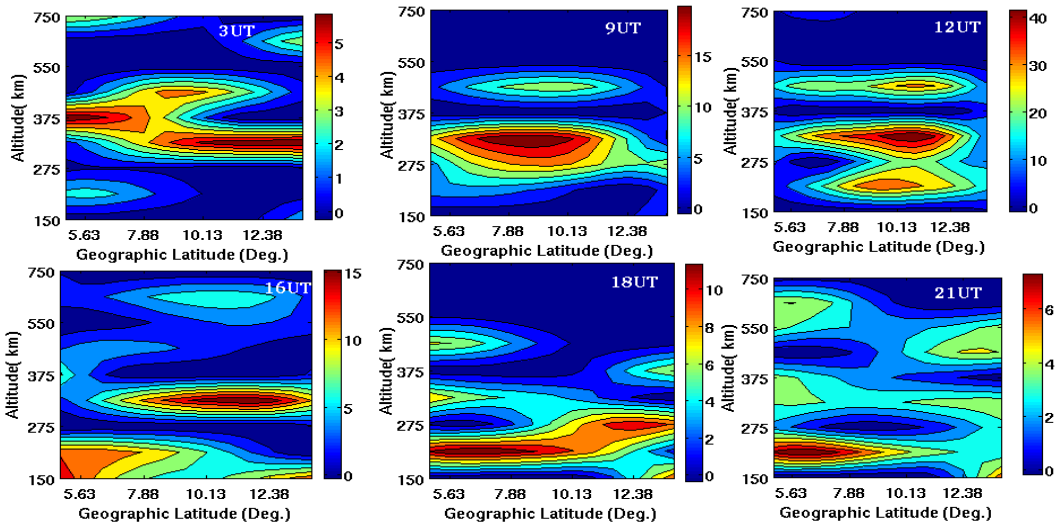

Figure 6.11: Samples of 2D ionospheric electron density inverted on 11^{th} July, 2008 over Ethiopia on a vertical plane at 40.03^0E longitude for different Universal Times.

Fig. 6.11 shows contour plot of the 2D ionospheric electron density over Ethiopia on a vertical plane (Section 6.3.1) at 40.03^0E longitude at different universal times on July 11, 2008. The numbers at the top right corner of each subplot represent the time in UT for which the electron density has been inverted and can be converted to the local (Ethiopian) Universal time by adding 3 hours. From Fig. 6.11, one can easily see that, on average the ionospheric peak electron density (F-layer) at an altitude of 350 km. However, the height of peak electron density is found to drop to about 200 km at 18 UT and 21 UT. We

observe that as the time evolves from 3UT to 12UT, the ionospheric electron density over the study area increases followed by a decrease from 12UT to 21UT. The Ethiopian noon is at about 9UT and since the ionosphere lags behind the solid Earth by about 2 hours in its co-rotation with the Earth, we expect the ionospheric noon on Ethiopia to be at 11UT. Hence, this is the reason for the maximum electron density to be found at 12UT, one hour after the Ethiopian ionospheric noon. The electron density values over Ethiopia shown in Fig. 6.11 agrees well with the day-night and temporal variation of ionospheric electron density

6.4.4 3D Ionospheric Tomography Results and Discussion

The 2D ionospheric tomography is conventional as it assumes a 2D vertical plane in a 3D ionosphere with the polar orbiting GPS satellites overhead the receivers on fixed longitude ideally. Hence, 2D CIT is weak in ionospheric electron density recovery and biases from the physical reality [28, 30]. The 3D ionospheric tomography describes the ionosphere in the full dimensions of space and consuently will have an apparent strength than 2D ionospheric tomography. In four dimensions the time evolution of the 3D ionospheric structure can be studied [28].

In this study, the 2D ionospheric tomography experiment over Ethiopia has been successfully extended to 3D ionospheric tomography incorporating latitude, longitude and altitude in the full dimension of space. That is, the ionosphere with in the geographic latitude and longitude ranges of Ethiopia and altitude from 100 km to 1100 km has been discretized into three dimensional slices, voxels. Details can be found in Section 6.3.2. The big challenge in 3D ionospheric tomography is determination of the distance traveled by each ray from a GPS satellite to a receiver (see Fig. 6.12) with in each voxel crossed. In this thesis, the GPS local coordinate system has been used in a selected 3D reference frame to calculate the distance traveled by the signal with in each voxel accurately. Voxels not crossed by a signal will have a zero value and consequently the coefficient matrix is

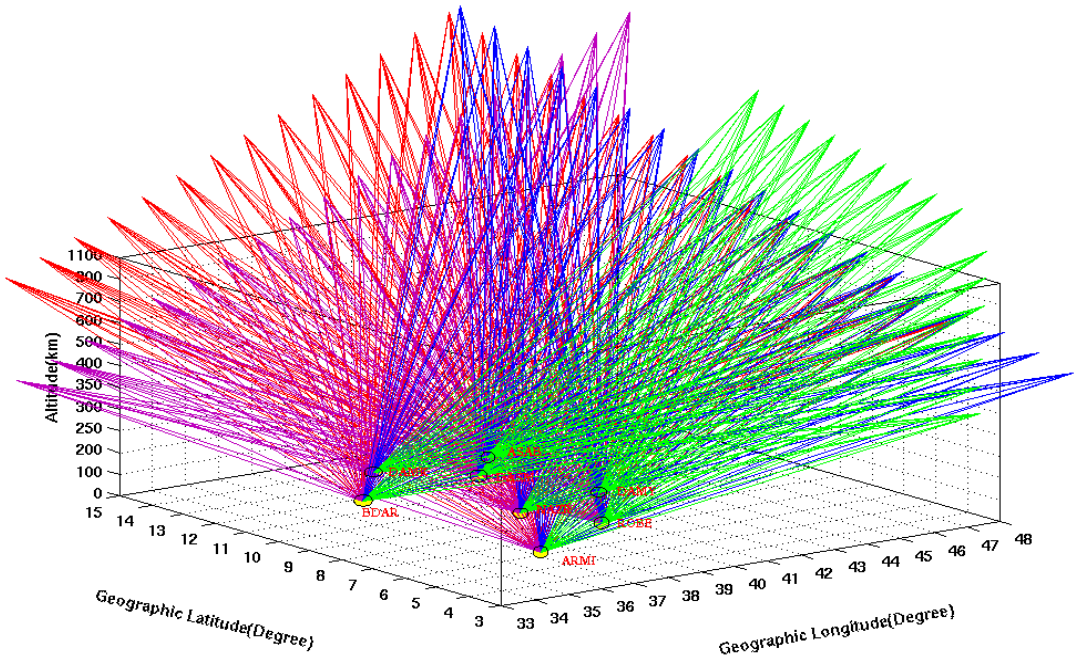

Figure 6.12: Schematic diagram depicting GPS satellites ray paths to GPS receivers spatially located on Ethiopia used for 3D tomographic scanning.

highly sparse, especially if we have large TEC measurement for large number of voxels.

Fig. 6.13 shows the number and spatial distribution of GPS receivers over Ethiopia employed in the measurement slant TEC for the 3D ionospheric tomography. Generally eight stations have been used in the slant TEC measurement with an elevation cut of 20^0. Satellites with line of sight elevation below 20^0 have been rejected to get rid off the multipath effect. In Section 6.4.2 we have validated that the second order Tikhonov regularized inversion in 3D recovers the electron density with high accuracy especially if the slant TEC calculation is free from any kind error (noise). Consequently, this inversion technique has been used to determine the ionospheric electron density (IED) in three dimensions over Ethiopia. Similar to the 2D ionospheric tomography the L-curve method is used to find the regularization parameter.

The problem with three dimensional spatial data is its graphic visualization in full space. Though Matlab has different 3D data visualization tools, we have found the patched

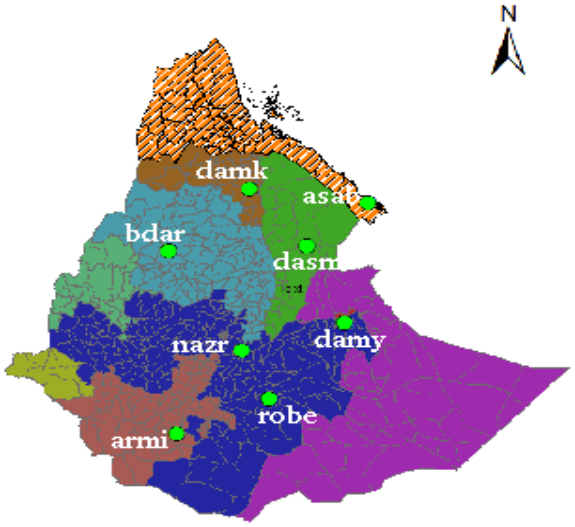

Figure 6.13: Location of GPS receivers over Ethiopia used in reading slant TEC for 3D tomography.

iso-surface graphics technique better for visualizing the 3D ionospheric electron density. Fig. 6.14 illustrates the inverted ionospheric electron density in 3D volume of space over Ethiopia at different universal times. The 3D geometry of this 3D tomographic inversion is discussed in Section 6.3.2. Each subplot in Fig. 6.14 represent patched isosurface plots with an isosurface value of 2 in IED units at different universal times as indicated at the top of each plot. That is IED below a value of 2 in IED unit are shown by empty (white) spaces with in each subplot. This graphics plot visualizes our 3D electron density better than other 3D visualization mechanisms such as 3D slices along fixed latitudes, longitudes and/or altitudes.

Carefully observing each of the sub-figures in Fig. 6.14, we see various electron density distributions over the Ethiopian ionosphere. Generally the ionospheric peak electron density occurs at an average altitude 350 km from the surface of the Earth with exceptions of the last two figures at 18UT and 21UT. As time evolves in Fig. 6.14 the electron density increases with a peak value around the Ethiopian ionospheric noon then the

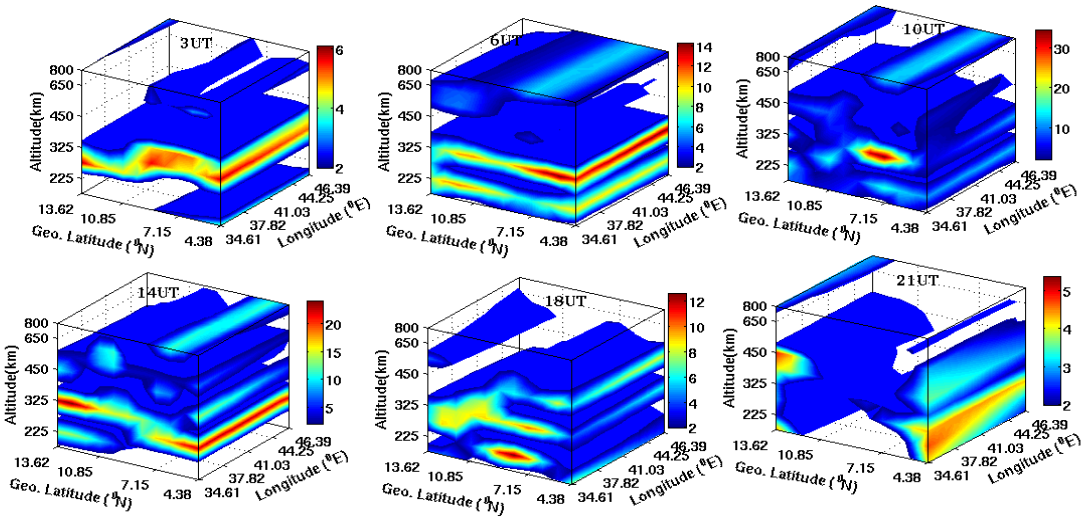

Figure 6.14: Patched iso-surface plots of 3D ionospheric electron density over Ethiopia with iso-surface value 2 in IED units at different universal times on 11^{th} July, 2008. Empty (white) spaces in each subplot represent IEDs below a value of 2 in IED units.

electron density decreases as dusk comes. This agrees with the 2D tomographic inversion and fulfills the global ionospheric reality. The second important IED distribution we see from from Fig. 6.14 is that as we go from 34^0E to 47^0E longitudinally over Ethiopia at all altitudes the ionospheric electron density exhibit no significant variation. Contrary to this , the electron density distribution over Ethiopia exhibit a high variation with latitude as we move from the lowest geographic latitude border (3^0N) of the country to the highest latitude border (15^0N). All of the spatial (latitudinal, longitudinal and vertical) and temporal (day-night) variations we see in Fig. 6.14 agrees well with the ionospheric physical fact and with the behavior of the global ionosphere.

Fig. 6.15 depicts the ionospheric electron density distribution of the 3D tomographic inversion on vertical slices at different fixed longitudes at 6UT on July 11, 2008. Fig. 6.15 a-e clearly show that ionospheric electron density over Ethiopia at the six longitudes, whose value is indicated at the top of each sub-figure, remain constant. However, Fig.

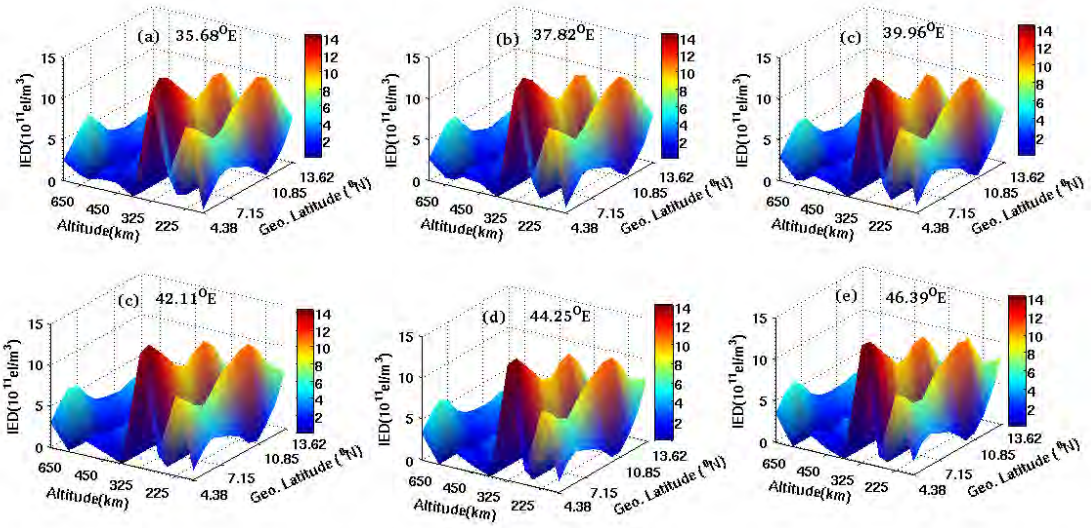

Figure 6.15: Surface plot of 3D ionospheric electron density along different fixed longitudes at 6UT (9 local UT) on 11^{th} July, 2008.

6.15 a-e show that the ionospheric electron densities along each fixed longitude vary with latitude with maximum electron density belonging in voxels at low latitudes. Another interesting feature observed in Fig. 6.15a-e is that the E-region peak ionospheric electron density is reconstructed well at the time indicated in addition to the the obvious F-region peak electron density.

Results of ionospheric electron density distribution from the 3D tomographic inversion along four fixed latitudes is shown in Fig. 6.16. From each of the subplots, Fig. 6.16a-c, one can easily see that the electron density over the Ethiopian ionosphere at fixed latitudes of 5.31^0N, 7.17^0N, 9.92^0N and 13.62^0N do not show any variation with longitude. We also see that the maximum electron density is found to be along 5.31^0N latitude and generally decreases as we move to high latitudes.

Ionospheric electron density distribution is maximum during day time than the corresponding night time. This is due to higher photo-ionization rate than the recombination

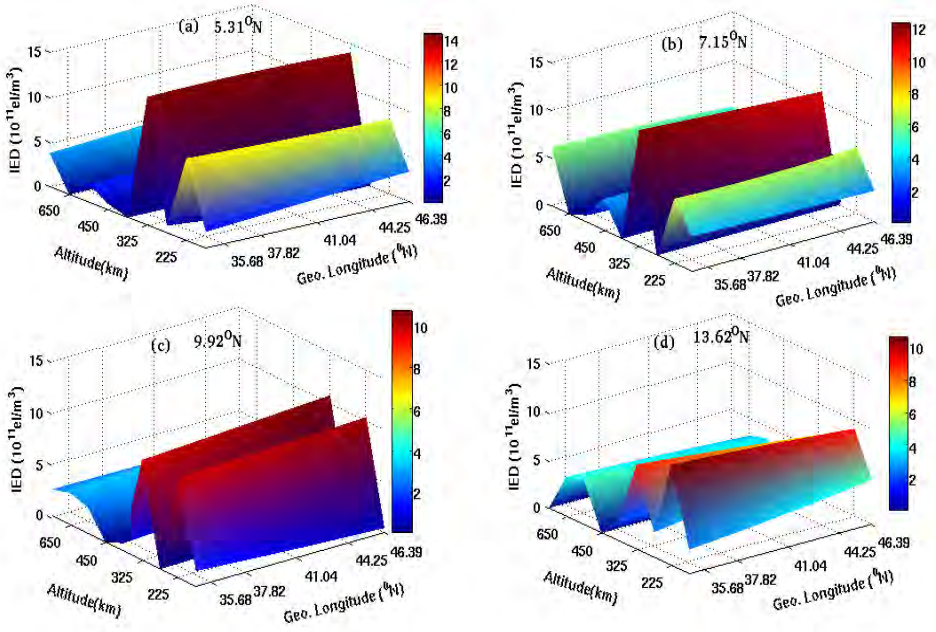

Figure 6.16: Surface plot of 3D ionospheric electron density along different fixed latitudes at 6UT (9 local UT) on 11^{th} July, 2008..

rate for the day side ionosphere with peak electron density occurring at about the ionospheric noon at a given location. The ionosphere co-rotates with Earth but lags behind by about 2 hours than the solid Earth [39]. As a result the Ethiopian ionospheric noon is expected to occur at 11UT (Ethiopian local noon occurs at 9UT).

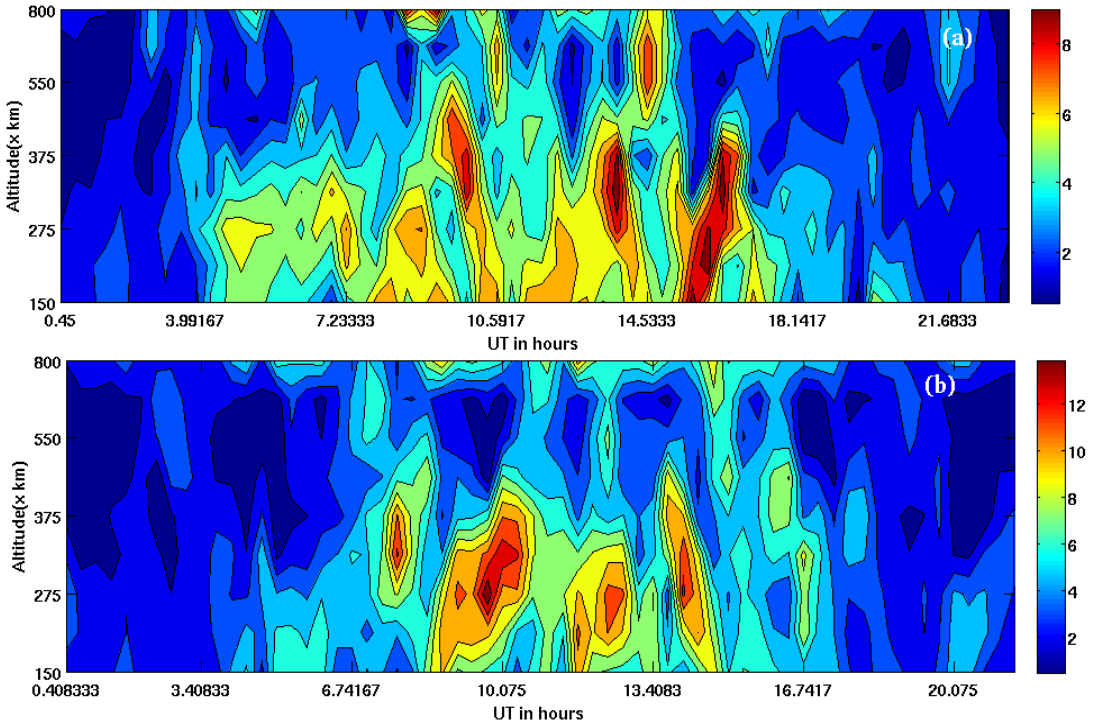

Figure 6.17: Vertically averaged 3D ionospheric electron density over Ethiopia for two days.(a) Reconstructed electron density for July 11, 2008. (b) Reconstructed electron density for May 01, 2009.

The reconstructed ionosphere in 3D has been averaged along latitude and longitude and hence we will have mean electron density at each height sliced for each observation time. Fig. 6.17 represent the mean electron density calculated at each height for two days. Fig. 6.17a depicts the mean electron density at each height for 11^{th} July, 2008; a month after the June 21 Solstice. In a similar manner the mean 3D ionospheric electron density at each height for May 01, 2009 is illustrated in Fig. 6.17b. In universal time, the Ethiopian day time is roughly from 3UT to 15UT. From Fig. 6.17a-b we clearly see that high ionospheric electron density distribution over Ethiopia occur at the day time for both dates at an average height of 350 km with a more pronounced electron density distribution at around the ionospheric noon. Moreover, the Sun is expected to be overhead for GPS receivers over Ethiopia at May 01, 2009 (41 days after the March 21 Equinox) on its way to the tropic Cancer and this is the reason for larger electron density distribution shown in Fig. 6.17b than in Fig. 6.17a. Note that in both cases the the regularization parameter is taken to be relatively fixed.

Chapter 7

CONCLUSION

Ionospheric slant TEC being the measurement along the line of sight from a receiver to a GPS satellite does not convey any information about the ionospheric vertical electron density profile. Ionospheric tomography is a recently popular method to reconstruct well the ionospheric electron density along the vertical. In CIT we invert the measurement TEC to obtain the ionospheric electron density.

In this work, TEC has been processed from GPS receivers across Ethiopia. In processing the TEC corrections have been made for phase ambiguity, satellite and receiver hardware delays. However, the TEC calculated is subject to discretization and measurement errors. Tomographic inversions in two and and three dimensions (2D and 3D) have been performed to image the ionospheric electron density distribution over Ethiopia.

For our 2D tomographic inversion, the ionosphere on a vertical plane at 40.03^0E longitude is discretized into pixels and the electron density with in each pixel is obtained using the damped least square inversion algorithm with second order Tikhonov regularization. The algorithm has been validated using simulated from IRI2007 model. It is found that the algorithm reconstructs the electron density well with high accuracy.

In the case of 3D tomographic inversion, the ionosphere over Ethiopia has been discretized into small voxels and the distance traveled by the satellite signals with in each voxel calculated setting an appropriate coordinate system. Then the ionospheric electron density at the center of each voxel is determined using damped least square inversion

algorithm. Analogous to the 2D inversion the algorithm has been validated using numerical TEC simulated by the algorithm itself with various random noise levels added. We have found that the algorithm works highly effectively with small or no noise introduced to the simulated TEC. In both 2D and 3D cases the regularization parameter has been determined by the L-curve method.

Assuming that the ionosphere does not exhibit much variation for 25 minutes, the validated algorithm has been applied to invert the TEC measurements to reconstruct the ionospheric electron density over Ethiopia in both two and three dimensions. TEC for 2D tomographic inversion has been taken from four GPS stations close to the plane at 40.03^0E longitude; Whereas in the 3D tomographic inversion TEC data from all GPS receivers over the study area has been used. Tomographic inversion results indicate that the ionosphere over Ethiopia has been reconstructed well with spatio-temporal variations in a good agreement with the realities of the equatorial global ionosphere.

Bibliography

[1] Michael C. Kelley (2009), *The Earths Ionosphere Plasma Physics and Electrodynamics*, International Geophysics Series, Academic Press,inc.

[2] Margaret G.Kilvson and Kristopher T.Russell (1995), *Introduction to Space Physics*, Cambridge University Press.

[3] John S. Seybold (2005), *Introduction to RF propagation*, John Wiley & Sons inc.

[4] Debao Wen et.al. (2007), *Ionospheric temporal and spatial variations during the 18 August 2003 storm over China*, Earth Planets Space, 59, 313-317.

[5] Attila Komjathy (1997), *Global Ionospheric Total Electron Content Mapping Using the Global Positioning System*, PhD Dissertation, University of New Brunswick, Canada.

[6] Y. Kamide and A. Chian (Eds.) (2007), *Handbook of the Solar-Terrestrial Environment*, Springer.

[7] Ja Soon Shim (2009) *Analysis of Total Electron Content (TEC) Variations in the Low- and Middle-Latitude Ionosphere*, PhD Dissertation, Utah State University, Logan, Utah.

[8] R. D. Hunsucker and J. K. Hargreaves (2003), *The High Latitude Ionosphere and its Effect on Radio Propagation*, Cambridge University Press.

[9] John D.Jackson (1962), *Classical Electrodynamics*, John Wiley & Sons inc. Third Edition.

[10] Robert W.Schunk and Andrew F. Nagy (2009), *Ionospheres Physics, Plasma Physics, and Chemistry*, Cambridge Atmospheric and Space Science Series, Cambridge University Press, Second Edition.

[11] Elliott D. Kaplan, Christopher J. Hegarty (Eds.) (2006), *Understanding GPS Principles and Applications*, Artecht House Inc, Second edition.

[12] Guochang Xu (2007), *GPS Theory, Algorithms and Applications*, Springer, Second edition.

[13] Ahmed El-Rabbany (2002), *Introduction to GPS the Global Positioning System*,Artecht House Inc.

[14] Ramjee Prasad and Marina Ruggieri (2005), *Applied Satellite Navigation Using GPS, GALILEO, and Augmentation Systems*,Artecht House Inc.

[15] T. A. Herring, R. W. King, S. C. McClusky (2009), *GAMIT Reference Manual GPS Analysis at MIT*, Department of Earth, Atmospheric, and Planetary Sciences Massachusetts Institute of Technology, Release 10.3,.

[16] Thomas Dautermann and Eric Calais (2008), *TEC Data Processing Software*, Purdue University.

[17] William Menke (1989), *Geophysical Data Analysis: Discrete Inverse Theory*, International Geophysics Series, Volume 45, Academics Press Inc, Revised Edition.

[18] Richard C. Aster, Brian Borchers, and Clifford H. Thurber (2005), *Parameter Estimation and Inverse Problems*, International Geophysics Series, Elsevier Inc.

[19] Samuli Siltanen (2009), *MAT-52506 Inverse Problems*, Lecture Note, Version 10.

[20] M.S Zhdanov (2002), *Geophysical Inverse Theory and Regularization Problems*, Elsevier Science B.V.

[21] K. Bhuyan , S. B. Singh , and P. K. Bhuyan (2004), *Application of Generalized Singular Value Decomposition to Ionospheric Tomography*, Annales Geophysicae, 22, 3437-3444.

[22] K.Bhuyan, S.B.Singh, and P.K.Bhuyan (2002), *Tomographic Reconstruction of the Ionosphere Using Generalized Singular Value Decomposition*, Research Communications, Current Science, VOL. 83, NO. 9, 1117-1120.

[23] Per Christian Hansen (2008), *Regularization Tools: A Matlab Package for Analysis and Solution of Discrete Ill-Posed Problems*, Technical University of Denmark, Denmark.

[24] Endawoke Yizengaw, Peter Dyson, and Elizabeth Essex, *Tomographic reconstruction of the ionosphere using ground-based GPS data in the Australian region*, Physics Departement, La Trobe University, Bundoora, Australia.

[25] Wen Debao, Kefei Zhang and Robert Norman, (2010), *GNSS-based tomographic reconstruction of the ionospheric electron density distribution using a combined algorithm*, Current Science, 99, 1233-1238.

[26] Debao Wen et.al., (2007), *Three-dimensional ionospheric tomography by an improved algebraic reconstruction technique*, GPS solut., 11, 251-258, Springer.

[27] Bruce M.Howe, Ray Runciman, James A. Secan (1998), *Tomography of the Ionosphere-Four Dimensional Simulations*, Radio Science, Vol. 33, No.1, 109-128.

[28] Cathryn N Mitchell, *Combining Radio Occultation Measurements with Other Instruments to Map the Ionospheric Electron Concentration*, Department of Electronic and Electrical Engineering, University of Bath, Bath, UK.

[29] E.J.Fremouw and James A. Secan, (1992), *Application of stochastic inverse theory to ionospheric tomography*, Radio Science, 27, 721-732

[30] Zhizhao Liu and Yang Gao, *Ionospheric Tomography Using GPS Measurements*, Department of Geomatics Engineering, The University of Calgary, Calgary, Alberta, Canada.

[31] ftp://igscb.jpl.nasa.gov/pub/data.

[32] http://sopac.ucsd.edu/pub/products/

[33] ftp://data-out.unavco.org

[34] ftp://ftp.unibe.ch/aiub/CODE

[35] Eric Sutton and Helen Na, (1996), *Ionospheric tomography using the residual correction method*, Departement of Electrical and Computer Engineering, University of Lowa, Lowa City, Preprint to reappear in Radio Science.

[36] Debao Wen et.al, (2008), *A Hybrid Reconstruction Algorithm for 3-D Ionospheric Tomography*, IEEE Transactions on Geoscience and Remote Sensing Vol. 46, No. 6, 1733-1739.

[37] Ersin Yavuz et.al, (2008), *Algorithms and Basis Functions in Tomographic Reconstruction of Ionospheric Electron Density*, Turkey.

[38] Wen Debao, Yuan Yunbin and Ou Jikun, (2007), *Monitoring the three-dimensional ionospheric electron density distribution using GPS observations over China*, J. Earth Syst. Sci. 116, No. 3, 235-44.

[39] BDL Opperman (2007), *Reconstructing Ionospheric TEC Over South Africa Using Signals From Regional GPS Network*, PhD Dissertation, Rhodes University.

Declaration

This thesis is my original work, has not been presented for a degree in any other University and that all the sources of material used for the thesis have been dully acknowledged.

Name: Gebreab Kidanu Zewdie

Signature:

Place and time of submission: Addis Ababa University,

This thesis has been submitted for examination with my approval as University advisor.

Name: Dr.Gizaw Mengitu Tsidu

Signature: