Ph.D. Thesis

Doctoral Program in Aerospace Science & Technology

Contributions to the 3D ionospheric sounding with GPS data

Miquel García-Fernández

Advisors:
Dr. J.Miguel Juan Zornoza
Dr. Manuel Hernández Pajares

Research group of Astronomy and Geomatics (gAGE)
Depts. of Applied Mathematics IV and Applied Physics
Universitat Politècnica de Catalunya (UPC), Spain

January 22, 2004

Manel, Miguel i Jaume: vull agraïr-vos tots els consells que m'heu donat per començar a fer ciència a partir de "*nada*". M'agradaria compartir amb vosaltres la il.lusió de ser el primer doctor que surt de gAGE.

Sabi i Xelu, aquesta tesi doctoral és un regal per vosaltres, com a infinitèssima mostra de gratitud per tot el suport entusiàstic que he rebut. Cada pas que puc fer en el coneixement respon a un valor que m'heu ensenyat: tenir inquietuds.

Mónica, todos os miles de horas de proceso que hay detrás de cada resultado, técnica ou "*algoritmo neperiano*" investigado e programado perderian o seu encanto si non foran acompañados dese teu sorriso asombrado. Gracias por aturar tantos intentos, espero que non frustrados, de explicarche en que demos consiste esta tese.

Txema, va per tots els contes, novel.les i guions de cinema que tenim planejats (i que farem algun dia, ja veuràs). Una part d'aquesta tesi també li dec als meus col.legues del I.B. Jaume Balmes: Àngel, Anna, Edu, Jordi i Rafa i també, és clar, a les respectives parelles i fills que van arribant al món.

Raul, Xevi i Angela: gràcies per tots els "No surprises" amanits amb disquisicions científiques i elaboracions de plans, tant els que ja s'han realitzat com els que queden pel futur.
Ànims pels nouvinguts a gAGE (i pels que han d'arribar encara)!!

Gostariame adicar tamen esta tese a miña familia, tanto ós que resisten na morrinha de Barcelona como ós que disfrutan con cada pinga da chuvia que cae nos prados de Castelo ou no mar das Rias Baixas. A todos, moitisimas gracias por ensinarme a importancia de "*traballar a feito*". Como vedes tirei bo proveito persoal das labouras do campo.

"*All we have to decide is what to do with the time that has been given to us.*" Gandalf the Grey

Contents

Introduction **3**

I Background on Ionospheric sounding 7

1 The ionosphere and its monitoring **9**
1.1 Structure of the ionosphere 9
1.2 Variability of Ionosphere 13
1.3 Ionosphere and electromagnetic signals 15
1.4 Ionospheric data 18

2 The Global Positioning System **27**
2.1 Fundamentals 27
2.2 Observables 29

3 Ionospheric tomography **35**
3.1 Introduction 35
3.2 Techniques 37
3.3 Limitations 44

4 Abel inversion **49**
4.1 Inversion Techniques 49
4.2 LEO GPS missions 55

II Results on algorithms for 3D ionospheric sounding 59

5 Combining Ionosonde and ground GPS data **61**
5.1 Complementarity of data 61
5.2 Model 62
5.3 Results 67

6 Improvement of Abel inversion using VTEC data 77
6.1 Including VTEC information 77
6.2 Upper Ionosphere and Plasmasphere estimation 83
6.3 Model 85
6.4 VTEC and spherical symmetry 87
6.5 Results with GPSMET data 88
6.6 Results with concurrent missions: SAC-C and CHAMP 102

Conclusions and guidelines for future research 115

APPENDIX A: Acronyms 119

List of Figures

1.1 Typical day/night Ionospheric electron density profiles 10
1.2 α-Chapman profiles 13
1.3 Sunspot Number from 1995 to 2003 15
1.4 Example Dst,A_p and K_p indices 16
1.5 Example of Real height analysis with POLAN 22
1.6 IRI and NeQuick profiles 23
1.7 Example of Global ionospheric maps 24

2.1 Schematic view of GPS signal structure 28
2.2 Example of cycle-slips in GPS Signal 30
2.3 Elevation and ionospheric delay 33
2.4 Projection from Slant to Vertical direction 33
2.5 Mapping function 34

3.1 Scheme of geometry involved in ionospheric tomography with satellite data 36
3.2 Basis functions with voxeled ionosphere 39
3.3 Block diagram of ART 41
3.4 Block diagram of Kalman filter 44
3.5 Concept of combining complementary data 46
3.6 Scheme of LEO lines-of-sight 47

4.1 Scheme of occultation geometry 50
4.2 Scheme of occultation geometry with straight line 53
4.3 Scheme of discrete inversion of STEC 55

5.1 Vertical description of electron density obtained before and after using ionosonde profiles 62
5.2 Scheme of vertical layers 64
5.3 Scheme of STEC modelisation in a ionosphere divided in voxels 65
5.4 Construction of constraints from Vertical profiles derived from ionosonde data 66

5.5 Map of IGS European GPS stations and ionosondes used for this scenario 69
5.6 Dst parameter during 1995 October 17th, 18th and 19th . . . 70
5.7 Profiles obtained combining GPS and Ionosonde data 71
5.8 Comparison with NeQuick model 74
5.9 Scheme of STEC Reconstruction for a LEO observation in a ionosphere divided in voxels. 75
5.10 Comparison with GPSMET: Single occultation 75
5.11 Comparison with GPSMET: Occultations of 1995 October 18th 76

6.1 Presence of VTEC gradients during an occultation 78
6.2 Comparison of electron density and shape functions profiles of two near locations 81
6.3 Comparison of electron density and shape functions profiles of two distant locations 82
6.4 Evolution of differences ($RMS_{profile}$ and Difference in peak) against distance and linear regression in the IRI simulated environment 84
6.5 Histogram with the number of occurrences for each bin of percentage difference of profile 85
6.6 Modeling of STEC assuming separability. 86
6.7 Constant gradients of VTEC simplify the Separability hypothesis inversion 88
6.8 Comparison of Plasmaspheric TEC estimated by F_p and the estimation with voxel model. 91
6.9 Slab thickness vs. LT for the period from 1995 October 10th to 1995 October 21st 92
6.10 Effect of slab thickness variation in ionosonde comparison of GPSMET occultations 95
6.11 Effect of co-location in ionosonde comparison of GPSMET occultations 96
6.12 GPSMET orbit in function of local time 97
6.13 Example of E layer improvement with Separability hypothesis 98
6.14 Inverted profiles using GPSMET data and separability hypothesis. The comparison is made with POLAN profiles obtained from ionogram data. 99
6.15 Latitude variation of the standard deviation of hmF2 101
6.16 Dst parameter from November 1st 2002 to November 16th 2002.103
6.17 Evolution of the RMS (of foF2 comparison) against distance for CHAMP and SAC-C satellites 105
6.18 Assumption of spherical symmetry affects topside estimation . 106

6.19 Latitude dependency of hmF2 error 108
6.20 Upper ionosphere and plasmaspheric estimation for SAC-C and CHAMP satellites 109
6.21 Co-located locations of SAC-C and CHAMP occultations in South America 112
6.22 Co-located vertical profiles of SAC-C and CHAMP in South America 113

List of Tables

2.1 Main contributions of Pseudoranges 30

5.1 Relative variation of electron density within a cell 63
5.2 Table of distances between ionosondes in kilometers 68
5.3 Comparison of estimated N_e profiles using combination with NeQuick profiles from test ionosondes 72

6.1 Expected figures of differences between Electron density profiles and Shape functions 83
6.2 Table for NmF2 and hmF2 comparison for GPSMET satellite using the IRI model 90
6.3 Table for foF2 comparison for GPSMET satellite 94
6.4 Table for foE comparison for GPSMET satellite 97
6.5 Inter-comparison between hmF2 estimators 100
6.6 Comparison with Dudeney formula for all available ionosondes. The values for the bias and standard deviation are expressed in km. 101
6.7 F2 layer critical frequency comparison for the SAC-C and CHAMP satellites 104
6.8 E layer critical frequency comparison for the SAC-C and CHAMP satellites 106
6.9 F2 layer peak height comparison for the SAC-C and CHAMP satellites 107
6.10 Shape Function peak inter-comparison for the SAC-C and CHAMP satellites 110
6.11 Shape Function peak height inter-comparison for the SAC-C and CHAMP satellites 111

Introduction

With the increasing number of Global Navigation Satellite System (GNSS) applications, the knowledge of the elements that affect their radio signals has become an important field of study. One of the most important effects is the one caused by the ionized layer of the atmosphere, the ionosphere, which causes a delay in the electromagnetic signals that journey through it, in particular to the GNSS signals. In fact, the knowledge of the state of the ionosphere is a key point in the development of navigation applications based on GNSS. Moreover, the ionosphere monitoring is important for the long-haul communications based on ionospheric reflections and, obviously, it is important by itself in order to provide with a deeper comprehension of the characteristics of this plasma and the phenomena that take place in it.

In the recent years, satellite signals, in particular those from the Global Positioning System (GPS, the American GNSS system), have become an important source to study the ionosphere. In this context, tomographic techniques similar to the ones used in medical sciences have been implemented in this field, leading to ionospheric tomography techniques, which have been the object of intense research.

This is the framework in which this doctoral thesis is included. Using GPS data as the main source of information, several techniques related to obtain a three dimensional description of ionosphere are studied.

Research objectives

The objectives of this doctoral thesis can be summarized in the main idea of obtaining a 3 dimensional description of the ionospheric electron density. Two different ways to achieve this goal are considered:

- The use of ionospheric data gathered by instruments placed in the earth surface only. If satellite data gathered by ground receivers are used alone, it appears the problem of correlations between vertical layers.

To overcome this drawback, a combination of GPS ground data with additional ionospheric data is proposed.

- Improvement of Abel inversion, which is a simple technique that allows to obtain vertical profiles of electron density from GPS observables. The second point of this thesis deals with the use of GPS data only (from ground and satellite receivers) in order to modify the classical Abel approach, which assumes that the electron density depends only on height (spherical symmetry assumption) and can lead to an important mismodelling. The main idea of this modification is the inclusion of horizontal gradients of electron density in order to obtain an improvement on the classical implementation.

Methodology

The methodology used in this work consists basically in the development of software applications (written in C, Fortran, Unix c-shell and gawk scripts) aimed at processing different types of ionospheric data. The processed data sets consist on data from ionosonde, ground and LEO GPS receivers. The programs involved in this research work are:

1. A tomographic software previously developed in the gAGE/UPC group. Modifications have been applied in order to include data from ionosonde measurements.

2. A software implementation of the improved Abel inversion. Additional scripts have been written to automate the data gathering and processing of radio occultations as well as scripts for posterior comparisons with external data (i.e. ionosonde).

Thesis structure

This thesis consists of 6 chapters, and it is divided in two parts. The first part describes the background for the research work and contains the following chapters:

- Chapter 1, *The ionosphere and its monitoring*. Contains a basic description of the ionosphere.

- Chapter 2, *The Global Positioning System*. Introduces the fundamental concepts of the main tool used in this work to sound the ionosphere: GPS.

- Chapter 3, *Ionospheric tomography*. Describes the state-of-the-art of the most common techniques used nowadays to obtain a tomographic description of the ionosphere using satellite data.

- Chapter 4, *Abel inversion*. Briefly describes the Abel transform technique used to process GPS data gathered by satellites at low orbits (Low Earth Orbiters or LEO). This technique allows providing with high resolutions vertical profiles of ionospheric electron density (as well as other parameters related to neutral atmosphere).

The second part contains the research work developed in this framework and the results obtained:

- Chapter 5, *Combining ionosonde and ground GPS data*. Describes how the use of two types of ground data (from GPS receivers and ionosonde) can provide with a vertical description of ionospheric electron content using a data driven model.

- Chapter 6, *Improvement of Abel inversion using VTEC data*. Describes a modification in the Classical Abel inversion scheme based on the inclusion of VTEC data and the estimation of the electron content above the LEO. Results are compared with the classical approach.

The thesis work ends with the *Conclusions and guidelines for future research* that contains the key points reached with this research work. It tries as well to establish a link for future guidelines based on the results obtained in this doctorate thesis.

Part I

Background on Ionospheric sounding

Chapter 1

The ionosphere and its monitoring

1.1 Structure of the ionosphere

The Earth atmosphere is divided in several layers, one of these being the plasma called ionosphere, which is generally accepted that begins at 60km approximately from the Earth surface, after the neutral atmosphere layer, and ends with the start of the plasmasphere at around 1000km.

The origin of the ionosphere is the ionization of several molecular species, being the most important the atomic oxygen (O, ionized to O^+), due to the ultraviolet (UV) and X radiation emitted by the Sun. Since the absorption of this UV radiation increases with decreasing altitude and the density of neutral atmosphere molecules increases towards the Earth surface, a maximum of absorption takes place, generating a maximum of ionization typically placed between 200 and 400km height. Considering that the overall charge of the ionosphere is 0 and since the ionization of the oxygen generates a single electron, the number of electron can be considered approximately equal to the ions. Therefore a maximum in the electron density takes place as well.

The ionosphere is generally considered to be divided in several stratified layers (see Figure 1.1), so a distinction on the source of ionization is made. From lowest to highest, the distribution is:

- D layer, from 60 to 90km, is generated basically by the hard X-radiation of the Sun.

- E layer, from 90km to 130km, is generated by soft X-radiation. This layer is present basically by day and decreases by night. Its shape can be only a simple inflexion in the profile, but a valley usually appears, thus marking a local maximum. An additional feature of this layer is the appearance of the so called Sporadic E-layer (or E_s layer), a thin layer that sometimes shows large enhancements of electron density.

- F layer is the predominant layer in an electron density profile and contains the bulk of electron densities. This layer is sub-divided in two:

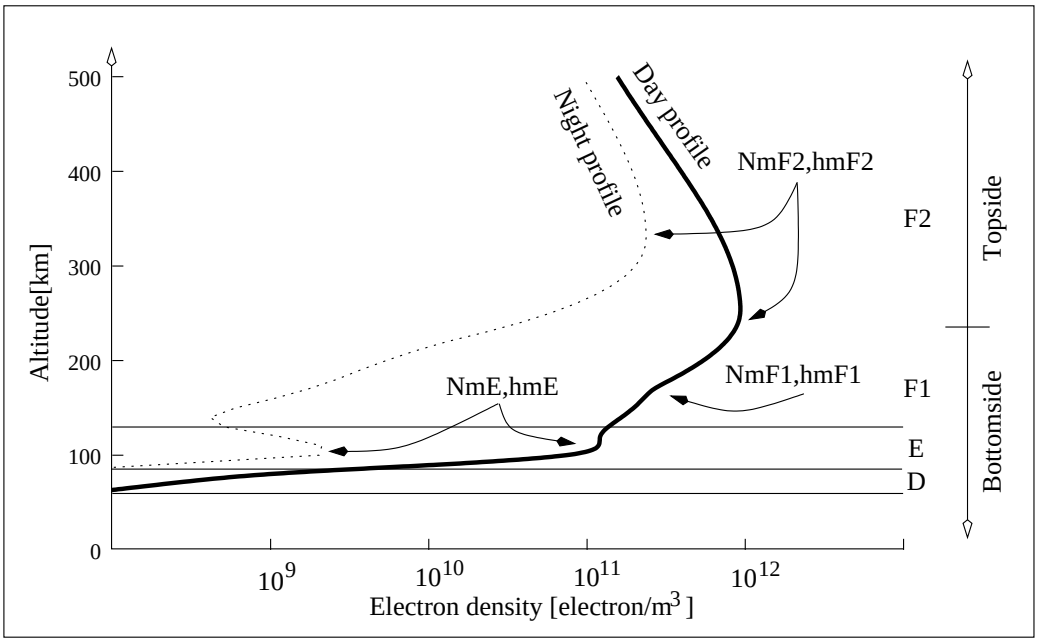


Figure 1.1: Typical day/night profiles of ionospheric electron density. E,D,F1 and F2 layers are indicated as well as the maximum of electron densities (Nm) and respective heights (hm). These profiles have been simulated with IRI model for a mid-latitude region (Barcelona, 2^oE 41.33^oN), the epoch corresponds to winter season.

- F1 layer, from 130km to 210km, which is present only during daytime and can be distinguished as an inflexion in the profile.
- F2 layer, from 210km to 1000km which contains the maximum of the electron density profile.

The Electron density maximum (NmF2) marks the border between the Bottomside (lower part of the ionosphere) and the Topside (higher part of the ionosphere).

The electron density above 1000km is mixed up in the plasmasphere, where the majority of ions correspond to H^+, therefore it is also known as protonosphere.

Chapman profile

To describe the distribution in height of the electron density, the Chapman function can be derived starting from the hydrostatic equilibrium assumption of a mass element placed at height h with respect to the earth surface.

Considering that this mass element has a density of particles n and mass m, the hydrostatic law can be derived from the force equilibrium between pressure p and gravity force (being g the gravity constant):

$$dp = -n\ m\ g\ dh \tag{1.1}$$

Assuming that the atmosphere is an ideal gas, which is a very good approximation for lower parts of the atmosphere ([Brekke, 1997]), the ideal gas expression ($p = nkT$, with k the Boltzmann constant and T temperature) is considered in order to modify the previous equation so it becomes:

$$\frac{dp}{p} = -\frac{dh}{H} \tag{1.2}$$

where H is the *scale height* and it is defined as:

$$H = \frac{k\ T}{m\ g} \tag{1.3}$$

where m is the mean molecular mass and g the Earth gravity constant. Integrating the differential equation of equation 1.2, an expression of pressure is obtained. Moreover, since the pressure and density are proportional, a similar relation with the densities can be obtained:

$$p = p_s\ e^{-\frac{h}{H}} \qquad n \propto p \quad \Rightarrow \quad n = n_s\ e^{-\frac{h}{H}} \tag{1.4}$$

where the s subscript denotes the value at the sea level.

The next step is to link the density of molecules with the radiation coming from the Sun (I_∞). Let us name $I(\lambda, h)$ the intensity of photon radiation at wavelength λ and height h. Assuming a monochromatic beam of solar radiation, $I(\lambda, h)$ becomes $I(h)$. If this radiation travels the distance ds through an atmospheric slab with an inclination of χ with respect to the zenith (for which χ=0), it will be reduced due to absorption (among other factors) by a quantity dI, which is expressed as:

$$dI = -n\ \sigma\ I\ ds \tag{1.5}$$

where the density of particles (n) is assumed to correspond to a single species and σ is the cross-section for ionizing the neutral particles. The wavelength dependency of σ is not considered since a monochromatic beam is assumed. If the radiation falls in towards the atmosphere by an angle χ , the slant distance differential ds is related to height differential dh by:

$$ds = -\frac{dh}{\cos(\chi)} = -dh \sec(\chi) \tag{1.6}$$

therefore the differential equation 1.5 becomes:

$$\frac{dI}{I} = n\ \sigma\ \sec(\chi)\ dh \tag{1.7}$$

the expression of radiation intensity can be obtained using the expression of particle density of equation 1.4 and integrating the previous differential equation:

$$I = I_\infty\ e^{-\sigma\ n_s\ H\ \sec(\chi)\ e^{-\frac{h}{H}}} \tag{1.8}$$

Since not all energy goes into ionization, the absorption will be modified by the ionization efficiency η, which denotes the number of (photo)electrons produced per photon absorbed. At this point, the ionization rate q can be defined as:

$$q = \eta\ n\ \sigma\ I = \eta\ n_s\ e^{-\frac{h}{H}}\ \sigma\ I_\infty\ e^{-\sigma\ n_s\ H\ \sec(\chi)\ e^{-\frac{h}{H}}} \tag{1.9}$$

As the altitude decreases, the intensity of radiation decreases and the mass density increases thus defining a maximum in the ionization rate that can be found solving $dq/dh = 0$. Therefore looking for the maximum of expression 1.9, the maximum of ionization rate (q_m) and the height for this maximum (h_m) are found:

$$q_m = \frac{I_\infty\ \eta}{H\ e} \cdot \frac{1}{\sec(\chi)} = q_{m,0}\ \frac{1}{\sec(\chi)} \tag{1.10}$$

$$h_m = H\ \ln(\sigma\ n_s\ H\ \sec(\chi)) = h_{m,0} + H\ \ln(\sec(\chi)) \tag{1.11}$$

where $q_{m,0}$ and $h_{m,0}$ are the peak of ionization rate and the height where it takes place when the beam penetrates the ionosphere with $\chi = 0$ (zenith direction). Arranging expression 1.9, the Chapman layer formula is obtained:

$$q(h,\chi) = q_{m,0}\ e^{1-\frac{h-h_{m,0}}{H}-\sec(\chi)\ e^{-\frac{h-h_{m,0}}{H}}} \tag{1.12}$$

Once the ion production rate is obtained, it is possible to link this quantity to the electron density. This relationship depends on the recombinations of the electrons with the molecular and atomic ions present in each layer. In the lower part of the ionosphere (i.e. in the E and F1 layers), the reactions are fast and the rate of the recombination of electrons depends quadratically on the electron density, that is, depends on the number of interactions per volume unit between the particles with different charge:

$$\frac{dN_e}{dt} = q - \alpha\ N_e^2 \tag{1.13}$$

where α is the mean dissociative recombination coefficient for the molecular ions. Assuming photochemical equilibrium (i.e. $dN_e/dt \simeq 0$), it is possible to obtain the expression for the electron density profile called α-Chapman layer:

$$N_e(h,\chi) = N_{e,0}\ e^{\frac{1}{2}\left(1-\frac{h-h_{m,0}}{H}-\sec(\chi)\ e^{-\frac{h-h_{m,0}}{H}}\right)} \tag{1.14}$$

where $N_{e,0} = (q_{m,0}/\alpha)^{1/2}$. Assuming a reference height ($h_{m,0}$) of 300km and a scale height (H) of 75km, the electron density profiles assuming $N_{e,0} = 10^{12} electron/m^3$ and different values for χ are in the upper panel of Figure 1.2.

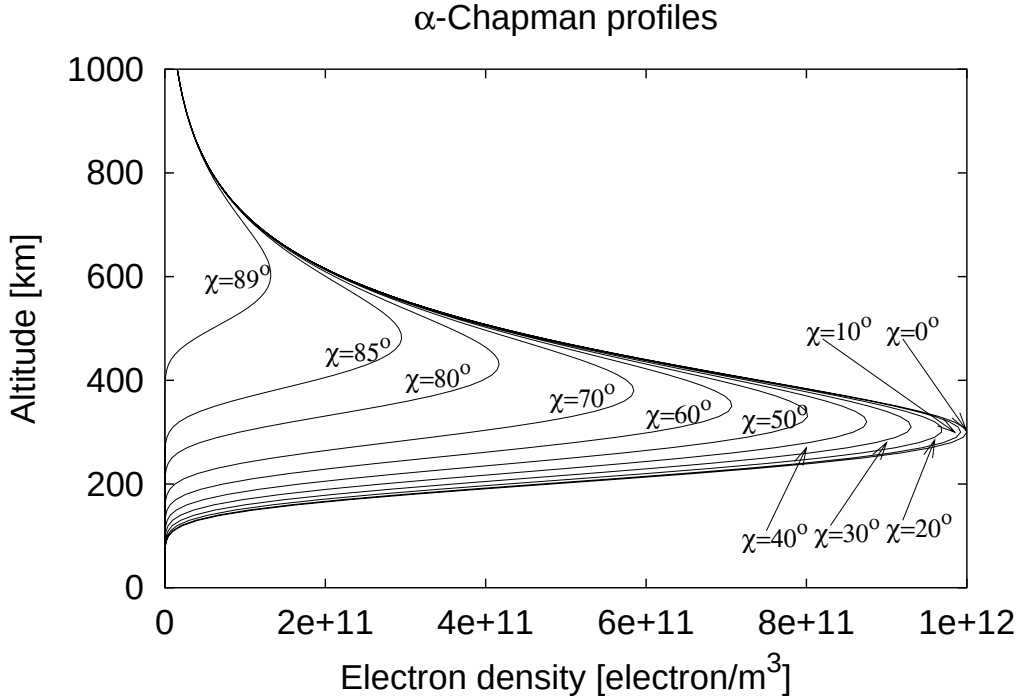


Figure 1.2: Set of different α-Chapman profiles corresponding to several values of χ (angle of the solar beam with respect to the zenith). The reference height is 300km, the scale height is 75km and the $N_{e,0}$ is 10^{12}

1.2 Variability of Ionosphere

Since the Sun is the main source of ionization of the ionosphere, any variation of the solar radiation or the relative geometry with respect to the Earth produce large dynamics in the electron content either in time and space.

- *Diurnal variation.* Due to the Earth rotation, the relative position of the Sun and Earth changes with time, following a daily cycle. This implies that the amount of ionization will be consequently dependent on this cycle.

- *Latitudinal variation.* The ionosphere shows certain behavior that are latitudinal dependent. In low latitudes, near the geomagnetic equator, the occurrence of the Appleton-Hartree anomalies (also known as Equatorial anomalies) take place. These anomalies are characterized as a depletion of electron density in the geomagnetic equator and large enhancements in the vicinity of $\pm 20^o$. This phenomena is caused by the "*fountain effect*": the presence of an eastward electric field in the geomagnetic equator causes an upward plasma drift, which rises until the pressure and gravity force are high enough and the plasma come back sideways through the magnetic field lines to higher latitudes (tropical ionosphere). In mid-latitudes the variations are lower, but the high latitudes, specially the regions comprised between 60^o and 70^o (i.e. auroral regions) are characterized by short-term variations more important than the lower latitudes. In the auroral regions, an interaction between the geomagnetic lines and the solar particle precipitation takes place. When the geomagnetic field is connected with the southward interplanetary magnetic field, geomagnetic storms are generated after solar ejection events (Coronal Mass Ejection or CME). In the polar caps the variation of the zenith angle of the Sun is much more smaller than other regions, therefore the variations in the electron densities are consequently smaller, although still detectable.

- *Solar cycle variation.* Several measurements of the sunspot number, which is an indicator of the solar activity, show that there is a main periodicity of 11 years (see Figure 1.3 for the sunspot number evolution since year 1995). This periodicity can also be detected studying the variability of the electron density.

Irregularities in the Ionosphere

Besides the expected variations of the electron density mentioned above, there are other unpredictable phenomena related to the Sun activity that depart from the usual state of the ionosphere. An important case are the geomagnetic storms, which are generated by a coupling of the solar wind and the Earth's magnetic field. This type of storms last from hours to several days and does not necessarily take place at a global scale. The typical geomagnetic

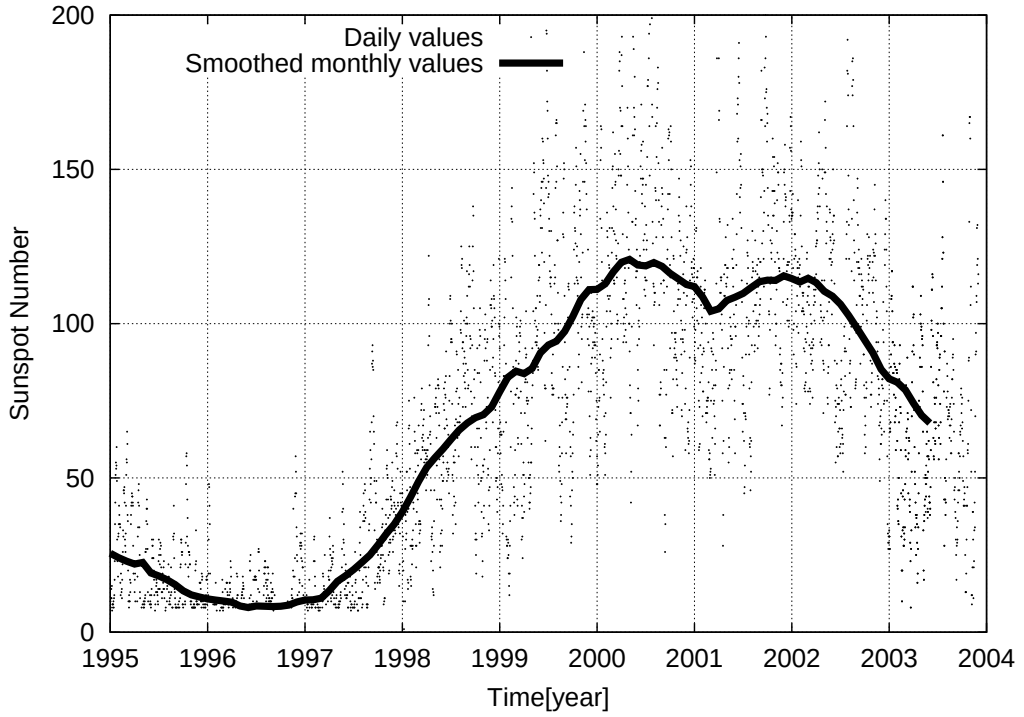


Figure 1.3: Sunspot Number from 1995 to 2003. The Sunspot number is an indicator of the solar activity. Note that the last maximum was bimodal with maximums placed at mid 2000 and beginning of 2002. The plot includes the daily measurements of the Sunspot number as well as the smoothed monthly median.

storm starts with an abrupt change followed by a recovery period that can last days. Several geomagnetic indices (for instance the K_p, A_p and Dst parameters) allow to monitor the virulence and evolution of a geomagnetic storm (see Figure 1.4 and [Davies, 1990]).

Another type of irregularities are the Traveling Ionospheric Disturbances (TID). These are wave-like irregularities related to perturbations of the neutral atmosphere, and can be classified according to its horizontal wavelengths (from 100km to 1000km), speeds (between $50ms^{-1}$ and $1000ms^{-1}$) and periods (from minutes to hours).

1.3 Ionosphere and electromagnetic signals

The presence of the ionosphere cause an effect on the electromagnetic signals that travel through it. Its effect will basically depend on the ionospheric free

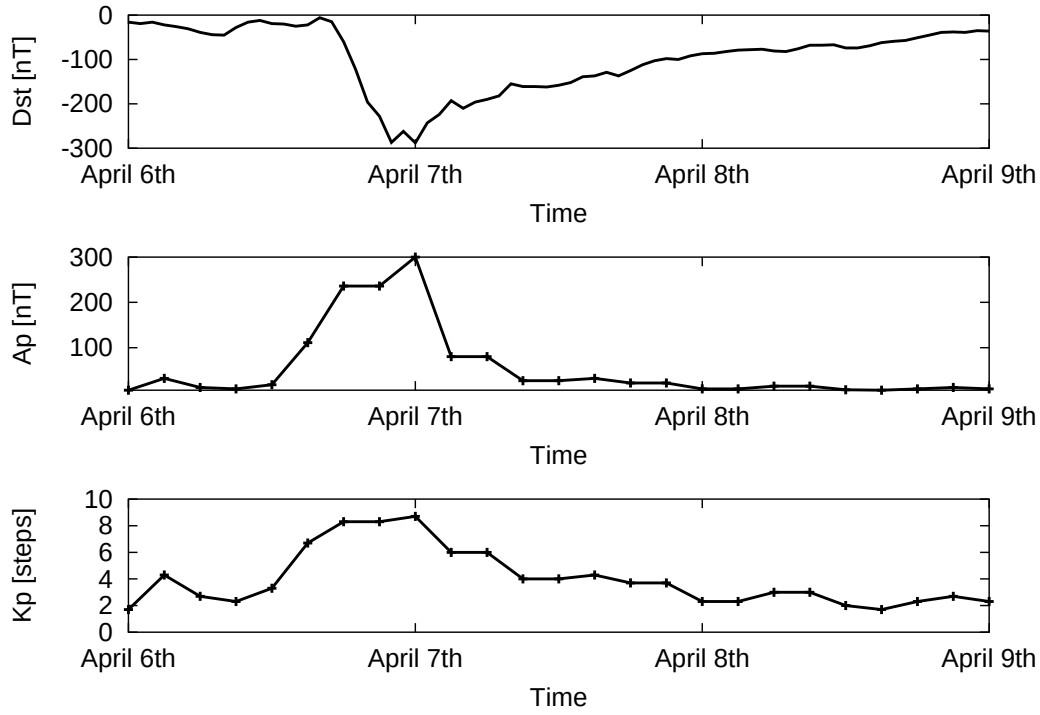


Figure 1.4: This plot show the Dst, A_p and K_p geomagnetic indices corresponding to a geomagnetic storm that took place in the beginning of April 2000. K_p values above 6 indicate that this is a severe storm. It can be clearly seen the abrupt beginning of the ionosphere (initial positive phase), the main negative phase and the recovery phase that last days, until normal ionospheric conditions are reached again.

electron density due to its most efficient interaction with electromagnetic signal (caused by the fact that the electron has a very high ratio charge to mass). Another important feature of the effect of the ionosphere on radio signals is that this effect depends as well on the frequency of the signal that traverses it (i.e. the ionosphere is a *dispersive medium*). This effect modifies the traveling speed of the signals with respect to the light speed, due to the presence of a refractive index different from 1. Depending whether one considers the group (where the information is carried) or the phase of the signal, this refractive index will be different. The following expression relates them both [Davies, 1990]:

$$n_{gr} = n_{ph} + f\frac{dn_{ph}}{df} \tag{1.15}$$

According to [Seeber, 1992] the phase refractive index can be approxi-

mated with a series expansion truncated after the quadratic term, therefore:

$$n_{ph} = 1 + \frac{c_2}{f^2} \tag{1.16}$$

Substituting expression 1.16 into relationship 1.15, the group refractive index is also obtained:

$$n_{gr} = 1 - \frac{c_2}{f^2} \tag{1.17}$$

The constant c_2 is a value that depends on the electron density and is defined in [Seeber, 1992] as:

$$c_2 = k \cdot N_e \qquad where\ k = -40.3\ m^3 \cdot s^{-2} \tag{1.18}$$

On the other hand the range between an emitter (Tx) and a receiver (Rx) is expressed as:

$$s = c \int_{Tx}^{Rx} \frac{dl}{v} = \int_{Tx}^{Rx} n ds \tag{1.19}$$

Assuming n=1, the straight line path is obtained instead:

$$s_0 = \int_{Tx}^{Rx} ds_0 \tag{1.20}$$

The ionospheric delay is defined then as the difference between the measured range s and the geometric range s_0:

$$I = s - s_0 = \int_{Tx}^{Rx} n ds - \int_{Tx}^{Rx} ds_0 \tag{1.21}$$

According to the definition of n, two different values of ionospheric refraction will be obtained, one for phase and the other for group. Simplifying the expression by considering the integration along the geometrical path (s_0) instead of the actual path (s), ionospheric contributions become:

$$I_{group} = \frac{40.3}{f^2} \int_{Tx}^{Rx} N_e\ ds_0 \qquad I_{phase} = -\frac{40.3}{f^2} \int_{Tx}^{Rx} N_e\ ds_0 \tag{1.22}$$

Therefore, the ionospheric effect is equal in value for phase and group but of opposite sign, that is an advance in phase and a delay in group.

At this point, the Slant Total Electron Content or STEC is defined as the integral of the electron density along the signal path, that is:

$$STEC = \int_{Tx}^{Rx} N_e\ ds_0 \tag{1.23}$$

The STEC is expressed in units of TEC Units (TECU), where 1TECU is defined as 10^{16} electron contained in a cylinder of cross section of 1 m^2 aligned with the ray path. Therefore the STEC quantities are in fact units of electrons per area unit (i.e. $electron/m^2$). If the STEC is in the vertical direction, one will have the Vertical TEC (VTEC, or simply TEC) instead of STEC.

1.4 Ionospheric data

1.4.1 Ionosonde

Since the early soundings of the ionosphere, ionosonde has been one of the most important instruments to obtain accurate description of the electron density distribution. This device relies in the fact that the signals under 20MHz are mainly reflected by the different layers of the ionosphere. The operation of this instrument consists in the emission and reception of signals from 0.1MHz to 30MHz in the vertical direction. Then, it measures for each frequency the travel time that took the signal to go upwards and downwards again due to the reflection of the ionosphere in the different layers.

The electron density present in the ionosphere defines the maximum frequency for which the signals are no longer reflected. This relationship is stated through the following expression:

$$f_c^2 = \frac{e^2}{4\pi^2 \epsilon_0 m_e} \cdot N_{max} \simeq 80.6 \cdot N_{max} \tag{1.24}$$

where e is the charge of the electron, m_e the mass of an electron and ϵ_0 the permittivity of free space. The electron density (N_{max}) is expressed in electrons per cubic meter and the critical frequency (f_c) is in units of Hertz. If the signal has a larger frequency than the critical frequency of the layer (f_c), it will not reflect the signal and therefore it will pass to a denser layer. When the critical frequency of the F2 layer is reached (it corresponds to the maximum of electron density or NmF2), signals with greater frequency will not be reflected anymore and will travel to the outer space (therefore the travel time measured by the ionosonde will be infinite).

With the travel time it is possible to obtain the apparent height of the ionized layer (*virtual height*), defined as:

$$h' = \frac{c\tau}{2} \tag{1.25}$$

where τ is the travel time and c is the speed of light. This height is not real but virtual since the pulse is mainly affected by the ionized layers,

instrumental delays and neutral atmosphere. The plots of virtual height against frequency are called *ionograms*.

With the ionogram and the relation 1.24 it is possible to obtain the electron densities of the different layers from the critical frequencies measured by the ionosondes, whose accuracy depends on various factors (see for instance [Piggot and Rawer, 1978] or [World Data Center-A, 2003]): (1) Inherent accuracy of the equipment, (2) accuracy of calibration method and (3) reading accuracy in the reduction of ionograms. The levels of accuracy of the ionosonde measurement of critical frequencies used in this work thesis are better than 5%.

One of the limitations of this instrument is that is not possible to obtain direct measurements of neither the valley between the maximum of the E-layer and the F-layer nor the topside ionosphere (i.e. measurements above the peak of the F2 layer).

Moreover, although the value of the virtual heights offers an orientation about the distribution in height of the ionospheric layers, it is desirable to have the profile of electron density referred to real height or, at least, the principal parameters such as the real height of the F2 layer peak (i.e. hmF2).

M(3000)F2 and hmF2

With the parameters provided by the ionosonde it is possible to apply straightforward expressions to obtain an estimation of the real height of the F2 layer peak using the M(3000)F2. This parameter is the maximum usable frequency for transmissions up to 3000km divided by the foF2 frequency (critical frequency of the F2 layer). In [Shimazaki, 1955] it was introduced a semi-empirical relation to compute an estimation of the hmF2, expressed in km, based on this parameter:

$$hmF2_{Shimazaki} = \frac{1490}{M(3000)F2} - 176 \qquad [km] \tag{1.26}$$

In fact, this value is an estimation of the hmF2 based on the assumption that the profile of electron density follows a parabola, but the departures to this assumption cause an important mismodelling. Further works ([Bradley and Dudeney, 1973],[Bilitza et al., 1979] and [Dudeney, 1983]) developed more accurate estimations based not only on the M(3000)F2 parameter, but on the ratio between the critical frequencies at the F2 and E layers. In particular, it is possible to obtain an estimation of the hmF2 with an accuracy of 4% or 5% ([Dudeney, 1983]) for good quality ionograms using the following expression:

$$hmF2_{Dudeney} = \frac{1490 \cdot F}{M(3000)F2 + \Delta M} - 176 \qquad [km] \tag{1.27}$$

where

$$F = M(3000)F2 \cdot \sqrt{\frac{0.0196 \cdot M(3000)F2^2 + 1}{1.2967 \cdot M(3000)F2^2 - 1}}$$

and

$$\Delta M = \frac{0.253}{\frac{foF2}{foE} - 1.215} - 0.012$$

being M(3000)F2, foF2 and foE the maximum usable frequency at 3000km divided by the foF2 (thus is an adimensional quantity) and the critical frequencies at the F2 and E layers (both expressed in MHz). Since the E layer is basically a diurnal layer and is difficult to measure during night using ionosonde, several authors apply conditions to this expression in order to obtain reliable values for the hmF2 parameter (see [McNamara et al., 1987] or [Rishbeth et al., 2000]). In particular, [Rishbeth et al., 2000] imposes two conditions:

1. $M(3000)F2 > 2.5$
2. $foF2/foE > 1.7$

This condition diminishes the number of hmF2 values, specially during night periods. It is generally accepted (see for instance [Zhang et al., 1999]) that the expected error in the hmF2 estimation using the Dudeney formula is between 20 and 30km.

The strength of these expressions rely in their straightforwardness. Therefore, since the number of ionosondes, and consequently the number of measured ionospheric parameters, has been increased during the last decades, it is possible to obtain, with a certain degree of sparsity, a high amount and world wide distributed values for hmF2. This allows performing studies about the morphology and temporal trends of the hmF2 parameter (see for instance [Rishbeth et al., 2000] or [Marin et al., 2001]).

True height profiles

The knowledge of the virtual heights is interesting for the applications that use the ionosphere for long-haul communications. Nevertheless, for the description of the ionosphere, it is more than interesting to have the whole

vertical profile of electron density referred to real heights rather than virtual heights, at least up to the hmF2. This problem requires an inversion of the raw ionogram given by the ionosonde by applying the *true height analysis*. Tools such as POLAN ([Titheridge, 1998]) or the "μ'fitting" technique ([Tsai et al., 2001]) are more complex approaches than the M(3000)F2 expressions, but allow to process ionogram data, expressed in function of virtual height, to obtain vertical profiles of electron density expressed in function of true height.

The true height analysis consist in the following process: The travel time corresponding to the lowest layers is almost unaffected by the ionization, therefore the deduced virtual height will be almost equal to the real height. As the frequency increases, the electromagnetic pulses will penetrate to denser layers, therefore the amount of ionization will be larger and the differences between the real and virtual height will increase as well. To invert from virtual to real height is necessary to reconstruct the ray in its travel through the ionosphere. To correctly perform this reconstruction is necessary to model (1) at which altitude the electron density becomes significant (*Starting position*) and (2) the valleys between layers that the ionosondes are unable to monitor (*Valley approximation*). These approximations can be performed, for instance, based on the geographical location of the ionosonde and epoch of measurement. An example of true height analysis performed with POLAN can be seen in Figure 1.5.

The POLAN method is considered to offer a high degree of accuracy in the true-height profiles, despite their high computational load. Depending on the quality of the ionograms it is generally accepted that the inaccuracies of those methods can be up to 20km-30km. Regarding the inter-comparison between the Dudeney formula and POLAN method, [McNamara et al., 1987] reported a discrepancy of 10km-20km between these two techniques in the estimation of the hmF2 parameter.

1.4.2 Ionospheric Models

Climatological models such as the International Reference Ionosphere (IRI) become handy to test ionospheric sounding algorithms because they provide with a controlled environment in which ionosphere is simulated. In this simulated environment the own difficulties of real data scenarios are not present. These problems are basically the sparsity of data and instrumental delays.

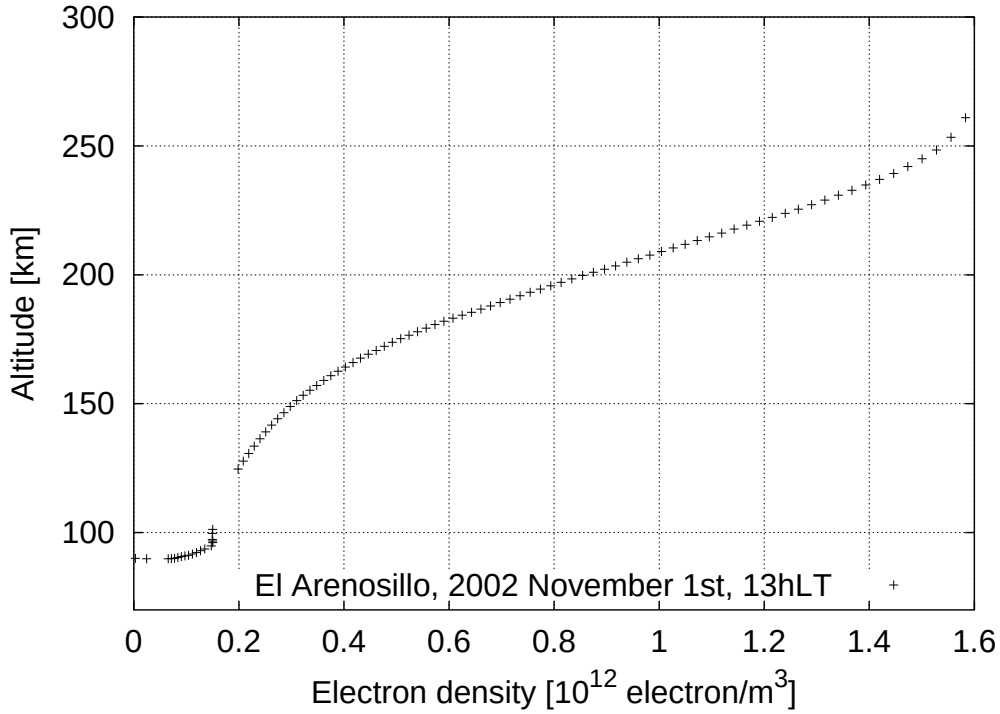


Figure 1.5: Example of electron density profile obtained with POLAN from ionogram data. The plot corresponds to the measurement from the ionosonde of El Arenosillo (Spain, 6.73^{o}W 37.1^{o}N) at 2002 November 1st. In order to correctly invert the raw ionosonde data, two assumptions regarding the starting height and the shape of the valley have to be made.

IRI

This climatological model comprises several height profiles for a group of plasma parameters: plasma density, plasma temperatures of electrons and ions and ion composition. The F2 plasma parameters are obtained from a series of coefficients given by the CCIR while the hmF2 are obtained using a method based on M(3000)F2 similar to the ones exposed before ([Bilitza et al., 1979]).

NeQuick

The NeQuick model (see [Hochegger et al., 2000]) is an empirical model of the ionosphere that generates profiles of electron densities based on parameters extracted from ionograms, that act as anchor points to the model. One of the particularities of this model is that it does not rely on Chapman profiles to construct the topside ionosphere, but uses semi-Epstein functions to

model it. In Figure 1.6 it is given an example of profile for the same epoch and geographical coordinates for both the IRI and NeQuick model.

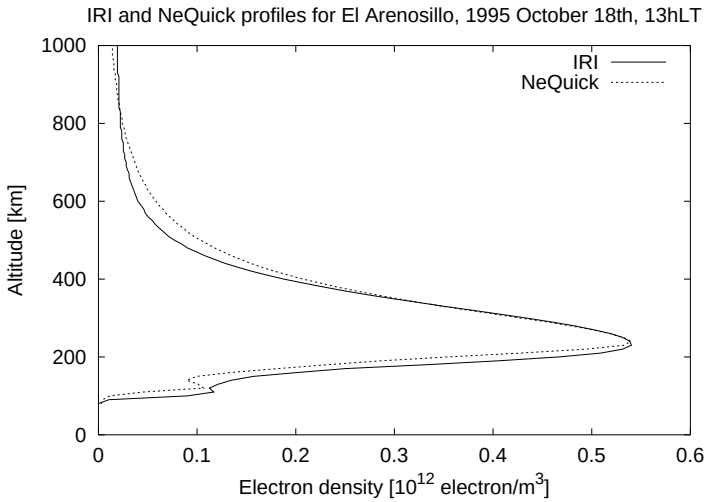


Figure 1.6: IRI and NeQuick profiles. This particular case corresponds to a profile of El Arenosillo ionosonde (Spain, 6.73^{o}W 37.1^{o}N) at the 13hLT of October 18th 1995.

1.4.3 Global ionospheric maps of VTEC

Since summer 1998, the International GPS Service (IGS) began to officially distribute, on a daily basis, Global Ionospheric Maps (GIM) of Vertical Total Electron Content (VTEC) computed using GPS data gathered in ground stations (see [Feltens and Schaer, 1998] and [Hernandez-Pajares et al., 1999] for the definition and for a description of global maps computed at UPC respectively). Those maps are distributed with a latency of one week. As the techniques improve and the computer capabilities increase, the IGS global VTEC maps are now computed as well with latencies better than 24 hours (rapid IGS ionospheric products, [Komjathy and Hernandez-Pajares, 2004]). The spatial resolution of these distributed global maps in the IONEX standard format is $5^o \times 2.5^o$ in longitude and latitude and 2 hours in time. An example of GIM map is shown in Figure 1.7. It can be seen the variation introduced by the Earth Rotation as well as the latitude dependence, specially the large enhancements of VTEC of the equatorial anomalies above and under the geomagnetic equator.

One of the applications of the GIM maps is to provide to navigation users with VTEC information. This type of data allows to correct the most part of

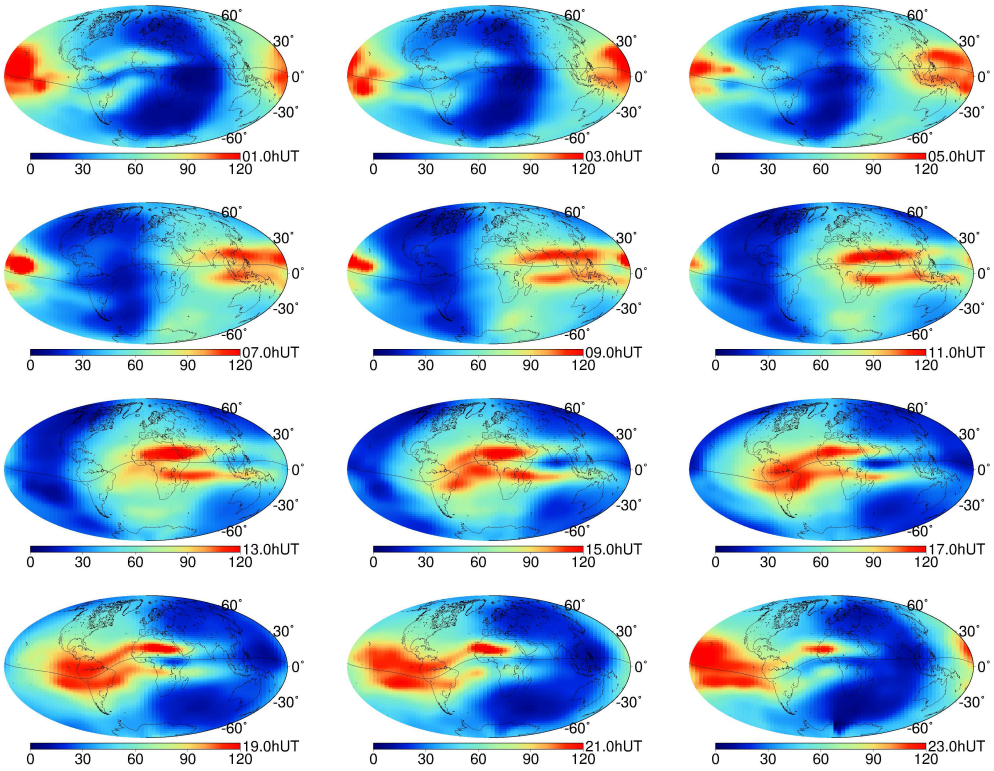


Figure 1.7: These series corresponds to the Global Ionospheric Maps in IONEX format computed and distributed by the Technical University of Catalonia (UPC, Spain) for the day May 16th, 2000.

the ionospheric effect on the GNSS signal and to obtain better precisions in positioning. The typical error of these maps (compared with external data, in this case of VTEC measurement provided by the TOPEX satellite) is of 15% approximately at northern mid-latitudes (see [Orus et al., 2002] and [Orus et al., 2003]).

VTEC interpolation

To obtain the VTEC in a given position from the global map in the IONEX grid, it is necessary to perform a spatial/temporal interpolation. It can be performed between the nearest two *rotated* VTEC maps as explained in [Schaer et al., 1998]. This rotation is related with respect to the Z-axis by an amount related to the time difference between the interpolation and map epochs. Essentially it consists on an interpolation in local time instead of longitude, which implies a minimization of the interpolation error due to the variability in longitude. This is justified by the significant stationarity of the

electron content with respect to a Sun fixed reference frame. Therefore, the interpolated VTEC follows the expression:

$$VTEC(\lambda,\phi,t) = \frac{T_{i+1}-t}{T_{i+1}-T_i} VTEC_i(\lambda'_i,\phi) + \frac{t-T_i}{T_{i+1}-T_i} VTEC_{i+1}(\lambda'_{i+1},\phi) \tag{1.28}$$

where $T_i \leq t < T_{i+1}$ and $\lambda'_i = \lambda + (t - T_i)$, which is the rotation to be applied to each map in order to obtain the *rotated* TEC maps. This expression comes from considering equal local times: $LT = T + \lambda = LT' = T_i + \lambda'_i$. $VTEC_i(\lambda',\phi)$ corresponds to the nearest *rotated* TEC map (i.e. the map at epoch T_i). The VTEC value for a given latitude and longitude within a map corresponds to a distance weighting of the nearest 4 grid points (see [Schaer et al., 1998]).

1.4.4 Internet servers offering ionospheric data

Extensive ionospheric data can be found in the Internet web or FTP servers:

- SPIDR web server (http://spidr.ngdc.noaa.gov, [O'Loughlin, 1997]) offers extensive data base of ionosonde parameters (critical frequencies, virtual heights among others).
- World Data Center A server (http://www.wdc.rl.ac.uk), offers ionosonde parameters as well as POLAN inverted profiles.
- CDDISA FTP server (ftp://cddisa.gsfc.nasa.gov/pub/gps/products /ionex) provides with an extensive data base of GIM maps in IONEX format.
- Server of the IGS Ionosphere Working Group, with a summary of the activities and related links, including those to download final and rapid IGS global VTEC maps at http://gage152.upc.es/∼ionex3/igs_iono /igs_iono.html.

Chapter 2

The Global Positioning System

2.1 Fundamentals

In the last years, Global Navigation Satellite Systems (GNSS), in particular the NAVigation Satellite Timing and Ranging Global Positioning System (NAVSTAR GPS), have led to an increasing number of applications such as precise navigation and timing, tropospheric and ionospheric sounding among others.

The GPS system consists in a set of satellites continuously transmitting a signal that is processed by the receiver. In order to obtain its coordinates, the receiver needs at least 4 satellites to solve for the 3 coordinates plus its own clock bias with respect to the GPS time scale.

Descriptions of GPS system can be found at [Wells, 1987], [Seeber, 1992], [Hoffmann-Wellenhof et al., 1994] and [Parkinson and Spilker Jr., 1996].

Satellites

The GPS constellation is constituted by a network of 24 satellites orbiting at 20200km approximately with respect to the Earth surface. They are evenly distributed within 6 orbital planes inclined 55^o with respect to the Earth's equator and equally spaced at 60^o. The orbital period is approximately 12 sidereal hours. This configuration guarantees a global 24 hours coverage with a visibility of at least 4 satellites, which are the ones needed to solve the position of a receiver.

Signal

The core of a GPS satellite is its high stable atomic clocks. These clocks oscillate at 10.23MHz and are used to generate the GPS signals, in particular the two L-Band carrier frequencies:

$$f1 = 154 \cdot 10.23\ MHz = 1575.42\ MHz$$
$$f2 = 120 \cdot 10.23\ MHz = 1227.60\ MHz$$

which correspond to approximate wavelengths of 19 cm (λ_1) and 24 cm (λ_2). The GPS signal is constructed in such a way so that a Coarse Ac-

quisition code (C/A) is modulated in the f_1 frequency, and a Precision (P) code and a navigation message are modulated in both frequencies. With this information, the receiver can compute its position. The codes are also known as Pseudo Random Noise (PRN) codes and there are more codes than potential number of satellites, so each code identifies the satellite.

The resulting transmitted signal corresponds to expression 2.1, whose schematic interpretation is shown in Figure 2.1.

$$\begin{aligned} S_{GPS}(t) = \; & A_c \cdot C(t) \cdot D(t) \cdot \sin(2\pi f_1 + \phi_c) + \\ & + A_p \cdot P(t) \cdot D(t) \cdot \cos(2\pi f_1 + \phi_{p1}) + \\ & + A_p \cdot P(t) \cdot D(t) \cdot \sin(2\pi f_2 + \phi_{p2}) \end{aligned} \tag{2.1}$$

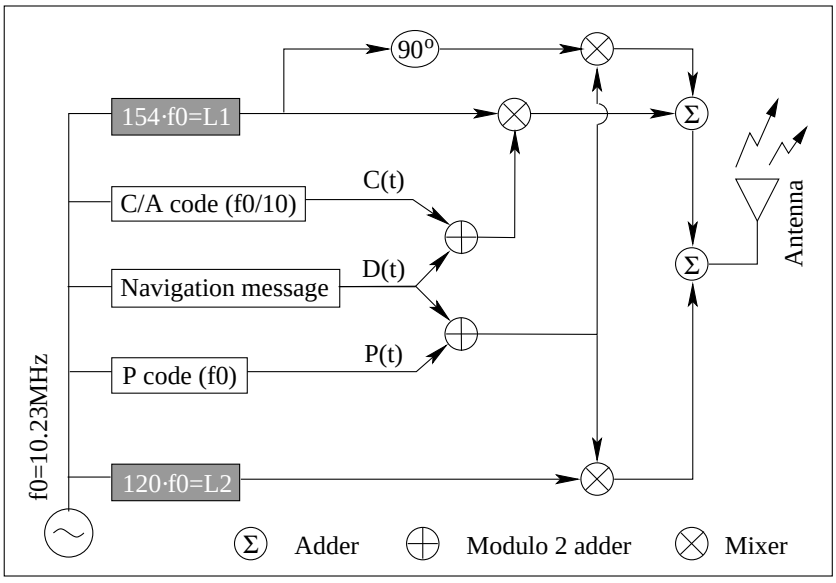


Figure 2.1: Schematic view of GPS signal structure

In order to limit the precision for which the civilian users can obtain its position, two modifications are applied to the GPS signal by the Unites States Department of Defense:

1. *Selective availability* (*SA*) consists in manipulating the satellite clocks. The ephemerides contained in the navigation message are also modified. The resulting effect is an increase of the positioning error from 10 meters to more than 100 meters approximately. From May 1st 2000, the SA was disconnected.

2. *Anti/Spoofing* (*AS*) consists in the encryption of the precision P-code (into the Y-code) so the unauthorized receivers are unable to use it,

forcing these users to rely in the worse C/A code in f_1, and in an indirect and noisier estimate of a code in f_2.

2.2 Observables

In order to obtain the GPS observables that will be used for processing, the GPS receivers correlate (compare) the incoming signal with an own generated copy. They basically measure the time or phase differences between both signals. If the differences are obtained from the PRN C/A or P codes (time differences) one will obtain the *Code Pseudorange*. Otherwise, by obtaining phase differences of the carrier frequency one will obtain *Phase pseudorange*. *Pseudoranges* are "ranges" because are an estimation of the geometric distance between the satellite and the receiver (traveling time multiplied by the light speed gives *apparent* distance), and are "pseudo" because this range is not the actual geometric range since it is affected by a set of errors and delays. The *Code Pseudoranges* (expressed in units of length) can be modeled for both frequencies as:

$$\begin{aligned} P1_i^j &= \rho_i^j + c(dt_i - dt^j) + I_i^j + T_i^j + rel_i^j + K1_i^j + M_{P1,i}^j + \varepsilon_{P1,i}^j \\ P2_i^j &= \rho_i^j + c(dt_i - dt^j) + I_i^j + T_i^j + rel_i^j + K2_i^j + M_{P2,i}^j + \varepsilon_{P2,i}^j \end{aligned} \tag{2.2}$$

Similarly, the *Phase Pseudoranges*, expressed in units of length as well, can be modeled as:

$$\begin{aligned} L1_i^j &= \rho_i^j + c(dt_i - dt^j) - I_i^j + T_i^j + rel_i^j + B1_i^j + w_{L1} + m_{L1,i}^j + \varepsilon_{L1,i}^j \\ L2_i^j &= \rho_i^j + c(dt_i - dt^j) - I_i^j + T_i^j + rel_i^j + B2_i^j + w_{L2} + m_{L2,i}^j + \varepsilon_{L2,i}^j \end{aligned} \tag{2.3}$$

The constant c is the speed of light in the vacuum and each pseudorange contribution are summarized in Table 2.1.

The term Bn_i^j in the phase pseudoranges is defined as $Bn_i^j = b_i + b^j + \lambda N_i^j$, thus including both instrumental delays and *integer ambiguity* term. Phase processing consists basically in measuring the accumulated difference between the incoming GPS signal and the receiver replica. Nevertheless it is not possible to measure the number of cycles between the GPS satellite and the receiver at the instance of first observation (this unknown quantity is the integer ambiguity). As a consequence, when the receiver loses visibility with the GPS satellite (for instance due to a building or vegetation) and re-locks afterwards, the phase observable shows discontinuities known as *cycle slips* (an example of cycle slips in GPS signal is given in Figure 2.2).

Geometric distance	ρ_i^j	$\simeq 20000km$
GPS sat $j \Leftrightarrow$ receiver i		
Receiver clock offset	dt_i	$< 300km$
Satellite clock offset	dt^j	$< 300km$
Ionospheric delay	I_i^j	1 to 50m
Tropospheric delay	T_i^j	1 to 20m
Relativistic effect	rel_i^j	$\simeq 10m$
Code Multipath effect	M	$0m - 3m$
Phase Multipath effect	m	$0cm - 5cm$
Code Noise	ε_P	$\simeq 3m(C/A)$
		$\simeq 30cm(P)$
Phase Noise	ε_L	$\simeq 3mm$
Wind-up effect (only phase)	w	$\simeq 200cm$
		per antenna rotation

Table 2.1: Main contributions of Pseudoranges. The multipath errors are difficult to quantificate since it highly depends on the environment. Additionally, the a priori broadcasted satellite position and clocks can cause an extra error of 2 meters approximately.

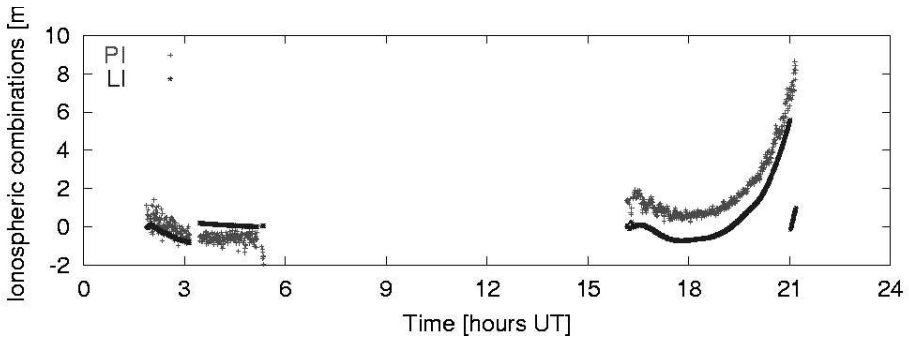


Figure 2.2: Cycle slips in GPS signal using Ionospheric combination of observables (defined later, it contains essentially the ionospheric delay and the instrumental delays and the additional phase ambiguity term in the case of LI). Note that the phase observable is reinitialized whenever a cycle slip or new signal lock takes place. It can be seen as well that, although being ambiguous, the phase observable is much more precise than the code observable.

The terms corresponding to clock bias (with respect to the GPS time scale) offer large errors unless they are dealt with accordingly. In the case of the receiver, since it usually uses a simple quartz clock used to generate the replica of the GPS signal, a large clock bias with respect to the GPS

time scale occurs. Therefore, at least 4 GPS satellites are needed to estimate this bias, simultaneously to the 3D position coordinates. In the case of the satellite, the satellite clock bias can be mostly corrected with the data included in the navigation message (D(t)). For more details of each pseudorange contribution can be found in [Hoffmann-Wellenhof et al., 1994] or [Parkinson and Spilker Jr., 1996].

Using these basic observables, it is possible to linearly combine them:

- *Free ionospheric combination (LC and PC)*: Taking into account that the ionospheric delay depends on the square frequency, it is possible to remove its effect by constructing this combination as follows:

$$PC = \frac{f_1^2 \cdot P_1 - f_2^2 \cdot P_2}{f_1^2 - f_2^2} \qquad LC = \frac{f_1^2 \cdot L_1 - f_2^2 \cdot L_2}{f_1^2 - f_2^2} \tag{2.4}$$

  This combination is basically used for navigation purposes in receivers that are able to process both GPS frequencies.

- *Narrow and Wide lane combinations (Pδ and Lδ respectively)*: The wide lane combination is used for cycle-slips detection since it provides with an effective long wavelength of λ_δ=86.2cm, which becomes very useful for this purpose. The Lδ and the corresponding combination for the code are constructed as follows:

$$P\delta = \frac{f_1 \cdot P_1 + f_2 \cdot P_2}{f_1 + f_2} \qquad L\delta = \frac{f_1 \cdot L_1 - f_2 \cdot L_2}{f_1 - f_2} \tag{2.5}$$

- *Ionospheric (or geometric free) combination (LI and PI)*: It cancels all terms that do not depend on frequency such as geometric range, troposphere and so on, leaving the ionospheric contribution, instrumental biases and wind-up among others. This combination is constructed as follows:

$$PI = P_2 - P_1 \qquad LI = L_1 - L_2 \tag{2.6}$$

  In order to respect the sign convention, the order of the observables are changed since the ionosphere causes a delay in the code and an advance in the phase in the same absolute amount (see Section 1.3).

In ionospheric sounding the information given by the ionospheric (or geometric free) observable becomes essential, therefore the next section offers a deeper insight to this combination and the effect of ionosphere to GPS signals.

2.2.1 Ionosphere and GPS: L_I and P_I

As seen in Section 1.3, the STEC plays a key role in determining the delay caused by the ionosphere to electromagnetic signals. From expressions 1.22 and 1.23, a relationship between these quantities can be established:

$$I_{group} = \frac{40.3}{f^2} \cdot STEC \qquad I_{phase} = -\frac{40.3}{f^2} \cdot STEC \tag{2.7}$$

being f the frequency, expressed in Hz, the STEC in $electrons/m^2$ and the ionospheric delay I expressed in units of meters of ionospheric delay.

According to each pseudorange observable (Definitions 2.2 and 2.3) and the definition of the ionospheric combination (Equation 2.6), P_I and L_I observables can be modeled as:

$$\begin{aligned} P_I &= \alpha_I \cdot STEC + K_I + M_{P_I} + \varepsilon_{P_I} \\ L_I &= \alpha_I \cdot STEC + k_I + \lambda_1 \cdot N_1 - \lambda_2 \cdot N_2 + m_{L_I} + w_{L_I} + \varepsilon_{L_I} \simeq \\ &\simeq \alpha_I \cdot STEC + b_I \end{aligned} \tag{2.8}$$

where

$$\alpha_I = \alpha_2 - \alpha_1 = \frac{40.3}{f_2^2} - \frac{40.3}{f_1^2} \simeq 1.05 \frac{m_{L_I}}{10^{17} electron/m^2} = 10.5 \frac{m_{LI}}{TECU} \tag{2.9}$$

and b_I contains the contribution of both the instrumental delays and phase ambiguities. As done in [Blewitt, 1989], the terms due to noise, multipath and higher-order ionospheric terms (whose error is less than centimeter) are not explicitly shown in the L_I expression, since being the remaining terms orders of magnitude larger.

Ionospheric delay will depend on the state of the ionosphere and its variability (see Section 1.2). The geometry of the ray (associated to the elevation for which the GPS satellite is viewed by the receiver) plays an additional important role in the ionospheric combination amount. The left panel of Figure 2.3 illustrates that the observations with high elevations correspond to rays that traverse less portion of ionosphere. Therefore the delay associated to ionosphere will be consequently lower, compared with rays with lower elevation. This can be seen with the right panel of the same figure, in which the L_I observable is lower for higher elevations.

This elevation dependency can be modeled in either 2D or 3D. In the case of 3D, it can be used, for instance, a voxel description as done in [Hernandez-Pajares et al., 1998]. In the case of a 2D approach, it is taken advantage of the fact that the electron density is mainly concentrated in a

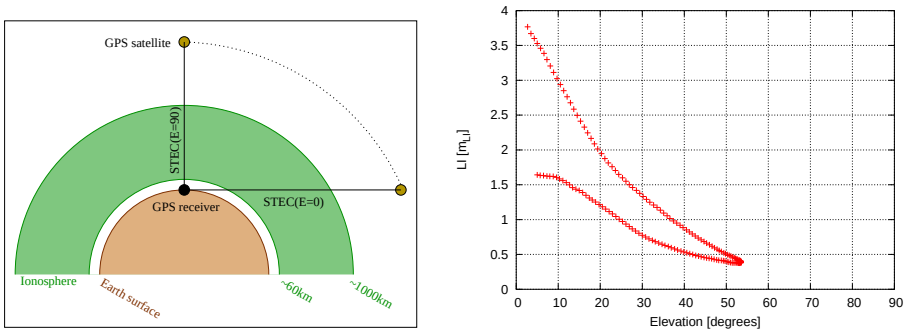
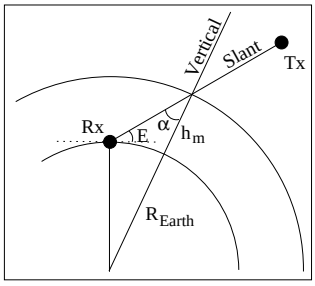


Figure 2.3: (Left) The contribution of the ionosphere will strongly depend on the elevation. As elevations decreases, the portion of ray traversing the ionosphere increases, thus the ionospheric delay increases as well. (Right) Example of dependency of ionospheric delay with elevation. This particular example corresponds to the IGS station of Herstmonceaux (0.4^{o}E 51^{o}N) that observed the GPS satellite PRN24 from 12hUT to 16.5hUT (same local time).

maximum between 200km and 500km (following the Chapman profile, see Section 1.1). Therefore it is assumed that the ionosphere is a thin layer that contains the electron density, and using a *mapping function* (m) it is possible to obtain the geometry factor that allows to project from the vertical to a slant direction (see Figure 2.4),



$$m = \frac{STEC}{VTEC} = \frac{1}{\cos(\alpha)} =$$

$$m = \frac{1}{\sqrt{1-sin^2(\alpha)}} = \frac{1}{\sqrt{1-\left(\frac{R_{Earth}}{R_{Earth}+h_m}cos(E)\right)^2}}$$

Figure 2.4: The STEC observation seen by a GPS receiver can be transformed to VTEC by means of a mapping function m that depends on the elevation (E, being $E = 0$ the horizon and $E = 90^o$ the zenith) and the height to which the thin ionospheric layer is placed (h_m) over the Earth surface (with radius R_{Earth}). Tx denotes the transmitter and Rx the receiver.

The value of h_m is usually comprised between 300km and 500km depending on the authors, nevertheless the choice of this height is only sensitive for

low elevations, as it can be seen in Figure 2.5.

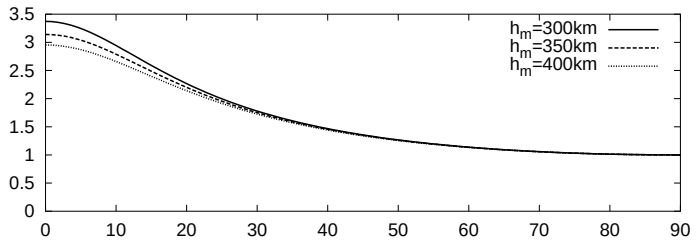


Figure 2.5: Geometric mapping function against elevation. It is depicted 3 different examples, considering h_m equal to 300km, 350km and 400km, it can be noted that the choice of h_m mainly affects to low elevations.

Note that for high elevations, the mapping function is close to 1. Recalling the expression of the L_I observable it can be seen that this makes it difficult to distinguish between the ionospheric contribution and the instrumental bias. This problem can be overcome using the phase instead of the code and estimating the bias in the process. This bias can be decorrelated as a consequence of the variation in the geometry (which causes a variation in the mapping function).

Chapter 3

Ionospheric tomography

3.1 Introduction

A possible definition of tomography is the one given by [Kak and Slaney, 1988]:

> *"Tomography refers to the cross-sectional imaging of an object from either transmission or reflection data collected by illuminating the object from many different directions."*

Tomography consists essentially in reconstructing an *image* (in this case the 3D distribution of free electron) from its *projections* (in this case the STEC), which are the integrals of the image in a given direction. The applications of this technique are extremely broad: from medical sciences to the mapping of underground resources. Ionospheric sounding using satellite data has also taken advantage of this technique to reconstruct the state of the ionosphere (by monitoring the distribution of its electron density). The first work that introduced the concept of ionospheric tomography is described in [Austen et al., 1988], although further improvements and studies were performed in following works by many authors (see for instance [Raymund et al., 1990], [Yeh and Raymund, 1991], [Fremouw et al., 1992], [Raymund et al., 1994], [Sutton and Na, 1996] and [Howe et al., 1998]).

3.1.1 Electron density and STEC

From the definition of tomography, and in an ionospheric context, the image would correspond to the electron density (N_e), and the *projections* or image integral in a given direction would correspond to the STEC, which is the integral of the electron density in the line-of-sight, as defined in expression 1.23.

The main idea inside the ionospheric tomography with satellite is that, once the transmitter and receiver coordinates are known, it is possible to derive the effect that the ionosphere cause to the electromagnetic signal in its journey from the satellite to the earth receiver. As this effect depends essentially on the STEC, as described in previous sections, the ionospheric

tomography consists basically in the retrieval of the electron density from the measurements of the satellite signal delay. A scheme of the geometry involved in the ionospheric tomography can be viewed in Figure 3.1. As the satellite journeys through its orbit, different receivers placed on the Earth surface gather observations affected by the ionosphere (which contain information about the STEC). Since the satellites work with sufficiently high frequencies, it is valid to assume that the ray bending due to ionosphere is negligible at the heights from the F layer and upwards and, in fact, the separation between the ray path of L1 and L2 is less than units of kilometers ([Hajj and Romans, 1998] and [Schreiner et al., 1999]). Therefore it is possible to assume that the ray follows a straight line between the transmitter (satellite) and the receiver. The observations of the different receivers are jointly processed in order to obtain a tomographic description of the ionosphere.

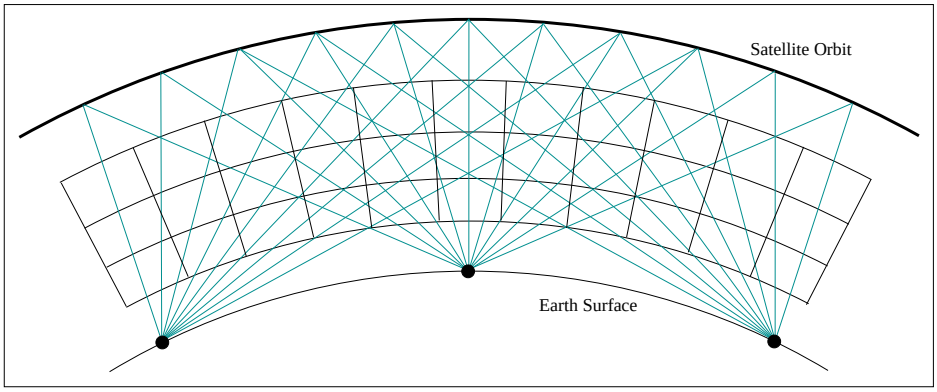


Figure 3.1: Scheme showing the geometry involved in a ionospheric tomography context. The STEC associated to the straight line-of-sights are processed in order to provide with the electron density in each box.

Before GPS was thought as a potential data source for ionospheric tomography, a previous satellite navigation system was used instead, the TRANSIT (or Navy Navigational Satellite System). The TRANSIT satellite system consisted in a set of satellites with nominal orbit of 1100km that were continuously transmitting at two frequencies 150MHz and 400MHz. Several works have been developed using these satellites (see for instance [Leitinger, 1994] or [Leitinger et al., 1997]) to test and implement the tomographic techniques that would be used afterwards with the GPS system.

By the end of 1996 the TRANSIT program was terminated and the GPS system substituted it. With GPS the possibilities of ionospheric tomography increased to a great extent. The global coverage of the GPS satellite

constellation converted this navigation system in a planetary scanner for 3D ionospheric sounding. In the case of GPS, the STEC information can be retrieved from the ionospheric combinations of phase and code (LI and PI respectively), as explained in Section 2.2.1.

3.2 Techniques

The impossibility to have the measurements of STEC evenly distributed in 180^o or 360^o, as in medical sciences, makes it impossible to apply methods based on Fourier transform that offer high resolutions ([Kak and Slaney, 1988]). In the case of ionospheric tomography the use of finite series expansion is considered instead. This technique, as explained in [Austen et al., 1988], consists in decomposing the STEC integral (equation 1.23) with a set of N_b basis functions that model the electron density, that is:

$$N_e(\lambda, \phi, h) \simeq \sum_{j=1}^{N_b} x_j \cdot b_j(\lambda, \phi, h) \tag{3.1}$$

where N_b is the number of basis functions considered, λ and ϕ are the horizontal coordinates (for which λ can be either local time or longitude) and x_j are the weighting coefficients to be applied at each basis function (b_j). Substituting this expression in equation 1.23, the STEC is expressed as the line integral of the electron density through the ray path, and can be approximated as:

$$\begin{aligned} STEC_i &\simeq \int_{T_x}^{R_x} \sum_{j=1}^{N_b} x_j \cdot b_j(\lambda, \phi, h)\ ds = \\ &= \sum_{j=1}^{N_b} x_j \int_{T_x}^{R_x} b_j(\lambda, \phi, h)\ ds \qquad\qquad i = 1, ..., N_p \end{aligned} \tag{3.2}$$

with N_p the number of STEC observations (or ray paths). The integral of the basis functions (B_{ij}) can be expressed as:

$$B_{ij} = \int_{T_x}^{R_x} b_j(\lambda, \phi, h)\ ds \tag{3.3}$$

the different STEC observation become:

$$STEC_i = \sum_{j=0}^{N_b} B_{ij} \cdot x_j + \varepsilon_i \qquad i = 1, ..., N_p \tag{3.4}$$

or its equivalent in vector notation:

$$\overrightarrow{STEC} = \mathbf{B} \cdot \vec{x} + \vec{\varepsilon} \tag{3.5}$$

where $\overrightarrow{STEC}$ is the vector with the STEC observations, B is the design matrix, $\vec{x}$ is the vector with the weighting coefficients and ε represents the noise, mainly due to unmodelled phenomena, discretisation error and measurement noise.

Therefore, a possible characterization of the tomographic techniques may be based in the selection of the *Basis functions*, which will determine the values of the B matrix, and the mathematical algorithm to solve the weighting coefficients vector $\vec{x}$.

3.2.1 Basis functions

The selection of the basis functions will determine how the model considers the ionosphere and will determine not only the values of the design matrix B, but the interpretation of the weighting coefficients. Due to the discrete nature of computer tomography, the final display of the ionospheric estimates (i.e reconstruction) will have to be expressed in *pixels* or *voxels*, depending whether a 2D or 3D tomographic scheme is performed. Pixels are rectangular elements that divide the ionosphere in a two dimensional grid. Analogously, voxels are the corresponding elements in a three dimensional environment.

The most simple approach is to consider pixels/voxels as the basis functions (afterwards reconstruction is immediate), nevertheless more complex approaches that consider alternative basis functions do exist as well, in this case the reconstruction is performed as a previous step to display the estimates. In fact, [Sutton and Na, 1996] classify the tomographic techniques in two broad classes related to the reconstruction technique: (1) techniques using pixels or (2) alternative functions other than pixels.

Pixeled ionosphere

In this case the ionosphere is divided in pixels or voxels so the basis functions would be defined as follows:

$$\delta(\lambda, \phi, h) = \begin{cases} 1 & if\ cell\ illuminated\ by\ ray \\ 0 & otherwise \end{cases} \tag{3.6}$$

A schematic view of this voxel based approach can be seen in Figure 3.2. Assuming straight line approximation, those cells "illuminated" by the ray will be associated to a basis functions equal to 1 and 0 otherwise.

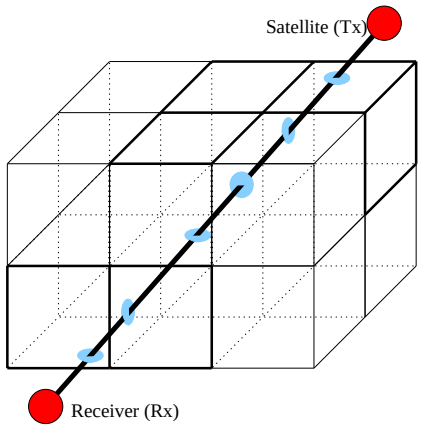


Figure 3.2: Scheme of ionosphere divided in voxels. The basis function (δ) equals 1 for the darker cubes (i.e. those "illuminated" by the ray) and 0 otherwise.

Therefore, assuming a three dimensional grid, expression 3.2 can be discretised and approximated by:

$$STEC \simeq \sum_{i=0}^{N_\lambda} \sum_{j=0}^{N_\phi} \sum_{k=0}^{N_h} x_{i,j,k} \cdot \Delta h_{i,j,k} \cdot \delta(i,j,k) \tag{3.7}$$

where $\Delta h_{i,j,k}$ are the lengths of the ray path at each "illuminated" cell and N_λ, N_ϕ and N_h are the number of cells in the longitude, latitude and height dimensions respectively. With this approach, the estimates (weighting coefficients, $\vec{x}$) are directly the electron densities in each cell. Since the estimated electron density is assumed to be constant within a cell, in a tomographic context it is important to focus in the compromise of the cell size: on the one hand it should be small enough in order to minimize the mismodelling due to the variations within the cell with respect the estimated value. On the other hand, it should be large enough so that the number of variables to solve (one for each cell in the grid) is kept reasonable and each of these cells are properly illuminated (by several rays with different orientation) to ensure its resolubility. This size will depend on various factors, being the most important the geomagnetic activity, solar cycle and latitude.

This modeling offers a direct and intuitively simple approach to reconstruct the ionosphere.

Empirical Orthogonal Functions (EOF)

An alternative way to describe the ionosphere is based on Empirical Orthogonal Functions, introduced in [Fremouw et al., 1992] and further developed in [Howe et al., 1998]. In this case the basis functions representing the electron density are expressed as:

$$N_e(\phi, h) = N_{e0}(h) + \delta N_e(\phi, h) \tag{3.8}$$

where $N_{e0}(h)$ is a background profile (with exclusive dependence on height h) and $\delta N_e(\phi, h)$ is the perturbation (with both latitude ϕ and height h dependency) with respect to $N_{e0}(h)$. While this background profile can be extracted from the rich data base of ionospheric observation that exist nowadays, the perturbation is modeled using harmonic functions analogous to Fourier series terms for the latitude and EOFs for the height variation ($\mathcal{H}(h)$). Thus, the perturbations are described with a set of coefficients (c_{jk} and s_{jk}) that multiply these EOFs. That is:

$$\delta N_e(\phi, h) = \sum_{j=0}^{N_h} \sum_{k=0}^{N_{EOF}} [c_{jk} \cos j\frac{2\pi}{L}\phi + s_{jk} \sin j\frac{2\pi}{L}\phi]\mathcal{H}_k(h) \tag{3.9}$$

where N_h and N_{EOF} are the number of harmonic terms and EOF functions considered to model the ionosphere respectively. The fundamental wavelength L is conservatively set to twice the latitudinal span of the region to be considered.

Therefore, the estimates (weighting coefficients, $\vec{x}$) of the tomographic method that uses this approach are not the electron density but the spectral description of the perturbation with respect to the background profile in the Fourier-EOF base of functions.

3.2.2 Mathematical algorithms

The other criteria to classify the tomographic approach is the type of mathematical algorithm used to obtain the estimates.

Back-substitution methods

Back-substitution methods are iterative procedures that consist in row action techniques to solve systems of the form:

$$\vec{y} = \mathbf{A} \cdot \vec{x} \tag{3.10}$$

The main feature of these techniques is the low memory requirements. The first algorithm applied in ionospheric tomography was the Algebraic Reconstruction Technique (ART) by [Austen et al., 1988]. The algorithm starts with a *first guess* ($\vec{x}^0$), and with an iterative scheme, the $\vec{x}^{k+1}$ vector of estimates is computed as follows:

$$\vec{x}^{k+1} = \vec{x}^k + \lambda_k \frac{y_i - \vec{A}_i \cdot \vec{x}^k}{\vec{A}_i \cdot \vec{A}_i^T} \vec{A}_i^T \tag{3.11}$$

where $\vec{A}_i$ is the i^{th} row of the design matrix $\mathbf{A}$ and the index i is set to k *module* N_p, being N_p the number of STEC observations (ray paths). The relaxation parameter is usually chosen to be $0 < \lambda_k < 2$ and equal to all k iterations. The process is carried on until the norm between the actual observation vector, $\vec{y}$, and the vector $\mathbf{A} \cdot \vec{x}^{k+1}$ is less than a threshold value, τ. A block diagram of this algorithm can be seen in Figure 3.3.

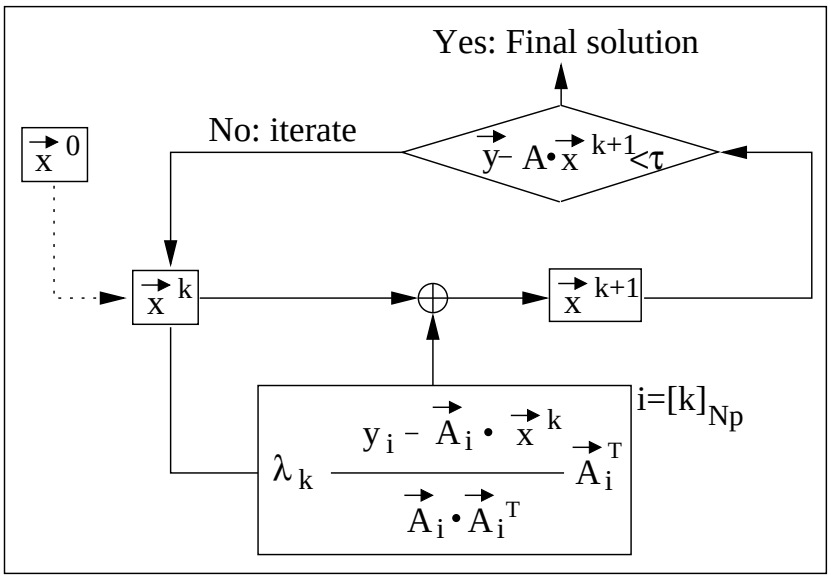


Figure 3.3: Block diagram of ART iterative method

An alternative implementation of the ART algorithm is the Multiplicative ART (or MART). In [Raymund et al., 1990], MART was introduced in a ionospheric context. Its main feature is that belongs to a maximum entropy algorithms, thus implying that MART minimizes the tendency towards any particular solution. The basis is exactly the same than the ART algorithm, the only difference is how the discrepancies between the estimates at k^{th} iteration ($\vec{x}^k$) are propagated to the next one. In the MART case this propagation is follows so that each element of $\vec{x}^{k+1}$:

$$x_j^{k+1} = x_j^k \left(\frac{y_i}{\vec{A}_i^T \cdot \vec{x}^k} \right)^{\lambda_k A_{ij}} \tag{3.12}$$

where the notation is the same as in Equation 3.11 and x_j^{k+1} is the j^{th} element of vector $\vec{x}^{k+1}$ and A_{ij} is the element of the matrix $\mathbf{A}$ placed in the i^{th} row and j^{th} column. In this case $i = (k \; module N_p) + 1$. The exponent is bounded so that $0 < \lambda_k A_{ij} \leq 1$.

Further variations of these algorithms may constitute others such as the Simultaneous Iterative Reconstruction Technique (SIRT) or an hybrid between the SIRT and the above mentioned ART (also known as Simultaneous ART or SART). Detailed description on these techniques may be found for instance in [Kak and Slaney, 1988].

Kalman filter

When an over-determined system of linear equations to estimate (Equation 3.10) is assumed stationary in time, it is possible to apply direct pseudo-inverse, the well known least mean square (LMS) approach. So the estimate of the state ($\hat{\vec{x}}$) is as follows:

$$\hat{\vec{x}} = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A^T} \cdot \vec{y} \tag{3.13}$$

or the corresponding version when each observation has its own weight, the weighted least mean square (WLMS):

$$\hat{\vec{x}} = (\mathbf{A}^T \mathbf{W} \mathbf{A})^{-1} \mathbf{A^T W} \cdot \vec{y} \tag{3.14}$$

Where W is a diagonal matrix containing the weights of each equation/measurement.

Nevertheless, when the system of linear equations shows time-dependence, the dynamics of this system must be taken into account. In this context, the Kalman filter is defined as a recursive filter that estimates the state ($\vec{x}$) of a linear time-varying system of equations driven by noise. The observations are built from a sequence of measurement ($\vec{y}$) in the presence of noise.

This section will describe the basic concepts involved in Kalman filtering since it is a key point in algorithms related to GPS data processing, in particular navigation algorithms and ionospheric tomography. A deeper insight of this algorithm and improvements on the classic approach are left to the reader, who can find further information at [Bierman, 1977] or [Welch and Bishop, 2003].

The basic equations that drive the Kalman filter are the following:

$$\begin{aligned} \vec{x}_{t+1} = &\ \Phi_t \cdot \vec{x}_t + \vec{w}_t \\ \vec{y}_t = &\ \mathbf{A}_t \cdot \vec{x}_t + \vec{\nu}_t \end{aligned} \tag{3.15}$$

where $\vec{x}_t$ is the state of the system at epoch t, Φ is the transition matrix (relates the state at epoch t to the state at the next epoch), $\vec{y}_t$ is the vector of predicted measurements at epoch t, $\mathbf{A}$ is the measurement or design matrix (relating the state to the measurement). The noise of the process and measurement ($\vec{w}$ and $\vec{\nu}$ respectively) are considered Gaussian white incorrelated noise, moreover they are independent to each other.

Basically, the Kalman filtering consists in a two-stage process, the *prediction* and *update*. Both are linked with the *Kalman gain*. The prediction stage accounts for the time variation of the state and is summarized, based on the classical formulation, with the following expressions:

$$\begin{aligned} \hat{\vec{x}}^-_{t+1} = &\ \Phi_t \hat{\vec{x}}_t \\ \mathbf{P}^-_{t+1} = &\ \Phi_t \mathbf{P}_t \Phi_t^T + \mathbf{Q}_t \end{aligned} \tag{3.16}$$

where $\mathbf{P}_t$ is the covariance matrix at time t and (Q) is the matrix representing the process noise. The matrix Φ_t is the transition matrix and indicates how the process evolves in time. The Kalman gain ($\mathbf{K}$) is defined as:

$$\mathbf{K}_{t+1} = \ \mathbf{P}^-_{t+1}\mathbf{A}^T_{t+1}(\mathbf{A}_{t+1}\mathbf{P}^-_{t+1}\mathbf{A}^T_{t+1} + \mathbf{R}_{t+1})^{-1} \tag{3.17}$$

Where the matrix $\mathbf{R}$ represents the measurement noise. The update stage corrects the estimation using the incoming measurements in such a way that:

$$\begin{aligned} \hat{\vec{x}}_{t+1} = &\ \hat{\vec{x}}^-_{t+1} + \mathbf{K}_{t+1}(\vec{y} - \mathbf{A}_{t+1}\hat{\vec{x}}^-_{t+1}) \\ \mathbf{P}_{t+1} = &\ (\mathbf{I} - \mathbf{K}_{t+1}\mathbf{A}_{t+1})\mathbf{P}^-_{t+1} \end{aligned} \tag{3.18}$$

An schematic view of the Kalman filter using the classical formulation can be seen in Figure 3.4. Note than expression 3.18 resembles to the one related to the ART algorithm (expression 3.11). In fact, the expression of the ART algorithm, taking $\lambda_k = 1$, is analogous to the "Scalar Filter" implementation of the Kalman filter (see [Bierman, 1977]).

Kalman filter has proven to be a very useful estimator for time-varying systems, in particular for navigation and ionospheric applications. Nevertheless it can not converge under certain conditions. The main reasons for divergence of the filter are due to: (1) an incorrect a priori and/or unmodelled parameters, (2) nonlinearities and (3) computer roundoff. In particular, extensive research has been made in order to avoid the problems derived from the last issue applying more robust techniques (alternative Kalman mechanizations tackling this topic can be found at [Bierman, 1977]). An additional

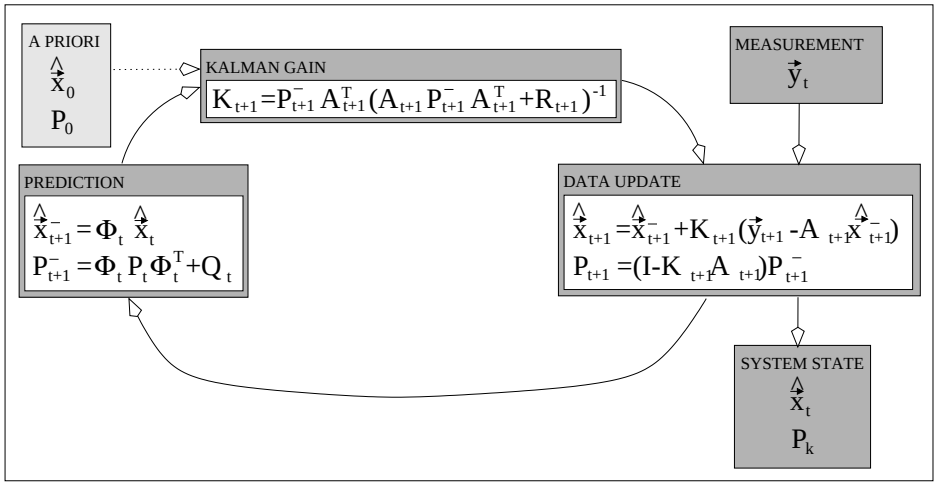


Figure 3.4: Block diagram of Kalman filter. This approach corresponds to the classical formulation.

point to take into account are the a prioris and weights chosen. The incorrect choice of both can cause dramatic effects on the estimation results.

3.3 Limitations

Although the great potentiality of tomography, the inherent geometry involved in the ionospheric case (i.e. it is not possible to have observations in all angles), the location of the receivers used and the discrete nature of data impose certain limitations pointed out in [Austen et al., 1988] and further developed and formulated in [Yeh and Raymund, 1991]. The different implementations described above try to diminish these limitations.

These limitations applied to ionospheric tomography with GPS can be summarized in three issues: (1) lack of vertical resolution, (2) sparsity of data and (3) computer load. The following subsections offer an overview to these limitations and solutions proposed in different works.

3.3.1 Vertical resolution

Satellite ionospheric tomography using ground receivers presents the problem of the vertical resolution. Although good estimates of VTEC can be achieved, it is very difficult to distribute properly the electron content in the different layers of the vertical. This is due to the fact that satellite STEC observations gathered by ground receivers do contain vertical info but to a very limited extent, being not enough to decompose the integral to obtain the different

contributions of the electron density. In this context, several solutions are devised.

A priori

A possible solution to the vertical resolution issue is the use of some kind of a prioris that inform about the vertical distribution of the ionospheric electron content. In the case of implementations with ART or MART, a first guess is needed to start the iteration. Nevertheless, this first guess can greatly determine the final solution if the number of observations is not enough.

An alternative implementation is using the EOFs approach. In this case the vertical resolution can be as large as required since it consists in a set of background profiles and perturbations built from the large information of ionospheric structures that exists nowadays. Nevertheless the drawback of these systems are similar to that of the ART and MART implementations. If not enough observations are obtained, these models can determine to a great extent the final solution. Moreover, the profiles and perturbations are built from average profiles, this may cause that local features, events or ionospheric irregularities may not be detected.

Summarizing, the problem of the a prioris relies in the relative importance with respect to the observations ([Sutton and Na, 1996]). If there are a large amount of a prioris or its relative weights are too large, the observations will be discarded by the algorithm in order to fulfill this previous information.

Combining complementary data

An alternative way to built a data driven model that improves the vertical resolution of the tomographic approach based on ground GPS receivers is the combination of different data sources. An illustrative scheme based on a naive example with voxels can be found in Figure 3.5, it illustrates that the knowledge of the sum in the vertical (equivalent to STEC measurements of ground receivers) is not enough to determine the content of each cell (equivalent to electron density) unless alternative information is given. The solution may rely in the use of an additional data in order to decorrelate the different contributions in the vertical and decompose the STEC integral in the electron densities. That is, to include data sources that provide with horizontal information.

With the advent of recent LEO missions that carry a GPS receiver on-board, a complementary data type to ground data is obtained. Its STEC observations are associated to line-of-sights that cross the ionosphere horizontally (see Figure 3.6), thus providing with an extra data type that allows

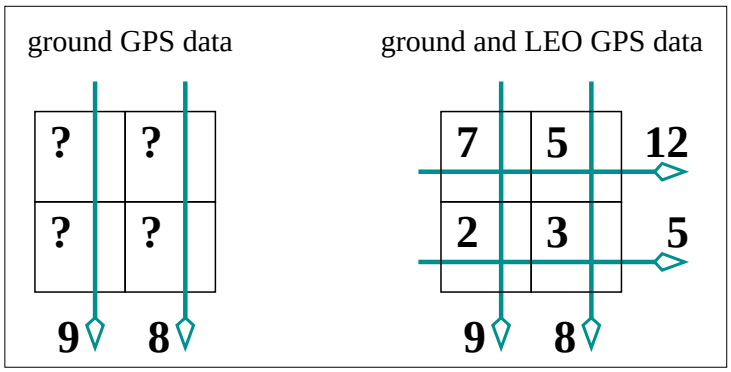


Figure 3.5: This simple example shows how the use of complementary data can help to obtain vertical resolution. If only vertical integrals are used (analogous to the information gathered by ground GPS receiver, left panel), it is not possible to decompose uniquely the integral in its parts. When complementary information is added to the system, it is possible to solve the system (right panel, which symbolizes the use of both GPS and LEO data).

to apply the idea of Figure 3.5.

The first works to develop simulated studies using combination of ground and LEO GPS data are described in [Hajj et al., 1994] and [Howe et al., 1998], showing promising results regarding the feasibility of this technique. With [Hernandez-Pajares et al., 1998] actual data gathered from both ground and LEO GPS receivers were jointly processed. The resulting estimates were compared with ionospheric parameters gathered from ionosondes, showing good agreement using real data.

In this context, several efforts are directed in what is known as *data assimilation*. Data assimilation consists in considering as much ionospheric information as possible (for instance ground and LEO GPS STEC retrievals, ionosonde data and Incoherent Scatter Radar (ISR) soundings) and assimilate them into a physical model of the ionosphere.

3.3.2 Sparsity of data

Although the number of GPS receivers has been increased in recent years, extensive networks of GPS stations such as the IGS one are unevenly distributed in the Earth globe. For instance, there are not receivers in oceans and seas save the ones placed in islands. In any case, the density of receivers of these areas and other continental regions such as Africa or Asia are much more less when compared with Europe, United States or Australia. These unevenly distributed GPS receivers cause that those regions with high den-

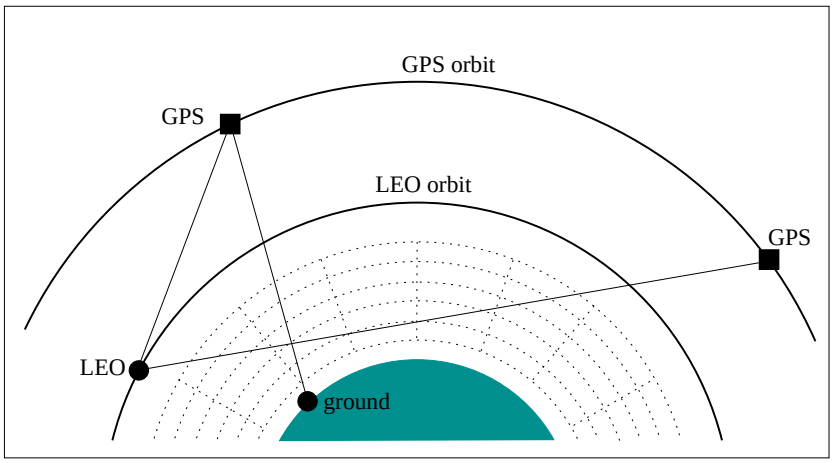


Figure 3.6: The lines-of-sight associated to LEO observations are mainly horizontal, in contrast with those gathered by ground GPS data, which are mainly vertical.

sity of GPS receiver can image properly the ionosphere, on the other hand, those regions with lack of receivers generate areas where the ionosphere is difficult to be monitored.

In order to solve for this problem, interpolation techniques become essential. The interpolation scheme may be aided with the use of a model as in the case of Global VTEC maps computed at the Technical University of Catalonia (UPC), where an interpolation aided with the IRI model was used to deal with this issue ([Hernandez-Pajares et al., 2002]).

3.3.3 Computational load

The number of variables involved in the resolution of the tomographic algorithm is a critical problem, regardless the tomographic or data assimilation technique used. In the case of Global VTEC maps (see Section 1.4.3), the ionosphere is divided in a horizontal grid of $5^o \times 2.5^o$ in longitude and latitude respectively. This corresponds to a total number of $72 \times 72 = 5184$ variables. If vertical resolution is required, this number has to be multiplied by the desired number of vertical layers. Assuming a desired vertical resolution of 50km, 20 vertical layers are required to image the ionosphere up to 1000km. Although this resolution is poor, the number of variables (N) increases to 103680, which is associated to huge design matrixes in the implementation. For the given example, the memory space needed for each matrix, which needs N^2 variables, is more than 85Gbyte.

The mathematical algorithms used to solve the estimates can be modified

in order to reduce the process time. For instance in [Hajj et al., 2002] it is introduced a method that performs ionospheric tomography at a global scale with a *band-limited* implementation of the Kalman filter. This approach follows the usual steps of the Kalman filter with a slight difference in the treatment of the covariance matrices. For a given voxel only the neighboring voxels within a *correlation scale* will be considered as voxels with non-zero covariance. With this approach the number of operations involved in the updating the covariance matrix are reduced, thus decreasing the process time. Nevertheless, the number of variables to solve is still an issue to keep the computational load relatively small. The following are possible alternatives to this problem:

1. An immediate alternative is to focus on a specific area to image (for instance a country or continent) instead of considering a global coverage. This allows to reduce the number of variables in the longitude/latitude plane so a better resolution in the vertical can be obtained.

2. If the horizontal (longitude/latitude plane) resolution is reduced, the number of variables reduces as well, nevertheless, care must be taken not to take large cells since it may cause that the variations within a cell are masked by the estimated value of electron density of the voxel (which is assumed to be constant in the whole cell).

3. Developing of alternative techniques that allow the reduction of the number of voxels in the longitude/latitude dimension. In this context, the Shape Functions (or Normalized Electron density functions), which are defined in Chapter 6, show smaller geographical variability than the Electron density functions. Therefore, a scheme based on these functions may be implemented to overcome this limitation.

These solutions may be combined in order to reduce in a substantial way the computer requirements to solve for the tomographic estimates.