Ionospheric Modeling for Precise GNSS Applications

Y. Memarzadeh

Ionospheric Modeling for Precise GNSS Applications

PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van de Rector Magnificus prof. dr. ir. J.T. Fokkema,
voorzitter van het College voor Promoties,
in het openbaar te verdedigen op dinsdag 8 december 2009 om 10.00 uur

door

Yahya Memarzadeh

Master of Science in Geodesy Engineering
K. N. Toosi University of Technology, Tehran, Iran
geboren te Tabriz, Iran

Dit proefschrift is goedgekeurd door de promotor: Prof. dr. ir. P.J.G. Teunissen

Samenstelling promotiecommissie:

Rector Magnificus,	voorzitter
Prof. dr. ir. P.J.G. Teunissen,	Technische Universiteit Delft, promotor
Dr. ir. H. van der Marel,	Technische Universiteit Delft, copromotor
Prof. dr. D.G. Simons,	Technische Universiteit Delft
Prof. dr. ir. R.F. Hanssen,	Technische Universiteit Delft
Prof. dr. M. Hernandez-Pajares,	Universitat Politècnica de Catalunya
Prof. dr. R. Warnant,	Université de Liège
Dr. B. Arbesser-Rastburg,	European Space Agency, adviseur

Memarzadeh, Yahya

Ionospheric modeling for precise GNSS applications

Delft institute of Earth Observation and Space systems (DEOS),
Delft University of Technology

Keywords: Ionospheric physics, Ionosphere modeling, Total Electron Content (TEC), Traveling Ionospheric Disturbances (TID), NeQuick model, Global Ionospheric Maps (GIM), Global Navigation Satellite Systems (GNSS), precise GNSS positioning

Citation: Memarzadeh, Y. (2009). Ionospheric modeling for precise GNSS applications. PhD thesis, Delft University of Technology,

ISBN-10 90-902468-5-1
ISBN-13 978-90-902468-5-7

This dissertation was published under the same title in the series: Publications on Geodesy, 71, Netherlands Geodetic Commission, Delft, the Netherlands,
ISBN-13 978-90-6132-314-3

Nederlandse Commissie voor Geodesie (NCG), Netherlands Geodetic Commission
P.O. Box 5030, 2600 GA Delft, the Netherlands
Tel.: +31 (0)15 278 28 19
E-mail: info@ncg.knaw.nl
Website: www.ncg.knaw.nl

Cover designed by the author.
Typeset by the author with the LaTeXDocumentation System.
Printed by Optima Grafische Communicatie, Rotterdam, the Netherlands.

This dissertation is dedicated to:
my parents, my wife and our son Ehsan.

Abstract

Ionospheric Modeling for Precise GNSS Applications

The main objective of this thesis is to develop a procedure for modeling and predicting ionospheric Total Electron Content (TEC) for high precision differential GNSS applications. As the ionosphere is a highly dynamic medium, we believe that to have a reliable procedure it is necessary to transfer the high temporal resolution GNSS network data into the spatial domain. This objective led to the development of a recursive physics-based model for the regular TEC variations and an algorithm for real-time modeling of the medium-scale Traveling Ionospheric Disturbances (MS-TID). The research described in this thesis can roughly be divided into three parts.

The main application of these developments can be found in Network RTK. Network-RTK is a technique based on a network of reference receivers to provide cm-level positioning accuracy in real time for users in the field. To get centimeter accuracy after a short (minutes) initialization period the ionospheric delay for the user's receiver needs to be predicted very precisely between the ionospheric pierce points of the reference receivers at the double difference level. Having the cm-level accuracy in the ionospheric interpolation is crucial for the carrier phase ambiguity resolution by the user. To achieve high precision in the ionospheric interpolation, regular and irregular variability of TEC in time and space should be taken into account. The regular TEC variation, which can reach several hundreds TEC units, is mainly a function of solar zenith angle. The irregular (or non-repeatable) variations are mainly wavelike effects associated with Traveling Ionospheric Disturbances (TID).

Although TID effects on the TEC are of the order of 0.1 TEC unit, MS-TIDs, with a typical wavelength less than a few hundred kilometers, is one of the main obstacles for accurate spatial interpolation of ionospheric induced delays in a medium-scale reference GPS network. Since most of interpolation methods either use spatial linear (or quadratic) interpolation or fit a lower-order surface, the methods are not capable to model the phase-offset, caused by MS-TIDs, at distinct ionospheric pierce points. There are two major complications. Firstly, interpolation must be done at the double-difference level, which involves taking single differences between ionospheric delays for the same satellite between two different receivers, followed by differencing single differences for different satellites. This means that two different patches of the ionosphere are involved, each related to a different satellite, and each possibly associated with different TIDs. Secondly, for operational network RTK, a real-time strategy for TID detection and modeling is needed.

In the first part the performance of several empirical ionosphere models for the regular TEC variation, such as Klobuchar, NeQuick, and the IGS Global Ionosphere Maps (GIM) are studied in the mid-latitude region using GPS data. Our results show that the GIM was able to correct the absolute slant ionospheric delay to better than 80% under different geomagnetic conditions of the ionosphere. The NeQuick model, which performed better than the Klobuchar model, could correct about 60% of the slant ionospheric delay. NeQuick is a real-time ionospheric correction model for the future European Galileo navigation system. A key input parameter for NeQuick is the effective ionization parameter (Az), which will be provided as a second order polynomial in the Galileo broadcast message to single-frequency users. The coefficients of the polynomial will be estimated daily from at least 20 permanent Galileo monitoring stations. As Galileo is under development, we propose an alternative approach for estimating Az using Global Ionospheric Maps (GIM). The main advantages of the alternative approach over the standard approach are: (1) the alternative approach is more reliable, because, each IGS GIM is based on data of up to 300 GNSS stations world-wide and each IGS GIM is the combination of results of up to four analysis centers, (2) the coefficients are more representative for all regions on the world because they are computed from a world-wide grid instead of about 20 distinct locations, (3) with the alternative procedure it is possible to provide Az in a different representation, for instance using a higher order polynomial, grid, or other function types, and (4) the computational effort is much smaller assuming the IGS GIMs have already been computed.

In the second part a normal ionosphere is defined using Chapman's ion production theory to approximate the regular variability of the Earth's ionosphere. The normal ionosphere consists of lower and upper region. The lower region is formed in a photochemical equilibrium resulting in a Chapman layer. The upper region is formed in a diffusive equilibrium, whilst ignoring the geomagnetic field, resulting in a new Chapman like ionospheric layer. Integration of the continuity equation of the normal ionosphere over height leads to a Boundary Value Problem (BVP) for the temporal evolution of VTEC. Solution of the BVP results in a novel recursive model for the regular TEC variation as a function of solar zenith angle. The main motivation for developing this model is that the empirical models of the first part were either ill-suited or too complicated to model and predict the regular variation of TEC for high precision differential GNSS applications. The performance of the new model is tested at local and global scales using GIM. In general, despite the geomagnetic field was ignored, the cases analyzed show that the model gives a good overall representation of the regular variation of VTEC in the mid-latitude region under a geomagnetically quiet ionosphere. This is an important result that shows the potential of the model for a number of applications. Since the model has a recursive form it is ideally suited to use as time update equation in a dynamic data processing or Kalman filter. Another application is to use it for removing the geometry-dependent trend from time series of GPS-provided ionospheric delays to provide a pure TID observation, which is carried out in the third part of this thesis.

In the third part, a new algorithm for the real-time detection and modeling of MS-TID effects is developed. In order to eliminate effects from large-scale TIDs, the algorithm uses between-receiver single-difference (SD) ionospheric delays in a medium scale GPS network. Although single-differencing also eliminates to some extend the geometry-dependent trend, the remaining part cannot be neglected. In this thesis, we fit the SD data to the recursive

model which was developed in the second part of the thesis. Any wavelike fluctuations in the data with respect to the model are assumed to be from MS-TID effects. The detrended SD data are the main input of the algorithm. The algorithm consists of six steps: initialization, detection, scraping, cross-correlation, parameter estimation, and ending. A MS-TID is assumed to be a planar longitudinal traveling wave with spatially independent amplitude that propagates in an ionospheric patch. All characteristic parameters of the MS-TID wave (e.g. period, phase velocity, propagation direction, and amplitude) are considered to be time dependent, while the Doppler-shift caused by the satellite motion is taken into account in the estimation step. The performance of the algorithm is tested with GPS data from a network. Although real TIDs are not perfect waves, the algorithm was able to model (in time and in space) the MS-TID to a large extend. The performance was found to be comparable with the Kriging interpolation method. This is an important first result, in part because these two methods are based on different principles, but also because there is still room for improvement in our algorithm. With our physics based model it is possible to avoid the planar wave approximation and take the phase-offset of the wave into account, something which is not possible with Kriging.

Samenvatting

Ionosfeer Modellering voor Nauwkeurige GNSS Toepassingen

De doelstelling van deze thesis is een procedure te ontwikkelen om de ionosferische "Total Electron Content" (TEC) te modelleren en te voorspellen voor zeer precisie differentile GNSS toepassingen. Aangezien de ionosfeer een zeer dynamisch medium is, geloven wij dat om een betrouwbare procedure te krijgen het noodzakelijk is om de hoge tijd resolutie van GNSS netwerkgegevens in het ruimtedomein over te brengen. Deze doelstelling leidde tot de ontwikkeling van een recursief op fysica-gebaseerd model voor de regelmatige TEC variaties en een algoritme voor real time modellering van de medium-scale "Travelling Ionospheric Disturbances" (MS-TID). Het onderzoek dat in deze thesis wordt beschreven kan ruwweg in drie delen worden verdeeld.

De belangrijkste toepassing van dit onderzoek is te vinden in GNSS RTK-Netwerken. Een RTK-Netwerk is een netwerk van GNSS referentieontvangers die gebruikers in staat stelt in real-time centimeternauwkeurigheid in hun plaatsbepaling te bereiken. Om na een korte initializatie periode (minuten) centimeternauwkeurigheid te bereiken moet de ionosferische vertraging tussen de referentiestations en de gebruiker zeer nauwkeurig voorspeld worden op het dubbel-verschilniveau. Dit is essentieel voor de oplossing van geheeltallige fase meerduidigheden door de ontvanger. Om de hoge precisie in de ionosferische interpolatie te bereiken moeten regelmatige en onregelmatige variaties van TEC in tijd en ruimte in acht worden genomen. De regelmatige variatie van TEC, met een bereik tot enkele honderden TEC-units, is hoofdzakelijk een functie van de zenithoek van de zon. De onregelmatige (of niet herhaalbare) variaties zijn hoofdzakelijk golfachtige verschijnselen verbonden aan "Travelling Ionospheric Disturbances" (TID).

Alhoewel het effect van een MS-TID op de TEC klein is, soms niet meer dan 0.1 TEC-unit met een typische golflengte minder dan een paar honderd kilometer, is de MS-TEC een belangrijke hindernis voor nauwkeurige ruimtelijke interpolatie van ionosferische vertragingen voor GPS referentienetwerken. Aangezien de meeste interpolatiemethoden lineaire (of kwadratische) interpolatie of een andere lage-orde techniek gebruiken zijn deze niet in staat de fase-offset tussen punten in ionosfeer, welke veroorzaakt wordt door MS-TID, te modelleren. Er zijn twee belangrijke complicaties. Ten eerste, moet de interpolatie op het dubbel-verschilniveau (het verschil in ionosferische vertraging tussen twee satellieten en twee ontvangers) worden gedaan. Dit betekent dat twee verschillende stukken van de ionosfeer, gerelateerd aan verschillende satellieten, in de verschillen betrokken zijn: de zogenoemde ionospherische patches, elk met een omvang ongeveer gelijk aan het RTK netwerk. Iedere patch kan verbonden zijn aan verschillen TIDs. Ten tweede, voor een

operationeel RTK-Netwerk, is een real-time strategie voor de detectie en modellering van TIDs een vereiste.

In het eerste deel worden de prestaties van verscheidene empirische ionosfeer modellen, zoals Klobuchar, NeQuick en de "Global Ionospheric Maps" (GIM) van IGS, bestudeert op gemiddelde breedtes gebruik makend van GPS gegevens. De resultaten tonen aan dat de GIM de fout ten gevolge van de absolute ionosferische vertraging met 80% kan reduceren onder verschillende geomagnetische omstandigheden. Met NeQuick, dat beter dan het Klobuchar model presteerde, kon ongeveer 60% van de fout worden verbeterd. NeQuick is het real-time ionosferisch correctiemodel voor het toekomstige Europese navigatiesysteem Galileo. Een belangrijke inputparameter voor NeQuick is de effectieve ionisatiegraad, or "effective ionization parameter" (Az), die als tweede graads polynoom via het Galileo navigatiebericht aan enkel frequentie gebruikers wordt verzonden. De cofficinten van dit polynoom zullen dagelijks vanaf minstens 20 Galileo referentiestations worden geschat. Aangezien Galileo nog in ontwikkeling is en deze parameters nog niet beschikbaar zijn, stellen wij een alternatieve benadering voor het schatten van Az voor die gebruik maakt van "Global Ionospheric Maps" (GIM). De belangrijkste voordelen van de alternatieve benadering zijn: (1) de alternatieve benadering is betrouwbaarder, omdat, de GIM op gegevens van 300 wereldwijde GNSS referentiestations gebaseerd is en elke IGS GIM een combinatie is van de resultaten van vier analysecentra, (2) de cofficinten representatiever zijn omdat gebruik gemaakt wordt van een wereldwijd netwerk van 300 stations in plaats van ongeveer 20 stations, (3) met de alternatieve procedure het mogelijk is om Az in een verschillende representaties te verstrekken, bijvoorbeeld gebruikend een hogere orde polynoom, een grid, of andere functietypes, en (4) de computerinspanning is veel kleiner er vanuit gaande dat IGS GIMs reeds zijn berekend.

In het tweede deel wordt een 'normaal ionosfeer' voor de regelmatige variaties gedefinieerd gebaseerd op Chapman's theorie voor ion productie. De normaal ionosfeer bestaat uit een lager en hoger gebied. Het lagere gebied wordt gevormd in een fotochemisch evenwicht resulterend in een Chapman laag. Het hogere gebied wordt gevormd in een diffuus evenwicht hetgeen resulteert in een nieuwe Chapman laag. Invloeden van het geomagnetische veld zijn hierin niet meegenomen. Integratie van de continuteitsvergelijking van de normaal ionosfeer over de hoogte resulteed in een grenswaarde probleem (BVP) voor de tijd evolutie van de vertikale TEC (VTEC). De oplossing van dit BVP resulteert in een nieuw recursief model voor de regelmatige TEC variatie als functie van de zenithoek van de zon. De belangrijkste motivatie voor het ontwikkelen van dit model is dat de empirische modellen uit het eerste deel niet geschikt en onnodig complex waren om de regelmatige variatie van TEC voor hoge precisie differentile GNSS toepassingen te modelleren en te voorspellen. Het nieuwe model is getest op lokale en globale schaal met behulp van GIMs. De geanalyseerde gevallen tonen aan dat het model, ondanks dat het geomagnetische veld werd genegeerd, de regelmatige variatie van VTEC op gemiddelde breedtes onder geomagnetisch rustige condities goed gemodelleerd kan worden. Dit is een belangrijk resultaat dat de mogelijkheden van het model voor een aantal toepassingen aantoont. Aangezien het model een recursieve vorm heeft is het geschikt om als tijdupdate vergelijking in een dynamische gegevensverwerking of een Kalman filter te gebruiken. Een andere toepassing is het verwijderen van de regelmatige geometrie afhankelijke trends uit TEC tijdreeksen om zodoende TID te kunnen bestuderen of te modelleren. Dit wordt in

het derde deel van deze thesis beschreven.

In het derde deel, wordt een nieuw algoritme ontwikkeld voor de real-time detectie en modellering van MS-TID in de TEC. Het algoritme gebruikt de ionosferische vertragingen van de enkel-verschillen (SD) tussen twee ontvangers in een middel groot GPS netwerk. Alhoewel een deel van de geometrisch afhankelijke TEC variatie in de SD wordt geelimineert, kan het resterende deel niet worden veronachtzaamd. Om de resterende trend te verwijderen wordt de techniek uit het tweede deel van deze thesis toegepast. Wat daarna overblijft wordt verondersteld gevolgen te zijn van MS-TIDs en ruis. De aldus detrended SD is de belangrijkse input voor het TID modellering algorithme. Het algoritme bestaat uit zes stappen: initialisering, detectie, pellen, kruiscorrelatie, parameterschatting, en beindiging. Een MS-TID wordt verondersteld in een ionosferische patch van enkele honderden kilometers een vlakke reizende golf te zijn met gelijkblijvende amplitude. Alle kenmerkende parameters van de MS-TID golf (b.v. periode, fasesnelheid, propagatierichting, en amplitude) zijn tijdafhankelijk, waarbij de Doppler-verschuiving veroorzaakt door de beweging van de satellieten in de schattingsstap in acht wordt genomen. De prestaties van het algoritme zijn getest met GPS gegevens van een netwerk. Hoewel een TID in het echt geen perfecte golf is, kan het algoritme MS-TID goed detecteren en modelleren in de tijd en ruimte. De prestaties op het gebied van interpolatie en predictie voor RTK-Netwerk zijn vergelijkbaar met die van de Kriging interpolatiemethode. Dit is een belangrijk eerste resultaat, niet alleen omdat beide methodes op verschillende principes gebaseerd zijn, maar ook omdat het MS-TID detectie en modellering algoritme nog verbeterd kan worden. Dit, op fysica gebaseerd model, maakt het mogelijk om de vlakke golfbenadering te vermijden en kan rekening houden met de fase-verschillen van de golf, iets wat met Kriging niet mogelijk is.

Curriculum Vitae

Yahya Memarzadeh was born on 29 May 1969 in Tabriz, East Azarbayjan Province of Iran. He graduated from high school in 1987. In 1988, he was admitted to K.N. Toosi University of Technology, Tehran, Iran. He graduated a Bachelor of Science degree in Surveying Engineering in 1993. A few month later, he started his master study in Geodesy at K.N. Toosi University of Technology. At same time, he was employed by the National Cartography Center (NCC) of Tehran, Iran, where he was responsible for the least-square adjustment of the Iranian first-order precise leveling network. This work has resulted in the first publication of orthometric heights in Iran. During this period he wrote his master thesis 'Iranian first-order precise leveling data analysis' and graduated in 1998.

In 2003, he was entitled to an overseas PhD scholarship from the Iranian Ministry of Science, Research and Technology. He worked under supervision of Prof. Peter Teunissen and Dr. Hans van der Marel at the Delft Institute of Earth Observation and Space systems (DEOS) of Delft University of Technology, the Netherlands. This thesis covers his PhD study.

Acknowledgments

This thesis could not have been realized without the support of many people. In the first place, I would like to thank my promotor, Prof. Peter Teunissen, for giving me the opportunity to perform my PhD research. I would like to express my deepest gratitude to my supervisor, Dr. Hans van der Marel, for his support, many valuable comments, and helping me finalize this thesis.

I would also like to thank the committee members for their reviews and useful comments on my thesis.

Financial support came from the Iranian Ministry of Science, Research, and Technology (MSRT) as well as the Department of Earth Observation and Space system (DEOS) at Delft University of Technology (TU-Delft). My deep appreciation and thanks go to Dr. Rahmati, Dr. Matkan, and Dr. Mohammad Hossein Abdollahi from MSRT and my supervisor, Prof. J. Fokkema, and Prof. Ramon Hanssen from TU-Delft.

I am grateful to Prof. Sandro M. Radicella for inviting me to visit the International Center of Theoretical Physics (ICTP) in Trieste, Italy. Thanks also go to Dr. Bruno Nava and Dr. Pierdavide Cosson from ICTP for productive discussions about the NeQuick model.

Many current and former colleagues at the department of DEOS and especially in the Mathematical Geodesy and Positioning (MGP) group have directly or indirectly contributed to this thesis. My thanks go to Ria Scholtes for administrative support. She had the skill to create a great working environment. She will always reside in my heart, God bless her. A special thanks goes to Dr. Dennis Odijk for teaching me to work with the *GPSveQ* and *Virint* softwares and for sharing his knowledge of network-RTK, especially at the beginning of my PhD studies. Dr. Alireza Amiri is kindly acknowledged for our discussions about least squares processing and for providing with me his LaTeXstylefile adapted for this thesis. I am grateful to Anh Quan Le for *Matlab* scripts to process some of the GPS data. I express my appreciation to my former roommate Dr. Roderik Lindenbergh for our discussions. I also greatly appreciate Peter Buist and my roommates: Peter de Bakker, Benoît Muth, and Junlin Yan for our discussions and for proofreading some chapters of this thesis. Furthermore, I would like to thank all (former) members of the MGP group at TU-Delft for the pleasant working environment.

Being away from home (Iran) proved much harder than first anticipated. But, it has been our fortune to have friends who have made us feel at home away from home. My special thanks go to Floris Roukens and his lovely wife Alena Sokolova for all the pleasant times

with them. I would like to express my gratitude to Prof. Hans Mooij and Mrs. Annelies Mooij and Prof. Gijs Kuenen and Mrs. Vibeke Kuenen for their support and kindness. Furthermore, I would like to warmly thank all my Iranian friends in the Netherlands.

I would like to express my sincerest appreciation to my brothers and sister and her husband who is like a good brother to me. I am very thankful to Mohammad Reza Mahdawan and his family. I will never forget their kind help and support during my studies in the Netherlands.

Last but not least, I would like to express my deepest gratitude to my beloved wife, Shabnam Vosoughi Oskouei, who has always inspired me with her endless love and unconditional support. I am deeply grateful to my lovely son Ehsan for filling our life with happiness. I also express my deepest love to Shabnam and Ehsan for their understanding when I was really busy with my work and for their great patience all those weekends and evenings that I was away. Dear father and mother, my mind fails to find words that can show how deeply I am indebted to you for your unconditional love and support. Dear father- and mother-in-law, you have always been supportive and kind to me, my warm regards!

Yahya Memarzadeh
8 December, 2009
Delft

Notation and Symbols

Physics of Ionosphere

h or h'	height (altitude)
A	area
p	pressure of the Earth's atmosphere
ρ	mass density of the Earth's atmosphere
F_p	pressure force
F_g	gravity force
m	particle mass
n	number of density of molecules by volume
k	Boltzmann's constant $= 1.3806 \cdot 10^{-23}\ \frac{m^2 kg}{s^2 K}$
k	a correction factor in the relative sunspot number
T	absolute temperature
ϕ_g	descent particle flux of atmosphere from gravity
ϕ_p	expansion particle flux from vertical gradient of pressure
R	the relative sunspot number (Wolf or Zurich number)
R_{12}	12-month smoothed relative sunspot number
g	Earth's gravitational acceleration
g	the number of groups of sunspots
f	the number of individual of sunspots
$d\omega$	probability of collision between photons and particles
σ	absorption cross-section
ϕ_S^E	radiative energy flux in unit area from Sun to top of atmosphere
τ	optical depth of a gas volume
τ_E	lifetime of electrons in E-region
τ_F	lifetime of electrons in F-region
q	ion production rate in ionosphere
L	ion disappearance (or recombination) rate in ionosphere
V	bulk transport velocity of ions
e	electron charge in Coulomb $= 1.60218 \cdot 10^{-19}$
m_e	electron mass in kg $= 9.10939 \cdot 10^{-31}$
χ	solar zenith angle
η	ionization efficiency
$Ch(\chi, h)$	Chapman grazing incidence function
$q(\chi, h)$	Chapman production function
q_{max}^0	Maximum rate of ion production when the sun is overhead $\chi = 0$
n^-	number density of electron and negative ion
n^+	number density of positive ion

n or n_e	electron density or number density of electron
$n_{e,E}$	electron density in E-region
$n_{e,F}$	electron density in F-region
α	quadratic recombination coefficient
β	linear recombination coefficient
$\vec{B}$	the geomagnetic induction vector
B	total intensity of the geomagnetic field
I	magnetic inclination or dip angle
D	declination angle of the geomagnetic field
g_n^m and h_n^m	Gauss coefficients of degree n and order m in the geomagnetic field
$\vec{F}$	vector of force
$\vec{E}$	vector of electric field
$\vec{v}$	velocity vector of a charged particle
$\vec{F}^{\parallel}$	component parallel to the magnetic field
$\vec{F}^{\perp}$	component perpendicular to the magnetic field
r	Larmor radius
ω	the angular gyro-frequency of a electric charge
$D(h)$	plasma (ambipolar) diffusion coefficient at height h
w	ionospheric plasma velocity along geomagnetic field
v_{drift}	velocity of charged particles by $\vec{E} \times \vec{B}$ drift
ϵ_0	the electric permittivity of vacuum
ω_p	the angular gyro-frequency of electron
H	horizontal intensity of the geomagnetic field
H	scale height of atmosphere
H_E	scale height of atmosphere at heights of E-region
H_F	scale height of atmosphere at heights of F-region
H_n	density scale height of atmosphere
foE	peak plasma frequency of the E-region
$foF1$	peak plasma frequency of the F1-region
$foF2$	peak plasma frequency of the F2-region
NmE	maximum electron density of the E-region
$NmF1$	maximum electron density of the F1-region
$NmF2$	maximum electron density of the F2-region
hmE	height of maximum electron density of the E-region
$hmF1$	height of maximum electron density of the F1-region
$hmF2$	height of maximum electron density of the F2-region
$\dot{A}$	Angstrom $(10^{-10}m)$
nm	nanometer $(10^{-9}m)$
μm	micron or micrometer $(10^{-6}m)$
UV	ultraviolet ray
EUV	extreme ultraviolet ray
λ	wavelength of an electromagnetic signal

GPS and Ionospheric models

φ	latitude
λ	longitude
f_j	L_j GPS frequency $j = 1, 2$
c	velocity of light $\approx 299792458\ m/s$
R_e or a	the radius of the earth $\approx 6371.2\ km$
A or a	amplitude
N	vertical total electron content in the normal ionosphere
N_E	vertical total electron content in the normal E-region
N_F	vertical total electron content in the normal F-region
ϕ_{∞}^{photon}	photon flux on top of Earth's atmosphere
ϕ^{photon}	photon flux at a certain height
$v_E = \frac{n_e}{n^0_{max,E}}$	rate of electron density w.r.t. maximum electron density (when $\chi = 0$) in E-region
$v_F = \frac{n_e}{n^0_{max,F}}$	rate of electron density w.r.t. maximum electron density (when $\chi = 0$) in F-region
$V_E = \frac{N_E}{n^0_{max,E}}$	rate of N_E w.r.t. maximum electron density (when $\chi = 0$) in E-region
$V_F = \frac{N_F}{n^0_{max,E}}$	rate of N_F w.r.t. maximum electron density (when $\chi = 0$) in F-region
Dip	modeled magnetic inclination
$Dipl$	geomagnetic dip latitude (or dip latitude)
μ or $Modip$	modified geomagnetic dip latitude
$\mu \approx 1.647$	a factor that depends on the frequency ratio
$\vartheta^s_{r,j}$	code pseudo-range observable between receiver r and satellite s on L_j
$\varphi^s_{r,j}$	carrier phase observable between receiver r and satellite s on L_j
$\tau^s_{r,j,g}$	signal traveling time between receiver r and satellite s on L_j for code observation
$\tau^s_{r,j,\varphi}$	signal traveling time between receiver r and satellite s on L_j for phase observation
$p^s_{r,j}$	code observation between receiver r and satellite s on L_j
$p^s_{r,j}(t_i)^c$	a priori corrections for code observation
$\phi^s_{r,j}$	carrier phase observation between receiver r and satellite s on L_j
$\phi^s_{r,j}(t_i)^c$	a priori corrections for carrier phase observation
$\phi_{r,j}(t_0)$	initial phase of L_j in the receiver r
$\phi^s_{,j}(t_0)$	initial phase of L_j in the satellite s
T^s_r	tropospheric error between receiver r and satellite s
ρ^s_r	geometric distance between receiver r and satellite s
g^s_r	summation of all frequency independent terms in the GPS observation equation
G^s_r	geometry-dependent trend of measured slant ionospheric delay
$d^s_{,j}$	instrumental code delay for satellite s on L_j
$d_{r,j}$	instrumental code delay for receiver r on L_j
$\delta^s_{,j}$	instrumental phase delay for satellite s on L_j
$\delta_{r,j}$	instrumental phase delay for receiver r on L_j
DCB^s	differential code bias for satellite s
DCB_r	differential code bias for receiver r

DPB^s	differential phase bias for satellite s
DPB_r	differential phase bias for receiver r
$N^s_{r,j}$	integer number of full cycle between receiver r and satellite s on L_j
$M^s_{r,j}$	carrier phase ambiguity between receiver r and satellite s on L_j
$M^{s'}_{r,j}$	computed biased carrier phase ambiguity on L_j
$\varepsilon_{p^s_{r,j}}$	code observation random noise
$\varepsilon_{\varphi^s_{r,j}}$	phase observation random noise
I^s_r or I	real slant ionospheric delay between receiver r and satellite s on L_1
I_v	real vertical ionospheric delay on L_1 at a ionospheric pierce point
$\tilde{I}^s_r$ or $\tilde{I}$	GPS-provided slant ionospheric delay between receiver r and satellite s on L_1
$D\tilde{I}^s_r$	TID observation (detrended slant ionospheric delay) between receiver r and satellite s
$I^o_{ij}(t_1 : t_k)$	time series of single-difference slant ionospheric delay observations for baseline l_{ij} in a time span from t_1 to t_k
$\check{I}_{ij}(t_k)$	time series of single-difference TID observations for baseline l_{ij} in a time span from t_1 to t_k
$I^s_{r,GIM}$ or I_{GIM}	slant ionospheric delay from GIM between receiver r and satellite s on L_1
$V^s_{r,GIM}$	vertical ionospheric delay computed from GIM between receiver r and satellite s
$VTEC_{GIM}$	vertical total electron content from GIM maps
$sd\tilde{I}^s_{r_1,r_2}$	between-receiver single-difference slant ionospheric delay
$Dsd\tilde{I}^s_r$	single-difference TID observation
$dd\tilde{I}^s_{r_1,r_2}$	double-difference slant ionospheric delay
$Ddd\tilde{I}^s_r$	double-difference TID observation
ξ_{IP} or ξ	satellite zenith angle at a ionospheric pierce point
ξ'	satellite zenith angle on the ground
ψ	geocentric angle
ι	propagation effect of refraction
κ	bending effect of refraction
v_g	group velocity
v_ϕ	phase velocity
V_{ph}	phase velocity of TID
V_{IP}	velocity of ionospheric pierce point
λ_{TID}	TID wavelength
T_{TID}	TID period
f_{TID}	frequency of TID wave
f_D	Dopplered-frequency of TID
T_D	Dopplered-period of TID wave
t	time
t_k	time of current epoch
t_S	start time of TID
t_D	time of TID detection
t_M	time for stopping TID parameter estimation
t_E	ending time of TID
τ_{Gmin}	time delay corresponding to global minimum of correlation function
τ_{Gmax}	time delay corresponding to global maximum of correlation function
τ_{SGmax}	time delay corresponding to second global maximum of correlation function

φ_0	initial phase of TID wave
γ	angle between TID propagation direction and moving direction of IP point
l	baseline length
$\vec{l}$	baseline vector
f_c	cut-off frequency
n_g	refractive index for group of the signal
n_ϕ	refractive index for phase of the signal
Ω	shape factor
Az or α	azimuth
Az	effective ionization level of NeQuick model
B	thickness of a Epstein layer
$n_{max,E}$	maximum electron density of E-region
$n_{max,F1}$	maximum electron density of F1-region
$n_{max,F2}$	maximum electron density of F2-region
$h_{max,E}$	height of maximum electron density of E-region
$h_{max,F1}$	height of maximum electron density of F1-region
$h_{max,F2}$	height of maximum electron density of F2-region
$B_{top,E}$	thickness of a semi-Epstein layer for top side of E-region
$B_{bot,E}$	thickness of a semi-Epstein layer for bottom side of E-region
$B_{top,F1}$	thickness of a semi-Epstein layer for top side of F1-region
$B_{bot,F1}$	thickness of a semi-Epstein layer for bottom side of F1-region
$B_{top,F2}$	thickness of a semi-Epstein layer for top side of F2-region
$B_{bot,F2}$	thickness of a semi-Epstein layer for bottom side of F2-region
$MUF(3000)$	maximum usable frequency refracted by the ionosphere
$M(3000)F2$ or M	a propagation factor of the ionosphere
$F10.7 = CI$	solar radiation index or Covington index
CI	Covington index
Kp	planetary K_p-index of the geomagnetic activities for every 3 hours
Ap	daily planetary A_p-index of the geomagnetic activities
f	frequency of wave
T	period of a wave
λ	wavelength of a wave
ω	angular frequency a wave
$\vec{K}$	wave vector
K	wave number
$\vec{V}_{ph}$	phase velocity vector of a traveling wave
V_{ph}	phase velocity of a traveling wave
G_r^s	geometry-dependent trend of the GPS-provided slant ionospheric delay
G_0 or $G_r^s(t_0)$	geometry-dependent trend of the GPS-provided slant ionospheric delay
sdG_r^s	geometry-dependent trend of between-receiver single-difference slant ionospheric delay
ddG_r^s	geometry-dependent trend of double-difference slant ionospheric delay
m	ionospheric mapping function
δ	discrepancy
δ	the solar declination angle
T	longitude of a point measured from the sun meridian
T_r	longitude of a point measured from the sun meridian at sunrise
T_s	longitude of a point measured from the sun meridian at sunset

Mathematical and Statistical Notation and Operators

$\sum(.)$	summation
$div(.)$	divergence operator
$\nabla(.)$	gradient operator
$e = 2.718\ldots$	the Neperian number
$\breve{P}_n^m$	the normalized associated Legendre function of degree n and order m
$\|.\|$	norm of a vector
$E\{.\}$	mathematical expectation operator
$D\{.\}$	mathematical dispersion operator
σ	standard deviation
ρ	correlation coefficient
$\mathbb{R}^m$	real Euclidean space of dimension m
$\mathbb{I}_j$	$j \times j$ identity matrix
$\Re(A)$	range space of the $m \times n$ design matrix A
$N(A)$	null space of the $m \times n$ design matrix A
$\mathrm{rank}(.)$	rank of a matrix (independent columns or rows of a matrix)
trace$(.)$	sum of elements on the main diagonal of a square matrix
dim$(.)$	dimension of a matrix or vector
$\oplus$	direct sum of two subspaces
$\perp$	orthogonal complement (is orthogonal to)
$(.)^T$	transpose of a matrix
$(.)^{-1}$	inverse of a matrix
N(x, Q_x)	normal distribution with mean x and covariance matrix Q_x
$Q_{\hat{a}}$	covariance matrix of $\hat{a}$
h_v	$n \times 1$ interpolation vector
λ_v	Lagrange multiplier
e_n	$n \times 1$ vector with all elements equal one
$S_{\underline{ee}}(f)$	auto-spectral density function
$R_{\underline{xx}}(\tau)$	auto-correlation function
$R_{\underline{xy}}(\tau)$	cross-correlation function
τ	time delay measured by cross-correlation function

Neutral and charged atmospheric gases

H	neutral Hydrogen
H^+	Hydrogen ion
He	neutral Helium
He^+	Helium ion
O	neutral atomic Oxygen
O^+	atomic Oxygen ion
O_2	neutral Oxygen molecule
O_2^+	Oxygen molecule ion
O_2^-	negative Oxygen molecule ion
N	neutral atomic Nitrogen
N^+	atomic Nitrogen ion
N_2	neutral Nitrogen molecule
N_2^+	Nitrogen molecule ion
NO	neutral nitric oxide
NO^+	nitric oxide ion
NO^2	neutral Nitrogen dioxide
e	electron
q	electric charge of a particle in Coulomb

Acronyms

ACP	Area Correction Parameters
ADOP	Ambiguity Dilution Of Precision
AGW	Atmospheric Gravity Waves
APME	Average Percentage Model Error
BLUE	Best Linear Unbiased Estimator
BLUP	Best Linear Unbiased Predictor
BVP	Boundary Value Problem
C/A	Coarse/Acquisition code
CBL	Correlated BaseLines
CCIR	Comite Consultatif International des Radiocommunications
CI	Covington Index
CODE	Center for Orbit Determination in Europe
EGNOS	European Geostationary Navigation Overlay Service
DEOS	Department of Earth Observation and Space system
DD	Double Differenced
DDM	Daytime Double Maxima
DCB	Differential Code Bias
DGR	Giovanni and Radicella model
DPB	Differential Phase Bias
ESA	European Space Agency
EUV	Extreme Ultra-Violet
GEO	GEOstationary orbit satellite
GIM	Real-Time Kinematic
GLONASS	GLObal NAvigation Satellite Systems (Russian)
GNSS	Global Navigation Satellite Systems
GPS	Global Positioning System
IAG	International Association for Geodesy
IAGA	International Association of Geomagnetism and Aeronomy
ICD-GPS-200C	GPS Interface Control Document, Item # (1634) Updated in April 2000
ICTP	International Center of Theoretical Physics in Trieste, Italy
IGRF	International Geomagnetic References Field
IGS	International GPS Service
ILS	Integer Least Squares
IONEX	IONosphere map EXchange format
IP or IPP	Ionospheric Pierce point
IRI	International Reference Ionosphere
ITU-R	International Telecommunication Union - Radiocommunication sector
JPL	Jet Propulsion Laboratory

LAMBDA	Least Squares AMBiguity Decorrelation Adjustment
LSTID	Large-Scale Traveling Ionospheric Disturbances
LUE	Linear Unbiased Estimator
MAC	Master-Auxiliary Concept
MGP	Mathematical Geodesy and Positioning
MSE	Mean Squared Error
MSRT	Iranian Ministry of Science, Research, and Technology
MS-TID	Medium-Scale Traveling Ionospheric Disturbances
MUF	Maximum Usable Frequency
NASA	National Aeronautics and Space Administration
NCC	National Cartography Center
NI	Numerical Integration
NOAA	National Oceanic and Atmospheric Administration
NPQ	spherical triangle between North pole and two other points (P and Q)
P	Precision code
PDF	Probability Distribution Function
PPS	Precise Positioning Service
PRN	Pseudo Random Noise
RBL	Reference BaseLine
RINEX	Receiver INdependent EXchange format
RMS	Root-Mean-Squared error
RTCM	Radio Technical Commission for Maritime Service
RTK	Real-Time Kinematic
SBAS	Satellite Based Augmentation System
SCIGN	Southern California Integrated GPS Network
SEC	Space Environment Center
SIS-ICD	Galileo open service, Signal In Space Interface Control Document
SPS	Standard Positioning Service
SSTID	Short-Scale Traveling Ionospheric Disturbances
STD	STandard Deviation
STEC	Slant Total Electron Content
TAD	Traveling Atmospheric Disturbances
TEC	Total Electron Content
TECU	Total Electron Content Unit
TID	Traveling Ionospheric Disturbances
UPC	the Technical University of Catalonia (UPC) in Barcelona, Spain
UT	Universal Time
UTC	Coordinated Universal Time (UTC, French: Temps Universel Coordonn)
UV	Ultra-Violet
VRS	Virtual Reference Station
VTEC	Vertical Total Electron Content
WASS	Wide Area Augmentation System
WLSE	Weighted Least Squares Estimator

Contents

1 Introduction

1.1 Background

The Earth's ionosphere is an important error source for Global Navigation Satellite Systems (GNSS) signals. Free electrons, in the ionosphere, have a strong impact on the propagation of radio waves resulting in range errors on the GNSS signals. For single-frequency GNSS receivers, the range error caused by the ionosphere is currently the largest component in the error budget affecting the positioning accuracy. The Total Electron Content (TEC) is an integrated quantity that represents the number of free electrons in a column of 1 m^2 cross-section along a path through the ionosphere. The ionospheric delay is almost proportional to TEC along the signal path and inversely proportional to the frequency squared. This dispersive property of the ionosphere allows dual frequency GNSS receivers both to compensate for the ionospheric delay error and to measure TEC. The measured TEC along the signal path, also known as slant TEC (STEC), is often mapped into vertical TEC (VTEC) using a mapping function. This mapping is not perfect due to horizontal variations in electron density in the ionosphere but is a useful approximation assuming all free electrons are concentrated in a single layer (single-layer ionosphere approximation).

The ionosphere is a highly dynamic medium and the electron density can vary significantly from minute to minute at a given location resulting in temporal and spatial variations of TEC. The temporal variation of TEC is a combination of *regular* and *irregular* variations . The regular variation of TEC is associated to the diurnal and seasonal changes in the Earth-Sun geometry (or in solar zenith angle) and changes in the solar ionizing radiation intensity over a solar cycle (11 years). The daily regular variation in TEC at a given location more or less repeats daily because of the earth's rotation, although there are significant differences from day to day due to other effects. The irregular (or non-repeatable) variation refers typically to the effects of the Traveling Ionospheric Disturbances (TID) and ionospheric or geomagnetic storms. TIDs are wavelike fluctuations in electron density that propagate through the ionosphere and cause wavelike fluctuations in TEC.

Over the past decade, several numerical and empirical models for the regular variations of TEC have been developed at regional and global scales. Numerical ionospheric models use different techniques, such as polynomial expansions, grid-based techniques and spherical harmonics in latitude and longitude, to model VTEC on one or more ionospheric layers, as function of time. The parameters for these models are estimated from slant TEC (STEC) measurements. STEC is converted into VTEC using a mapping function, and vise versa. An example of such model is the GIM. Another example is the model broadcasted by SBAS (Satellite Based Augmentation System) systems.

Empirical models try to model the regular variations in space and time using fewer parameters than the numerical models. These empirical models are often build from the analysis of long time series and some elementary physics of the ionosphere. An example of an empirical model is the NeQuick model which uses one parameter, the effective ionization level. The irregular variations in TEC neither numerical nor empirical models are capable of capturing. Although irregular variations and disturbances in TEC are usually quite small, they degrade the reliability of modern GNSS techniques.

Dual frequency GNSS user can in single-point (absolute) positioning application eliminate the ionosphere delay by forming a simple linear combination of the dual-frequency observations. However, single frequency users must always rely on ionospheric delay models. The Global Positioning System (GPS) uses the Klobuchar model, which is able to correct for $50\% - 60\%$ RMS (Root Mean Square) error of the total ionospheric delay. The parameters for the Klobuchar model are broadcasted by the GPS navigation message. The future European GNSS system Galileo will support single-frequency users using the more complicated NeQuick model. The effective ionization level will be broadcasted as part of the Galileo navigation message.

In case single frequency users have a SBAS (WASS, EGNOS) enabled receiver they can also use the numerical model provided by the SBAS system. For post processing applications also GIM (Global Ionospheric Maps) can be used.

Absolute positioning with GPS is generally much less accurate than relative positioning. Many of the errors occurring in GNSS are highly correlated between receivers and these can be eliminated or reduced by differencing the observations of the two receivers. For high precision (cm-level) relative GPS positioning, it is also essential to use the very precise, but ambiguous, carrier phase observables instead of the pseudorange observations. The ionospheric delay is one of the errors that is reduced by relative positioning. The differential ionospheric delay error depends on the distance between both receivers. Although on the one hand the ionospheric delay error is reduced, on the other hand, the increase in accuracy of relative positioning makes the application also more sensitive to ionosphere delay errors. Only for very short ($< 10 - 30$ km) baselines the ionospheric delay errors can be ignored. For longer baselines ($> 50 - 100$ km) the ionospheric delay must be modeled or estimated using a dual frequency receiver. Also, relative positioning applications are more sensitive to irregular variations and disturbances in the ionospheric delay than absolute positioning applications.

In high precision relative positioning applications the ambiguous carrier phase observations are used. Therefore, in addition to the receiver position and clock error, also the so-called double difference carrier phase ambiguity has to be estimated. This requires the receiver satellite geometry to change significantly which results in a long observation time span (at least 30-50 minutes), unless the double differenced ambiguities can be resolved by some other method.

In the fast precise (cm-level) GPS applications, such as engineering surveying, cm-level precision has to be achieved using very short time span GNSS observations otherwise it would not be economically usable. This is feasible only when the double-differenced (DD) ambiguities are fixed on their integer values. This can be achieved through special integer estimation techniques such as the LAMBDA method developed at TU-Delft. However,

ambiguity resolution may fail, or result in the wrong integer estimates, in the presence of biases at the double difference level. It is well known that the ionosphere decorrelates as function of the baseline length. For a sufficiently short baseline (typically less than 10 km) the ionosphere becomes fully correlated and results in a very small relative ionospheric delay at the double difference level. For a larger baseline, due to the ionospheric decorrelation, the ionospheric delay at the double difference level is increased and may cause ambiguity resolution to fail. In this case, it is therefore necessary to compute a priori precise (cm-level) correction, for the DD ionospheric delay.

Modern GNSS applications, such as Network-RTK, provide such precise ionospheric corrections. The corrections at the user site are obtained by spatial interpolation of ionospheric delay estimates at GNSS reference stations from a regional permanent GPS network in the vicinity of the user's receiver (van der Marel, 1998). When the interpolated ionospheric corrections are sufficiently precise, instantaneous ambiguity resolution becomes feasible for the user.

In the past, several methods have been developed for spatial ionospheric interpolation. All of these methods use either a sort of linear interpolation or fit a lower-order surface approach. There is not any method which is able to provide a spatial prediction of ionospheric correction with an acceptable level of accuracy under all ionospheric conditions. Most methods only work well under quiet ionosphere conditions, up to several 10s of kilometers distance. In order to achieve sufficient accuracy for spatial ionospheric prediction, and increase the reliability of the Network-RTK systems, under almost all ionospheric conditions either a dense network of reference stations must be deployed or the method for spatial prediction of ionospheric delays has to be improved.

Medium-Scale TIDs, with a typical wavelength less than a few hundred kilometers, are one of the main obstacles for accurate spatial interpolation of ionospheric induced delays in a medium-scale reference GPS network (baseline length less than a few hundred kilometers). This is due to the fact that the spatial ionospheric prediction methods are not capable of handling the phase-offset caused by TID, between the different ionospheric pierce points. To achieve the cm-level level in accuracy, TIDs should be taken into account in the spatial interpolation. There are two major complications of the TID mitigation for Network-RTK that should be addressed. Firstly, mitigation must be taken care of at the double-difference level, which involves taking single differences between ionospheric delays for the same satellite between two different receivers, followed by differencing single differences for different satellites. This means that two different patches of the ionosphere are involved, each related to a different satellite, and each possibly associated with different TIDs. Secondly, for operational Network-RTK, a real-time strategy for TID detection and modeling is needed.

1.2 Research objectives

The main objective of this thesis is to develop an algorithm to model the regular and irregular variations of the ionospheric total electron content with the goal to predict ionospheric total electron content in space and over time. The algorithm should be suitable for ionospheric delay mitigation in Network-RTK systems. At the onset of this research we believe that, in order to improve the reliability of the Network-RTK, it is necessary

to transfer the high temporal resolution of TEC data from GNSS reference stations into the spatial domain. Furthermore, we believe that the modeling and prediction can be improved by including more information on the ionospheric physics in the algorithms. These two believes have been important guidelines during the remainder of the research. Other guidelines were of a more practical nature:

- Concerning to the regular variation, the model should be developed preferably in a recursive form. Then it can be used for ionospheric forecasting with the Kalman-filtering technique. The forecasting model should be able to model the regular ionosphere variation as a function of the solar zenith angle with high relative accuracy in the spatial domain.
- Concerning to the irregular variation, which are defined as fluctuations with respect to the forecast model, the algorithm should be able to detect and model at least Medium-Scale TIDs in real-time.

Two restrictions were applied during the research:

- The research focused on ionospheric conditions for the mid-latitude regions. The ionosphere is more active in equatorial and auroral regions and consequently the ionospheric errors in GPS observations collected in those regions are larger and more variable, in general. Possible adaptation of the model for these regions have not been considered, was not the model tested for these regions.
- Another restriction is that the distance between GPS reference stations is not allowed to be unlimited. The research is restricted to a baseline which is at maximum a few hundred km long.

1.3 Outline of the thesis

This thesis consists of 8 chapters, *Chapter 1* is an introduction.

Chapter 2 gives an overview of the physics related to the Earth's atmosphere and the Sun. A brief introduction to the earth's real magnetic field and its activity indexes are given, since the magnetic field plays a central role in the transport process in the ionosphere.

In *Chapter 3*, theory of the Earth's ionosphere formation is explained based on the Chapman grazing incidence function. Three principal photochemical processes: production, disappearance and transport of ions and electrons involving in the ionospheric plasma continuity equation are reviewed. The combination of these three processes explains the temporal variation of the electron density in the ionosphere. The last part of this chapter also discusses ionospheric storm and the traveling ionospheric disturbances.

In *Chapter 4* the propagation of the GPS signals through the ionosphere is reviewed. The total electron content plays a central role in the GNSS signal propagation in the ionosphere. Extracting ionospheric information from GPS data is described as well and different types of the ionospheric models are addressed.

Chapter 5 investigates the empirical three dimensional NeQuick model of ionospheric electron density. The NeQuick model has been proposed as real-time ionospheric correction model for single-frequency users of the future Galileo navigation system. In this chapter, the NeQuick's formulation and its characteristic parameters are described. The only model parameter of NeQuick is the effective ionization level (called Az parameter) that introduces the solar daily activity into the model. As the Galileo system is under development, daily values of the Az parameter are not yet available in the standard approach. Because of this, we present an alternative approach for estimation of the Az parameter using the Global Ionospheric Maps (GIM). In order to investigate the validity of the alternative approach, the performance of NeQuick model is studied using the Az parameter estimated by the standard and alternative approach. For comparison, the Global Ionospheric Maps (GIM) and Klobuchar model are included in this investigation.

Chapter 6 focuses on a novel method for modeling of the vertical total electron content (VTEC) based on ionospheric physics, introduced in chapter 3. A normal ionosphere, including an ideal E-region (lower ionosphere) and ideal F-region (upper ionosphere), is defined by considering an isothermal atmosphere with an exponentially density distribution in height. For simplicity, the geomagnetic field is ignored in definition of the normal ionosphere. The integration of the continuity equation of each region leads to a differential equation for VTEC for the corresponding region. The solution of this differential equation results in VTEC as a function of solar zenith angle. The solutions for each of the ideal E- and F- regions are discussed separately. Summation of the two solutions leads to a recursive model as function of the solar zenith angle. The temporal and spatial variations of the solar zenith angle results in similar VTEC variations. The model consists of three unknown parameters: the linear recombination coefficient, the solar radiation intensity and VTEC at t_0, these parameters can be estimated by he least-squares fitting the model to a time series of VTEC observations. The performance of the model is tested locally and globally using VTEC observations provided from GIM maps. The results show that the model gives a good overall representation of the regular VTEC variation versus the solar zenith angle.

In *Chapter 7*, an algorithm is developed for the irregular variation of the TEC that is particularly associated with wavelike fluctuations due to Traveling Ionospheric Disturbances (TID). Time series of TID observations are computed by removing the regular variations (geometry-dependent trend) from the slant ionospheric delays measurements. The detrending is carried out using the recursive model that developed in Chapter 6. The algorithm uses the cross-correlation between many pairs of the time series of single-difference (between-receiver) TID observations in a GPS network for TID detection and estimation of the TID wave parameters. In the estimation of the wave parameters, the Doppler-shift due to the satellite motion is taken into account. The real-time application of the algorithm is discussed and the performance of the algorithm is tested using real TEC observations for two different satellites.

Finally in *Chapter 8*, the conclusions of this thesis are summarized and recommendation are given for future research.

1.4 Contributions of this research

The main contributions of this research can be summarized as follows:

- Global Ionospheric Maps (GIM) are used to estimate the effective ionization level of the NeQuick model. This is an alternative approach, that provided valuable opportunities to investigate temporal and spatial variations of the effective ionization level. It is shown that a second order polynomial, as proposed for the Galileo program, is not the best-fitting function to describe spatial distribution of the effective ionization level. The correlation between the effective ionization level and solar radiation intensity index (F10.7) was not very high. In order to use the NeQuick model as an ionospheric correction model for single-frequency GNSS data, it is possible to use the predicted GIM to provide daily model parameters.

- A normal ionosphere is defined to express the regular variation of electron density in the Earth's ionosphere as a function of the solar zenith angle. Integration of the electron density in the normal ionosphere over height leads to a novel recursive model for the regular variation of VTEC as a function of the solar zenith angle. The model is physic-based and consists of two physical parameters that represent the solar radiation intensity and the recombination coefficient. The main advantage of the novel model is that it is physic-based (not an empirical or a numerical model), in a recursive form, and therefore can be used as a forecasting model. Another advantage of the model is that the integration term of VTEC is tabulated for grid points and does not require the use of a numerical integration method and is therefore computationally efficient.

- The novel recursive model of VTEC can be used to produce time series of TID observations from GPS slant ionospheric delays. The model parameters are computed using a least-square estimation. The time series of single-difference (between-receiver) TID observations has been used to analyse TID wave propagation. Equations for estimation of the TID wave parameters are derived for the case that the single-difference TID observations are used.

- A real-time algorithm for TID detection and modeling is developed using TID observations in a reference GPS network. The algorithm is in a form that can be used for mitigation of TID in Network-RTK. This algorithm can also be applied for real-time TID monitoring along the satellite tracks belonging to a moving ionospheric patch associated to the network of ionospheric pierce points for a GNSS satellite.

The Earth's Atmosphere, Sun, and Geomagnetism

2

This chapter deals with the the physical background of the Earth's atmosphere, the geomagnetism and the Sun. These are the three components that involved in the formation of an ionized region in the Earth's atmosphere. Understanding the features of these components is essential to an understanding of the formation of the Earth's ionosphere which will be discussed in the next chapter.

2.1 The Earth's Atmosphere

The *Earth's atmosphere* is a mixture of different gases and small particles. The atmosphere can roughly be defined as the region from sea level to about 1000 km altitude around the Earth, where neutral gases can be detected, although traces of atmospheric gases have been detected far into space. 99% of the mass of the atmosphere lies below about 30 km altitude. Above 80 km altitude the atmosphere contains ionized molecules and free electrons.

The Earth's atmosphere can be coarsely subdivided in several concentric spherical layers based on different characteristic features such as temperature, ionization, and propagation. Based on the vertical variation of temperature the Earth's atmosphere is subdivided into four layers: *troposphere* (from sea level to about 10 km), *stratosphere* (from 10 km to about 50 km), *mesosphere* (from 50 km to about 80 km) and *thermosphere* (from 80 km to about 400 km). In general, the temperature in the troposphere decrease with height and increases in the stratosphere, but decreases again in the mesosphere. In the thermosphere the temperature increases again. Above the thermosphere, the temperature is constant, is the *exosphere* which gradually merges into space. The exosphere is the uppermost layer of the Earth' atmosphere that the particles can escape to space. The temperature variation through the atmosphere layers is shown in figure 2.1.

With respect to signal propagation, the atmosphere is subdivided into two main layers of *troposphere* (also referred to as neutral atmosphere) and *ionosphere*. The troposphere, the lower part of the atmosphere, extends from the see level to about 40 km which is *non-dispersive* medium (the propagation delay is not frequency dependent), see figure 2.1. In the troposphere, signal propagation depends mainly on the water vapor content and on temperature. The ionosphere, the upper part of Earth's atmosphere, is a *dispersive* medium which starts about 80 km altitude. In the ionosphere, signal propagation is mainly affected by free charged particles.

The two most abundant atmospheric gases are Nitrogen (78% by volume) and Oxygen

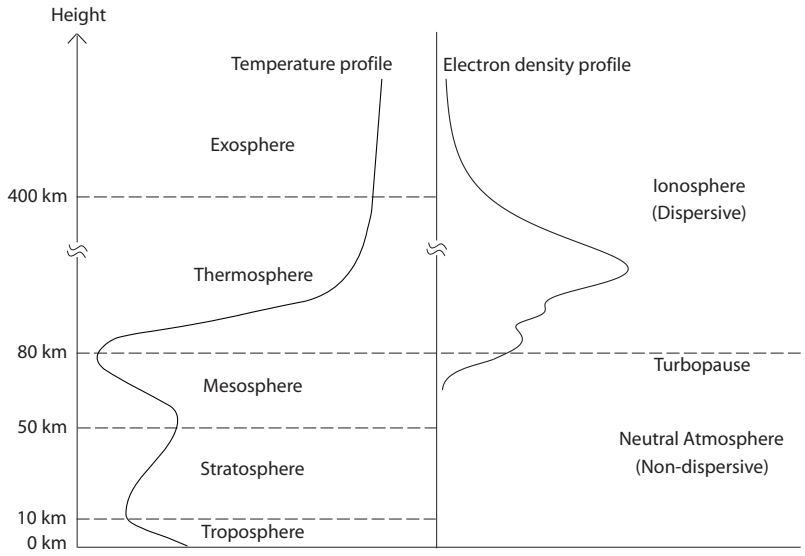

Figure 2.1: Subdivision of the Earth's atmosphere

(21% by volume), and together they compose over 99% of the lower atmosphere. The remaining 1% of the atmospheric gases includes the noble gases (Argon, Neon, Helium, Krypton and Xenon) and so-called *greenhouse* gases (Carbon Dioxide, Methane, Nitrous oxide and Water vapor). Hydrogen is also present in the atmosphere, but because it is so light, over time much of it has escaped the Earth's gravity into space.

Below about 85 km, the atmospheric gases are mixed well by turbulence and their relative proportions remain constant. At greater heights, where turbulence does not play role, each atmospheric component is in *diffusive equilibrium*. The height at which turbulent mixing is replaced by diffusive equilibrium is called the *turbopause*.

2.1.1 Pressure, temperature and density variations

The atmosphere is not physically uniform and has significant variations in temperature and pressure with altitude.

Hydrostatic Equation The *hydrostatic* or *aerostatic equation* describes the change of pressure p with height as a function of mass density ρ and Earth's gravitational acceleration g. To derive the hydrostatic equation, consider an area A at height h that is subject to the pressure from the gas below on the bottom side and to the weight of gas column above on the upper side (see figure 2.2). The pressure force from below at height h is

$$F_p = A\,p(h) \tag{2.1}$$

and the weight of gas column above h is calculated from

$$F_g(h) = A \int_h^\infty \rho(h')\, g(h')\, dh' \tag{2.2}$$

In hydrostatic equilibrium the pressure and gravity forces are equal and of opposite direction,

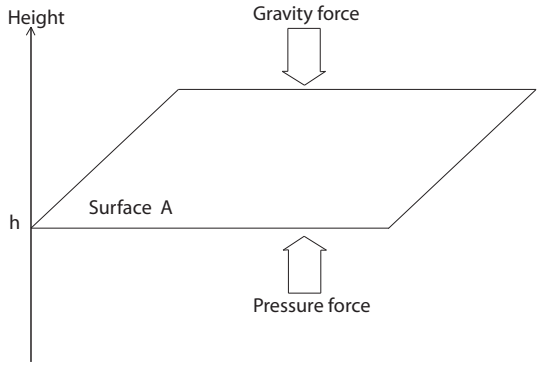

Figure 2.2: Deriving the hydrostatic equation

i.e. $F_p = F_g$. The pressure as a function of height is then

$$p(h) = \int_h^\infty \rho(h')\, g(h')\, dh' \tag{2.3}$$

The differential form of equation 2.3 is known as the *hydrostatic* equation

$$\frac{dp(h)}{dh} = -\rho(h)\, g(h) \tag{2.4}$$

The hydrostatic equation is valid for a single gas atmosphere as well as an atmosphere composed of various gases, regardless of whether the atmosphere is well mixed or gravitationally separated.

Barometric equation The hydrostatic equation expresses the relationship between pressure and mass density, which are both height dependent. To provide individual functions for pressure and density with respect to height, the vertical temperature profile need to be known and with help of the ideal (perfect) gas law $p = knT$, with n is the number density of molecules by volume, T is absolute temperature and k is, the mass density in a unit volume is equal to

$$\rho(h) = m(h)\; n(h) = m(h)\frac{p(h)}{k\; T(h)} \tag{2.5}$$

with m is the molecular mass. Inserting equation 2.5 into the hydrostatic equation yields

$$\frac{dp(h)}{dh} = -\frac{m(h)g(h)}{kT(h)}p(h) = -\frac{p(h)}{H(h)} \tag{2.6}$$

with

$$H(h) = \frac{kT(h)}{m(h)g(h)} \tag{2.7}$$

the *scale height* of the atmosphere (which is height dependent).

Separating p and h in equation 2.6 and integrating over the height interval from h_0 to h, where h_0 is an arbitrary height and h is any given height, results in the following relation

$$\int_{p(h_0)}^{p(h)} \frac{dp(h')}{p(h')} = \ln\frac{p(h)}{p(h_0)} = -\int_{h_0}^{h} \frac{dh'}{H(h')} \tag{2.8}$$

or

$$p(h) = p(h_0) \exp\left\{ - \int_{h_0}^{h} \frac{dh'}{H(h')} \right\} \tag{2.9}$$

Equation 2.9 is the well-known *barometric equation* for the vertical pressure profile in an atmosphere. An analogous barometric equation for the number density profile can be given if equation 2.5 is used to replace the pressure with the number density

$$n(h) = n(h_0) \frac{T(h_0)}{T(h)} \exp\left\{ - \int_{h_0}^{h} \frac{dh'}{H(h')} \right\} \tag{2.10}$$

which can be re-written

$$n(h) = n(h_0) \exp\left\{ - \int_{h_0}^{h} \frac{dh'}{H_n(h')} \right\} \tag{2.11}$$

with H_n the *density scale height*. The density scale height is defined by

$$H_n(h) = \left(\frac{|dn(h)/dh|}{n(h)} \right)^{-1} \tag{2.12}$$

similar to the definition of the in equation 2.6. The relationship between the two scale heights is

$$\frac{1}{H_n(h)} = \left(\frac{1}{H(h)} + \frac{1}{T(h)} \frac{dT(h)}{dh} \right) \tag{2.13}$$

usually at high altitudes (above 200 km) the second term on the right hand side is small compared to the first term, so that the density scale height corresponds closely to the pressure scale height. In an isothermal atmosphere, the two scale heights are equal.

2.1.2 Diffusive equilibrium

The Earth gravity field causes a downward acceleration to particles in unit volume of atmosphere whereas the pressure vertical gradient between lower and upper faces of the volume accelerates the particles in the direction of decreasing density, i.e. away from the Earth. A static density distribution can only be maintained when both fluxes, the descent flux from gravity ϕ_g and outward expansion flux ϕ_p are equal (see figure 2.3). This condition is known as *transport* or *diffusive equilibrium*. Each volume in atmosphere is in diffusive equilibrium. Below the turbopause, where the gases are completely mixed, all atmospheric gases have the same height variation, so that the mixture is distributed based on equation 2.11 with m the average mass of the molecules or atoms. Above the turbopause, the vertical distribution of each atmospheric gas is considered separately based on its own scale height (see figure 2.4). In the absence of chemical reactions, the density of each atmospheric gas diminishes exponentially with height. Since the scale height is inversely proportional to the mass and proportional to the temperature, the relative vertical distribution of the atmospheric gases above the turbopause is function of height. The heavy particles (like O_2 and N_2) dominate the lower altitudes and the lightest particle, hydrogen, dominates the higher altitudes.

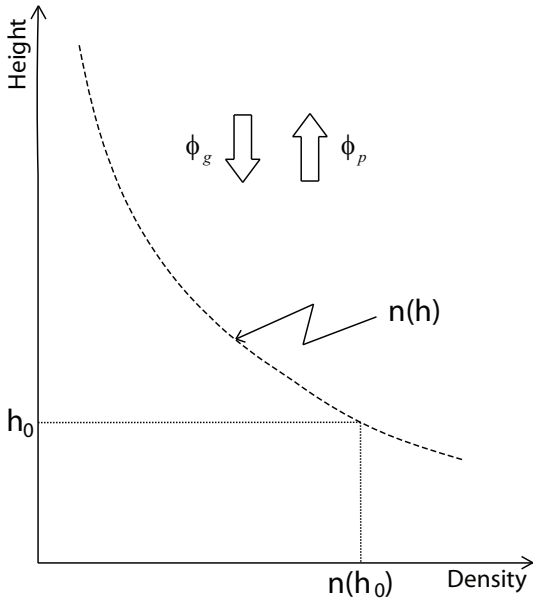

Figure 2.3: Diffusive equilibrium: ϕ_g and ϕ_p are respectively the descent and expansion fluxes

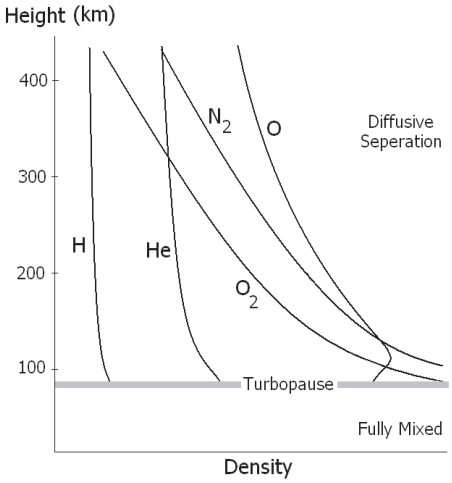

Figure 2.4: Typical vertical density profile of the neutral atmospheric gases

2.1.3 Upper atmosphere

Due to solar radiation the molecules in the upper atmosphere are heated and electrons are released from them. It results in an ionized medium that is called the ionosphere. Not all the particles in the ionosphere are ionized, but the ionosphere consists of a mix of uncharged and charged atmospheric particles. Although less than 1% of the mass of atmosphere lies in the upper Earth's atmosphere, the free electrons in ionosphere are sufficiently numerous to influence the propagation of radio waves. Because of this the ionosphere plays important role in radio communication and satellite navigation systems. The rate of ionization depends on the density of gas molecules and the intensity of the radiation (see chapter 3).

2.2 The Sun

The Sun is the largest object in the solar system. The Sun with a radius of 696000 km contains more than 99.8% of the total mass of the solar system. The Sun is composed mainly of hydrogen and helium. The Sun emits huge amounts of energy and mass (roughly $4 \times 10^{33}\ erg/sec$) that is produced by *nuclear fusion* reactions ($Hydrogen$ converted to $Helium + gamma$ rays) in its interior. The Sun consists of the solar interior (*core*), the visible surface (*photosphere*), the lower solar atmosphere (*chromosphere*), and outer solar atmosphere (*corona*). The upper layers of the Sun rotate with an angular velocity which depends on the *heliocentric latitude*. In the photosphere the rotation is faster at the *solar equator*. The rotation period at the solar equatorial latitudes is about 27 days. The temperature at the Sun's core is about 15×10^6 K and decreases toward the Sun' surface. Temperature in the photosphere is about 5800 K, but there are regions with lower temperature (4000 K) which are seen by observers on the Earth as dark sunspots.

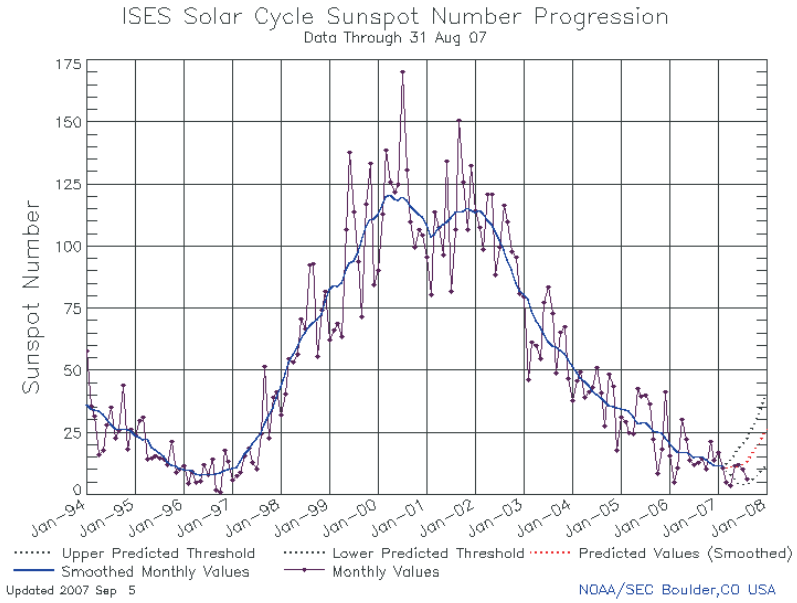

Figure 2.5: Solar cycle 23: Measured and predicted sunspot numbers (Graph courtesy of NOAA Space Environment Center (SEC), USA)

Sunspots Sunspots are cool regions on the photosphere which emit significantly less radiation than their hot surroundings and therefore appear dark only by comparison with the surrounding regions (Prolss, 2004). Sunspots can be very large, as much as 40000 km in diameter. Sunspots are caused by complicated and not very well understood interactions with the Sun's magnetic field. Sunspots tend to group together. In order to account for the statistical appearance of sunspots, the relative sunspot number (also called the *Wolf* or *Zurich number*) is used $R = k(10g + f)$ where g is the number of groups of spots (a group may include one or more sunspots), f is the number of individual spots, and k is a correction factor, usually less than unity, which depends on the observer and is intended as a correction to get the original scale by Wolf (Davies, 1989). The most widely used index in ionospheric work is the 12-month smoothed relative sunspot number $R_{12}(n)$ defined by

$$R_{12}(n) = \left[\sum_{i=n-5}^{n+5} R_i + 0.5(R_{n+6} + R_{n-6}) \right] \Big/ 12 \tag{2.14}$$

in which R_i is the mean value of R for month i.

Sunspots appear and disappear over time and R displays systematic variations that provide useful information on the state of the Sun. The magnitude of R varies between zero and 200, with a period of approximately 11 years; the so-called *sunspot cycle* or *solar cycle*. In figure 2.5 R is plotted for the current solar cycle (no. 23). Solar Minimum refers to the several years when the sunspot numbers are the lowest and Solar Maximum refers to the years when sunspots are the most numerous. During Solar Maximum, activity on the sun and its effects on our terrestrial environment are high. For example, the frequency and intensity of geomagnetic storms and radiation showers in the earth's atmosphere increases during solar maximum.

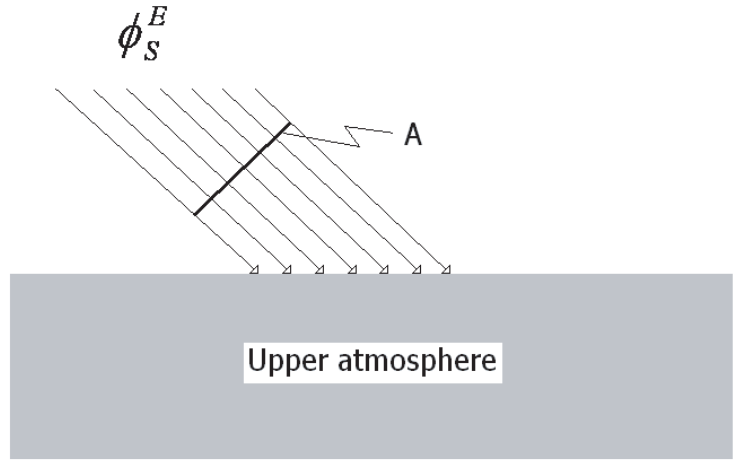

Figure 2.6: Energy flux of the solar radiation

2.2.1 The Solar radiation

Huge amounts of energy are continuously released from the Sun by both electromagnetic radiation (photon) and particle outflow (*protons* and electrons).

Photon radiation The energy transported from the Sun to Earth by the electromagnetic radiation is described by the radiative energy flux ϕ_S^E as the energy per second and passing through a unit area perpendicular to the Sun's rays above the atmosphere (see figure 2.6).

The Sun radiates electromagnetic waves over a wide range of wavelengths, including the X-ray, ultraviolet, visible, infrared and radio waves. The total radiated energy per second in all wavelengths is approximately constant. The total energy at top of the Earth's atmosphere (at a Sun-Earth separation of 1 astronomical unit) is 1370 watts per meter squared and is called the *solar constant*. Table 2.1 shows the contribution of the different wavelengths to the solar constant. The major energy contributions are from the infrared (52%), visible (41%), and ultraviolet (<7%) spectral regions. The radio and X-ray wavelengths make only a minor contribution (<1%) to the solar constant (Schunk and Nagy, 2000; Tascione, 1994).

Table 2.1: The contribution of the solar spectral regions to the solar constant (Note: $\dot{A} = 10^{-10} m$)

Solar spectral region	Wavelength range	Contribution
Radio	$1\,mm <\ \lambda$	$< 0.1\%$
Far Infrared	$10\,\mu m <\ \lambda\ < 1\,mm$	52%
Infrared	$0.75\,\mu m <\ \lambda\ < 10\,\mu m$	
Visible	$0.3\,\mu m <\ \lambda\ < 0.75\,\mu m$	41%
Ultraviolet (UV)	$1200\dot{A} <\ \lambda\ < 3000\dot{A}$	$< 7\%$
Extreme Ultraviolet (EUV)	$100\dot{A} <\ \lambda\ < 1200\dot{A}$	0.1%
X-rays	$\lambda\ < 100\dot{A}$	$< 0.1\%$

Particle radiation The Sun releases about one billion kilograms of the energetic charged particles (mainly protons and electrons) every second into space. This is the *solar wind*. The velocity of the particles is about 300 km/sec so it takes about 4-5 days to reach the Earth. The solar wind is mainly sustained by a continuously outflow of plasma from the Sun's corona. A portion of the solar particle radiation is related to powerful events in the Sun's atmosphere above sunspots, called *solar flares*. The received total energy from solar particle radiation at top of the Earth's atmosphere is only about one-tenth of the solar photon radiation energy from the X-ray and EUV spectral regions (Ratcliffe, 1972).

2.2.2 Variation of the radiation intensity

The total solar radiation per second in all wavelengths is approximately constant. Within the spectral range of maximum energy flux (the visible and neighboring infrared and ultraviolet ranges), the intensity of the solar radiation is practically constant, varying by less than 0.3% (Prolss, 2004). The minor radiation sources (the radio, extreme ultraviolet and X-ray spectral regions) display large fluctuations depending on solar activity.

Solar flares and rapidly moving plasma clouds in the Sun's corona are more frequent during solar maximum, and therefore solar particle radiation showers near the earth are also more frequent and more intense during the period of solar maximum. There is also a general increase in solar wind during the solar maximum. Solar activity varies not only during the solar cycle, but also from day to day. Therefore radiation will also vary from day to day. The sunspot number is a useful way to describe solar activity in a quantitative sense and provides an approximate measure of the solar short wavelength (the X-ray and UV) radiation intensity. Another index that is very useful is the so-called solar radiation index $F10.7$.

2.2.3 Solar radiations index ($F10.7$)

The solar flux on the radio wavelength of 10.7 cm (2800 MHz) is well correlated with X-ray, EUV, and UV fluxes. The flux on wavelength of 10.7 cm is measured daily with a reflector of 1.8 m diameter at the Algonquin Radio Observatory, near Ottawa at 17:00 UT in units of $10^{-22}Wm^{-2}Hz^{-1}$ (solar flux unit). The 10.7 cm radio flux is known as F10.7 index or *Covington index* (CI) and varies from a minimum near 65 (corresponding to sunspot number zero at solar minimum) to a maximum of about 200 corresponding to a sunspot number of about 150 to 160 (Davies, 1989). Because of correlation with X-ray, EUV and UV fluxes the F10.7 is one of the most commonly used indicator for solar activity.

The F10.7 index displays similar variations as the sunspot number (see figure 2.7). An empirical formula to convert the smoothed relative sunspot number (see equation 2.14) into monthly averaged F10.7 index is provided by the *Radiocommunication* sector of the *International Telecommunication Union* (ITU-R) as follow (Leitinger et al., 2005):

$$F_{10.7} = 63.7 + (0.728 + 0.00089R_{12})R_{12} \tag{2.15}$$

And wise versa

$$R_{12} = (167273 + (F_{10.7} - 63.7)1123.6)^{0.5} - 408.99. \tag{2.16}$$

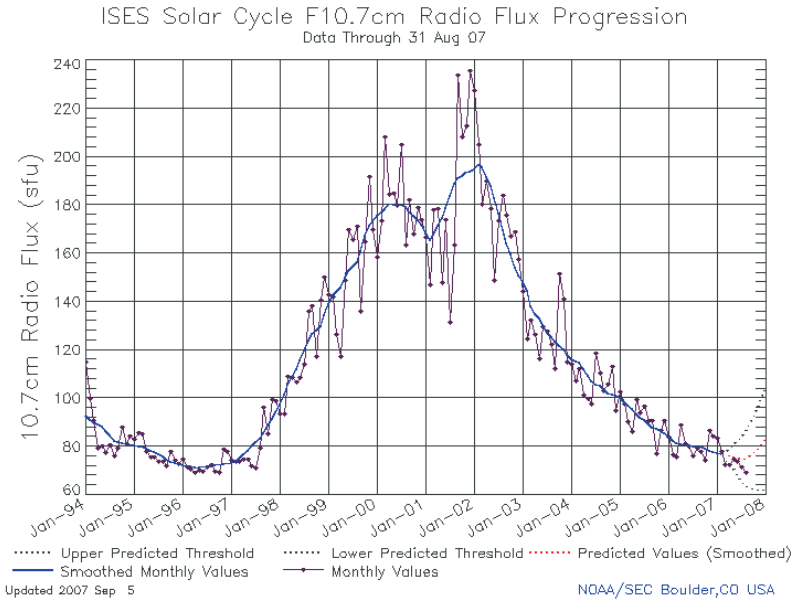

Figure 2.7: Progression of solar cycle 23: Measured and predicted F10.7 (Graph courtesy of NOAA Space Environment Center (SEC), USA)

2.3 Geomagnetism

2.3.1 The earth's magnetic dipole field

Electric currents in the Earth's core produce a magnetic field around the Earth. In the absence of any external forces (solar wind), the geomagnetic field near the earth can be approximated by a simple that located in the earth's center with an axis tilted about $\delta = 11.5°$ degrees from the earth's rotation axis (see figure 2.8) (Davies, 1965; Prolss, 2004). The geocentric dipole axis intersects the surface of the Earth at the so called north and south geomagnetic poles.

The *geomagnetic field vector*, or the *geomagnetic induction vector*, $\vec{B}$ is related to the so-called *magnetic elements* B, H, I, D, X, Y, and Z (see figure 2.9). X, Y, and Z are three field components along orthogonal directions with positive values for geographic north, east, and down. B and H are the total and horizontal field intensities (unit: $Tesla = VoltSecond/Meter^2$). The *declination angle* D is the angle between geographic North and the geomagnetic North. The angle that $\vec{B}$ makes with the horizontal plane is called the *magnetic inclination* or *dip angle* I. The *geomagnetic latitude* φ is given by

$$\varphi = \arctan(\frac{1}{2}\tan I) \tag{2.17}$$

The magnetic inclination is independent of the height and is positive in the Northern Hemisphere. At the geomagnetic poles the inclination is $90°$. The circle on the surface of the Earth with inclination $0°$ is called the *magnetic* or *dipole equator*.

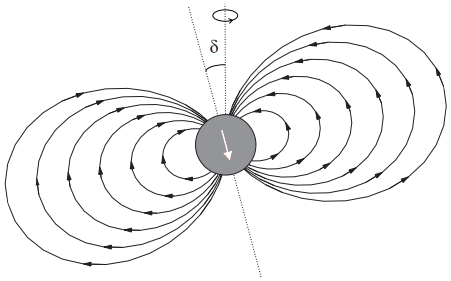

Figure 2.8: Geocentric magnetic dipole field lines around the Earth

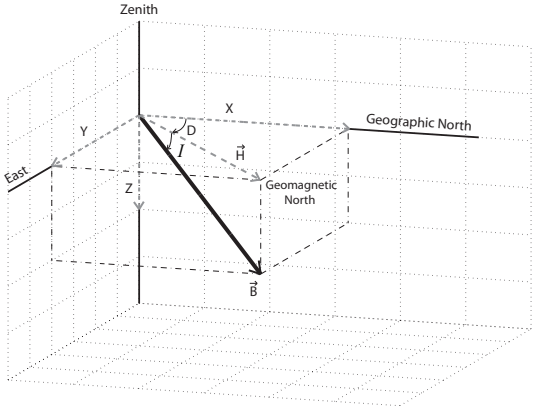

Figure 2.9: Components of the geomagnetic field in a point located the Northern Hemisphere (total field vector inclined into the Earth

2.3.2 The real geomagnetic field

The real geomagnetic field departs significantly from the geocentric magnetic dipole field. The real geomagnetic field is a conservative force-field and can be written as the negative gradient of a scalar geomagnetic potential field

$$\vec{B}(\varphi, \lambda, r, t) = -\nabla V(\varphi, \lambda, r, t) \tag{2.18}$$

The geomagnetic potential can be expanded in terms of spherical harmonics (NOAA, 2005)

$$V(\varphi, \lambda, r, t) = a \sum_{n=1}^{\infty} \left(\frac{a}{r}\right)^{n+1} \sum_{m=0}^{n} \left[g_n^m(t) \cos(m\varphi) + h_n^m(t) \sin(m\varphi)\right] \breve{P}_n^m(\sin(\varphi)) \tag{2.19}$$

where $a = 6371.2km$ is the standard Earth's magnetic reference radius and φ, λ, and r are the latitude, longitude and radius in a geocentric spherical reference frame. $\breve{P}_n^m(\sin(\varphi))$ is the *normalized Associated Legendre function*. The $g_n^m(t)$ and $h_n^m(t)$ are the time-dependent *Gauss coefficients* of degree n and order m that from degree 1 to 8 are assumed to have a quadratic dependence on time

$$g_n^m(t) = g_n^m + \dot{g}_n^m(t - t_0) + \frac{1}{2}\ddot{g}_n^m(t - t_0)^2 \tag{2.20}$$

$$h_n^m(t) = h_n^m + \dot{h}_n^m(t - t_0) + \frac{1}{2}\ddot{h}_n^m(t - t_0)^2 \tag{2.21}$$

Where g_n^m, h_n^m, $\dot{g}_n^m$, $\dot{h}_n^m$, $\ddot{g}_n^m$, and $\ddot{h}_n^m$ are constants and time is given in decimal year and t_0 is the reference date of the model. From degree 9 up to 12 a linear dependence on time of the Gauss coefficients is assumed and for higher degrees, the Gauss coefficients are assumed to be constant with time.

Spherical harmonics up to degree and order 10 have been fitted by the *International Association* of *Geomagnetism and Aeronomy* (IAGA) to the geomagnetic field observations

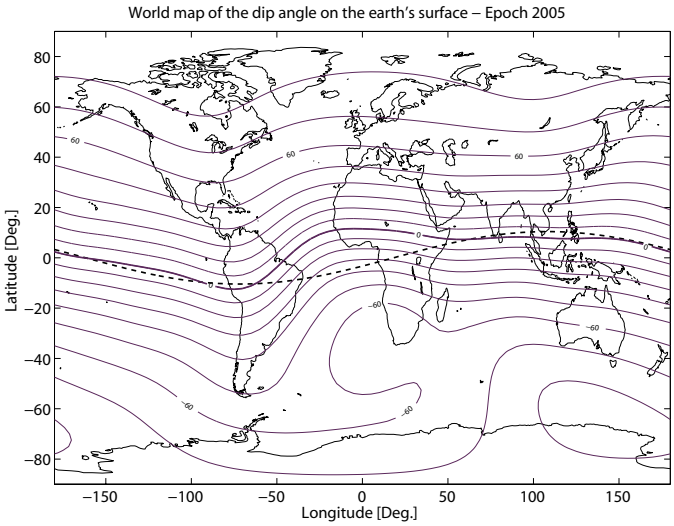

Figure 2.10: The magnetic inclination (dip angle) curves over the world, black dashed curve is geomagnetic equator and black solid curve is the dip equator

from satellite, land, and marine surveys every five years. These set of spherical harmonics is known as the *International Geomagnetic Reference Field* (IGRF).

Using the IGRF, the all seven geomagnetic elements of the geomagnetic field can be computed at a given location and time. The worldwide variation of the inclination, or dip angle, computed from IGRF-2005 is shown in figure 2.10. The locus of points where the geomagnetic inclination is zero is called the *dip equator*, which is a different from the *geomagnetic equator* introduced in the previous section. The difference between the geomagnetic and dip equators is quite large in the African continental region, but small in the South American region.

The position on the Earth's surface where the geomagnetic field line is vertical ($I = 90°$) are called the *dip poles*. The north and south dip poles are not *antipodal*. The dip poles computed from IGRF are called the *model dip poles*. The model dip poles are also not antipodal. Due to the pressure of the continuous solar wind on the Earth's magnetic field, the actual dip poles move considerable distances over one day, tracing out approximately oval-shaped loci on a daily basis, with large variation from one day to the next depending on solar activity.

The *geomagnetic dip latitude* or *dip latitude* (denoted by $Dipl$) is computed from equation 2.17 using the dip angle (Matsushita and Campbell, 1967), i.e.

$$Dipl = \arctan(\frac{1}{2}\tan(Dip)) \tag{2.22}$$

where Dip stands for the computed dip latitude from the IGRF.

The geomagnetic latitude and dip latitude of a given point on the Earth generally are not identical. In 1963, Rawer proposed a new geomagnetic parameter that is close to the magnetic dip angle at low latitudes, but becomes closer to the geodetic latitude ϕ as latitude increases (Bilitza, 1990). This new parameter is called the *modified dip latitude*

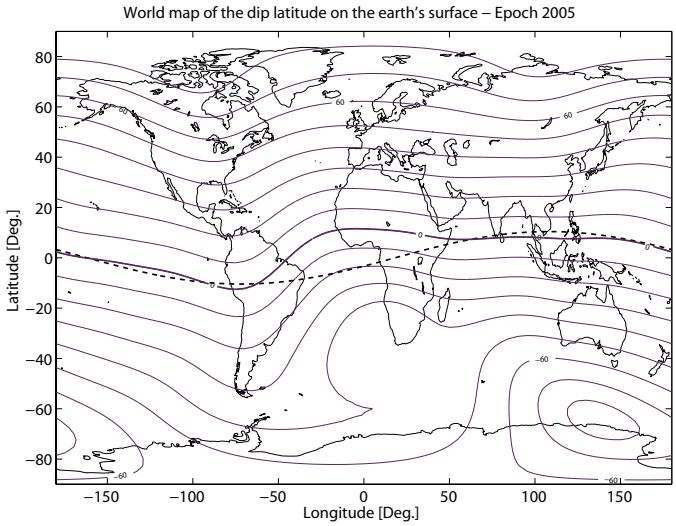

Figure 2.11: The magnetic dip latitude curves over the world, black dashed curve is geomagnetic equator and black solid curve is the dip equator

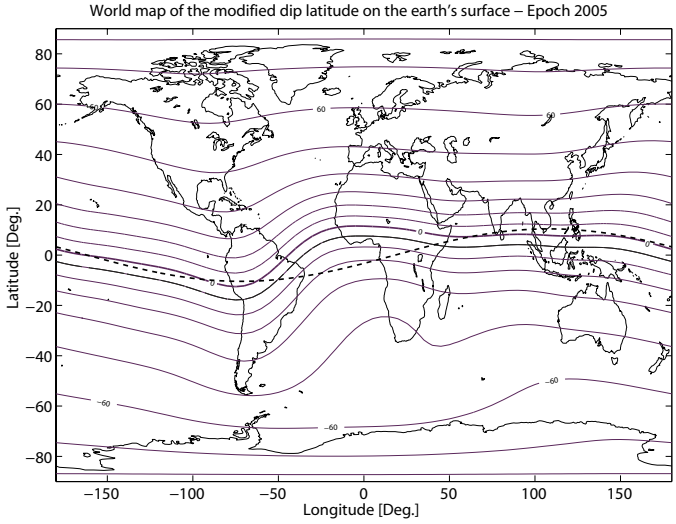

Figure 2.12: The modified dip latitude curves over the world, black dashed curve is geomagnetic equator and black solid curve is the dip equator

or *modified dip* (denoted by $Modip$ or μ) and is defined as follows

$$\tan(\mu) = \frac{Dip}{\sqrt{\cos(\phi)}} \tag{2.23}$$

The dip latitude and modified dip latitude are plotted in figures 2.11 and 2.12 respectively. The modified dip latitude is widely used for description of spatial dependency of the ionospheric parameters.

2.3.3 Geomagnetic storm

The geomagnetic field elements are varying in time mainly due to electric currents flowing above the earth's surface under influence of the continuous solar wind. A day is geomagnetically *quiet day* when the field elements at the earth's surface are changing gradually and smoothly due to the daily changing geometrical relationships between the Earth, Sun, and Moon. Sometimes, due to an suddenly increase in the solar particle radiation over the earth's polar upper atmosphere, additional currents circulate in the ionosphere and lead to *irregular* and rapid variations of the field elements and causing geomagnetic disturbances at the Earth's surface that known as *geomagnetic storm*. Geomagnetic storms may last from a few hours to several days. Days with the geomagnetic storms are called *geomagnetically disturbed days*. Geomagnetic storms especially occur in the northern Polar Regions (Hunsucker and Hargreaves, 2003), charged particles flow along the geomagnetic field lines and interact with the neutral atmosphere causing colored displays (the well-known *Aurora*). Around solar maximum, the intensity and frequency of geomagnetic storms increases.

2.3.4 Geomagnetic indices

A global picture of the geomagnetic activity is provided by different indices. The most important measure of the geomagnetic activity level, specifically in the field of ionospheric research, is called the *planetary Kp-index* which identifier solar particle radiation by its magnetic effects. The planetary Kp-index is derived using an average of the horizontal field intensity H (or D element if it is more disturbed than H) observations from a network of about 12 geomagnetic observatories distributed around the world between geomagnetic latitudes 48° and 63°. The planetary Kp-index is computed in three-hourly intervals. It is a global measure of the magnetic deviations from the regular daily variation during a 3-hour period (Schunk and Nagy, 2000). The Planetary Kp-index is provided through a semi-logarithmic numerical code that varies between 0 and 9, with the different numbers corresponding to different geomagnetic activity levels. Table 2.2 gives a classification of geomagnetic storms by means of the Kp-index in which $Kp = 5$ and $Kp = 9$ are corresponding to *minor* and *extreme* geomagnetic storms, respectively. Severe and extreme geomagnetic storms occur only rarely.

Table 2.2: Classification of the geomagnetic storms [NOAA, 2007]

Scale	Kp-index	# of storm days per solar cycle (11 years)
G5 (extreme)	9	4
G4 (severe)	8	60
G3 (strong)	7	130
G2 (moderate)	6	360
G1 (minor)	5	900

As an example, figure 2.13 shows the musical Kp-index diagram from October 27 until October 30, 2003. The horizontal axis is Universal Time (UT) and each bar is a 3-hour interval. The vertical axis shows the Kp index; the bars are green when the Kp≤ 3, yellow when Kp $= 4$, and red when the Kp> 4. The red bars coincide with an extreme geomagnetic storm that occurred at the end of October 2003.

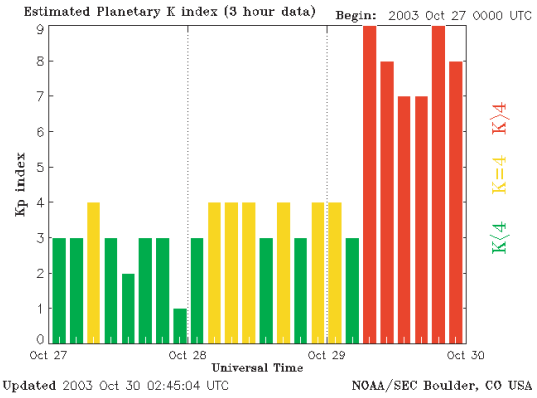

Figure 2.13: Kp index from October 27, till October 30, 2003 (Graph courtesy of NOAA Space Environment Center, USA)

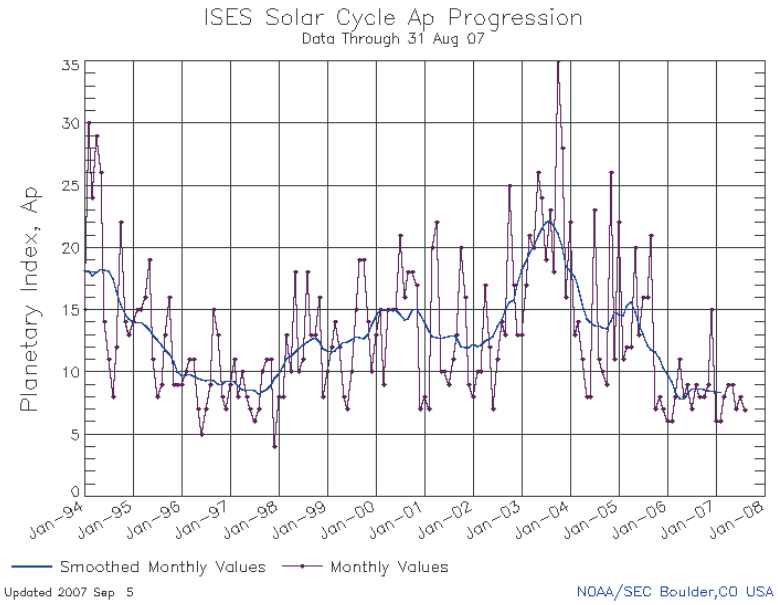

Figure 2.14: Progression of solar cycle 23: Measured and predicted Planetary Ap index (Graph courtesy of NOAA Space Environment Center (SEC), USA)

Since the Kp index is a semi-logarithmic scale it is not suitable for simple averaging to obtain a daily index. In order to obtain a daily measure of the geomagnetic activity, the Kp-index must first be converted into a roughly linear scale before it can be averaged over the day. This daily value is called the *planetary Ap-index*and used as a daily global measure of the geomagnetic activity for a given day. In figure 2.14 the monthly averaged planetary Ap-index for the current solar cycle (no. 23) is shown.

3 Physics of the Earth's Ionosphere

In this chapter the physics of the Earth's ionosphere is reviewed. Although many books, such as for example (Prolss, 2004; Schunk and Nagy, 2000; Dieminger et al., 1996; Hall et al., 1996; Tascione, 1994; Davies, 1989; Ratcliffe, 1972) give a detailed description of the ionosphere, a condensed description of the physical theory of ionosphere formation and its variations in time and space is needed. In this chapter first a brief explanation is given for interaction of solar radiation with the Earth's upper atmosphere that leads to production of electrically charged particles. Then the theory of the ionosphere formation based on the plasma continuity equation and transport process in the ionosphere are described. After that the formation of the stratified ionosphere is reviewed. Regular variations of the ionosphere, classical anomalies, ionospheric storm and the traveling ionospheric disturbances are also addressed.

3.1 Interaction of solar radiation with the Earth's upper atmosphere

Photons associated to solar UV, EUV, and X radiation collide with particles in the Earth's upper atmosphere resulting in the absorption of energy by the particles. The absorbed energy may lead to three important processes: *photodissociation*, *photoionization*, and the combination of these two, *dissociative photoionization* (Prolss, 2004). The absorption processes of interest are

- Photodissociation
  $O_2 \ + \ \{photon\} \ (\lambda \leq 242\,nm) \ \rightarrow \ O + O$
- Photoionization ($\lambda \leq 103\,nm$)
  $O \ + \ \{photon\} \ (\lambda \leq 91\,nm) \ \rightarrow \ O^+ + e$
  $N_2 \ + \ \{photon\} \ (\lambda \leq 80\,nm) \ \rightarrow \ N_2^+ + e$
  $O_2 \ + \ \{photon\} \ (\lambda \leq 103\,nm) \ \rightarrow \ O_2^+ + e$
- Dissociative Photoionization ($\lambda \leq 72\,nm$)
  $N_2 \ + \ \{photon\} \ (\lambda \leq 49\,nm) \ \rightarrow \ N^+ + N + e$

Absorption of solar radiation has a significant influence on the properties of the upper atmosphere. For instance, the photodissociation process produces atomic oxygen which not only changes the chemical composition of the upper atmosphere but also its absorption

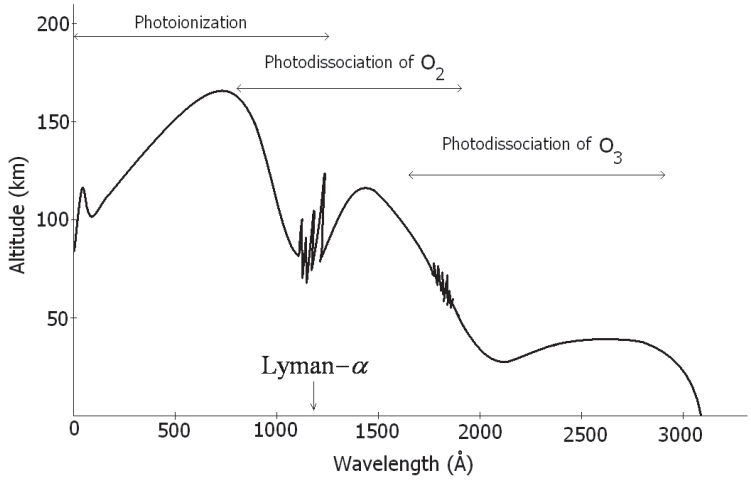

Figure 3.1: A typical representation of the depth of penetration of vertically incident solar radiation into the atmosphere

characteristics. Photoionization causes the upper atmosphere to transform into a conducting medium. During absorption processes heat is generated, which raises the temperature of the upper atmosphere significantly.

The different gas particles are dissociated or ionized only by a certain minimum of amount of energy from incident photons. Therefore, dissociation and ionization processes of different gases are performed by different wavelengths. This implies that the depth of penetration of incident solar radiation into Earth's atmosphere depends on wavelength. Figure 3.1 shows the depth of vertically incident solar radiation penetration into the atmosphere as a function of wavelength. For $\lambda > 3100\dot{A}$ the radiation reaches the ground, but for $\lambda \in [2000\dot{A}, 3000\dot{A}]$ all radiation is absorbed by photodissociation of *ozone* at about 40 km, for $\lambda \in [1000\dot{A}, 1700\dot{A}]$ is absorbed by photodissociation of O_2 around 100 km altitude, and photoionization of the upper atmosphere above 90km is caused by the portion of the spectrum with $\lambda < 1400\dot{A}$ approximately.

Absorption of radiation in a gas Solar radiation is attenuated while propagating through atmosphere. The attenuation over a path length ds in a gas is calculated based on the probability of collision between photons and particles. This probability can be visualized from figure 3.2. The sum of the gas particle absorption cross section projected onto the area A is a measure of the probability for a collision over the path length ds. The area remaining on A after subtracting this sum is a measure of the probability that a photon traverses this length without a collision. If it is assumed that there is no overlapping between the projection areas the collision probability $d\omega$ is then

$$d\omega = \frac{\sum \sigma}{A} = \frac{\sigma\, n(s) A\, ds}{A} = \sigma\, n(s)\, ds \tag{3.1}$$

where n is number of particles in volume Ads and σ the *absorption cross section*. The absorption cross section is the ability of a particle to absorb a photon of a particular wavelength.

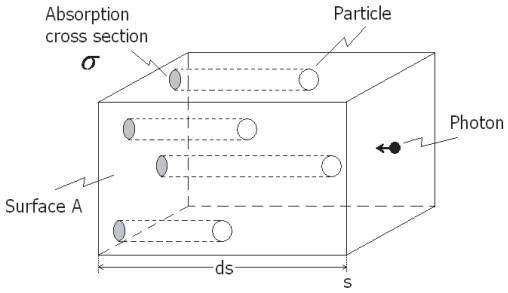

Figure 3.2: Probability of an absorption for a photon traversing the path in a gas, adapted from (Prolss, 2004)

Using the above collision probability, the attenuation of a photon flux in a gas volume can be calculated

$$d\phi^{photon} = -\phi^{photon}(s)\, d\omega = -\phi^{photon}(s)\, \sigma\, n(s)\, ds \tag{3.2}$$

where $\phi^{photon}(s)$ is photon flux passing at the entrance surface of the gas volume. The change in photon flux is proportional to the initial photon flux, the size of the absorption cross section, the density of the absorbing gas particles and length of the propagation path. Integration of equation 3.2 over path length leads to the *Lambert-Beer* law

$$\phi^{photon}(s) = \phi^{photon}(s_0)\, e^{-\tau(s)} \tag{3.3}$$

with

$$\tau(s) = \int_{s_0}^{s} \sigma\, n(s')\, ds' \tag{3.4}$$

the *optical depth* of the gas volume. A large value of τ corresponds to a strongly absorbing and optically thick gas volume, a small value of τ indicates a weakly absorbing and optically thin gas volume.

The rate of ion production, in the Earth's ionosphere, depends on the absorption of solar radiation in the Earth's upper atmosphere. This topic will be discussed in the next chapter.

3.2 Ionosphere formation theory

The ionosphere is a region of the Earth's atmosphere where significant numbers of free electrons and ions are present. The free electrons and ions are produced via ionization of the neutral particles by solar radiation and by collisions with energetic particles (solar wind) that penetrate the upper atmosphere. The ionization produced by solar wind is usually small compared with that produced by photons, and this source of ionization only plays a role at polar latitudes. In this chapter, we will only consider solar ionizing radiation (solar extreme ultraviolet and X-ray radiations) in the formation of ionospheric plasma in the Earth's atmosphere by *photoionization* of the neutral atmospheric gases.

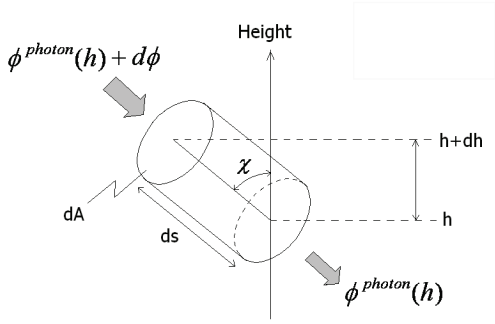

Figure 3.3: Absorption of incident radiation in the vertical direction

Positive ions and electrons in the upper atmosphere are usually well mixed and form a ionospheric plasma. A *plasma* contains equal amounts of positive and negative charges in a volume, such that each volume is electrically neutral (*charge neutrality*). A plasma is the fourth state of matter. In many respects it behaves like a gas, but when electric and magnetic forces are present, it has specific properties, quite unlike those of ordinary gases (Dendy, 1993).

3.2.1 Plasma continuity equation

The temporal variation of the ionized matter in a unit volume is described by the so-called *density balance* or *continuity equation*, which consists of three principal photochemical processes: *production*, *disappearance* and *transport* of ions and electrons. Under conditions of charge neutrality of the ionospheric plasma, the continuity equation may be written as follows (Ivanov-Kholodny and Mikhailov, 1986)

$$\frac{\partial n_i}{\partial t} = q_i - L_i - div(n_i V_i) \tag{3.5}$$

In which q_i and L_i are respectively rates of production and disappearance of ions (per unit volume and time), n_i is the ion concentration, and V_i is the bulk transport velocity of the i^{th} ionized species. $n_e = \sum n_i$ is the electron density. $div(n_i V_i)$ represents the change of the ion concentration by transport process.

3.2.2 Ion production

The rate of ion production depends completely on absorption of solar radiation in the upper atmosphere. In order to determine absorption of solar radiation in the upper atmosphere, an *idealized atmosphere* with a single gas and horizontal stratification is considered. The photon flux at hieght h for a parallel beam of *monochromatic* ionizing radiation entering the idealized atmosphere at a zenith angle χ (see figure 3.3), can be obtained from the Lambert-Beer law (see equation 3.3, exchanging the upper and lower integration limits $s = h$ and $s_0 = \infty$)

$$\phi^{photon}(h) = \phi_{\infty}^{photon}\, e^{-\tau(h)} \tag{3.6}$$

Here, ϕ_{∞}^{photon} denotes the solar photon flux at the top of the atmosphere. The optical depth takes the following form after equation 3.4

$$\tau(h) = \sec(\chi) \int_{h}^{\infty} \sigma \, n(h') \, dh' \tag{3.7}$$

with $n(h')$ the gas particle density and σ the absorption cross section.

Taking the hydrostatic equation (equation 2.6) and the ideal gas law into account, the optical depth is approximately

$$\tau(h) \approx \sec(\chi) \, \frac{\sigma \, p(h)}{m \, g(h)} = \sec(\chi) \, \sigma \, n(h) \, H(h) \tag{3.8}$$

where $H(h)$ is the scale height. To obtain the rate of ion production in a given volume, first, the number of photons absorbed in the volume should be calculated. For this purpose, consider a cylinder with base surface area dA and length ds with central axis parallel to the direction of incident radiation (see figure 3.3). The number of photons entering the cylinder per time interval dt is $(\phi^{photon}(h) + d\phi) \, dA \, dt$, and the number of photons leaving at bottom side per time interval dt is $\phi^{photon}(h) \, dA \, dt$. According to equation 3.2, the number of absorbed photons in the unit volume in the height of h over time interval dt is

$$\left.\frac{d\phi}{ds}\right|_{h} = \sigma \, n(h) \, \phi^{photon}(h) \tag{3.9}$$

Substituting equation 3.6 into 3.9 yields

$$\left.\frac{d\phi}{ds}\right|_{h} = \sigma \, n(h) \, \phi_{\infty}^{photon} \, e^{-\tau(h)} \tag{3.10}$$

From figure 3.3 follows that ds (measured towards the earth) is related to dh (measured upwards) by $ds = \sec(\chi) \, dh$, so that

$$\left.\frac{d\phi}{dh}\right|_{h} = \sigma \, n(h) \, \phi_{\infty}^{photon} \, e^{-\tau(h)} \, \sec(\chi) \tag{3.11}$$

Chapman production function Chapman was the first to clearly describe a theory of formation of ionized regions in an *idealized atmosphere*. His equation is now known as Chapman's formula or Chapman production function for the rate of ion production with respect to height (Tascione, 1994). To obtain Chapman's formula, let η be the *ionization efficiency*, which gives the number of ion pairs (ion and electron) produced per single absorbed photon. For an atomic gas and for wavelengths less than the minimum wavelength of the ionizing radiation of the atom, the ionization efficiency is $\eta \approx 1$. The rate of ion production is the number of ion pairs produced per unit volume per second at height of h. It is the product of ionization efficiency and number of absorbed photons, given by equation 3.10,

$$q(\chi, h) = \eta \left.\frac{d\phi}{ds}\right|_{h} = \eta \, \sigma \, n(h) \, \phi_{\infty}^{photon} \, e^{-\tau(h)} \tag{3.12}$$

The product of the ionization efficiency and the absorption cross section is known as the *ionization cross-section*. By substitution $n(h)$ from equation 3.8 into equation 3.12 yields

$$q(\chi, h) = \eta \frac{\phi_{\infty}^{photon}}{H(h)\ \sec(\chi)}\ \tau(h)\ e^{-\tau(h)} \tag{3.13}$$

The maximum rate of ion production is obtained from dq/dh or $d(\ln(q))/dh$ which leads to following equation

$$\frac{d\tau}{dh}\left(\frac{1}{\tau} - 1\right) = 0 \tag{3.14}$$

Here Chapman assumed that the idealized atmosphere is an isothermal atmosphere. Neglecting the variation of the earth's gravity with height, the height scale is independent of height in the isothermal atmosphere. So from now on we write H instead of $H(h)$. In an isothermal atmosphere the barometric equations (see equations 2.9 and 2.10) for pressure and density are, respectively,

$$p(h) = p(h_0) \exp\left(-\frac{h - h_0}{H}\right) \tag{3.15}$$

$$n(h) = n(h_0) \exp\left(-\frac{h - h_0}{H}\right) \tag{3.16}$$

with h_0 is an arbitrary reference height. Equations 3.8 and 3.16 can be used to calculate the vertical gradient of the optical depth

$$\frac{d\tau}{dh} = \sec(\chi)\ \sigma\ n(h_0)\ \exp\left(-\frac{h - h_0}{H}\right) \tag{3.17}$$

where $d\tau/dh$ vanishes only for $h \to \infty$. It follows that $\tau(h_{max}) = 1$. The maximal rate of ion production is

$$q_{\max} = q(\chi, h_{\max}) = \eta \frac{\phi_{\infty}^{photon}}{H\ e\ \sec(\chi)} \tag{3.18}$$

with $h_{\max}$ the height at which the maximum occurs. Equation 3.8 at h_{max} becomes

$$\tau(h_{\max}) = \sec(\chi)\ \sigma\ n(h_0)\ H\ \exp\left\{-\frac{h_{\max} - h_0}{H}\right\} = 1 \tag{3.19}$$

where equation 3.19 leads to

$$\exp\left\{\frac{h_{\max} - h_0}{H}\right\} = \sec(\chi)\, \sigma\, n(h_0)\, H \tag{3.20}$$

Substituting equation 3.16 into equation 3.7 and taking equation 3.20 into account results in

$$\tau(h) = -\exp\left\{-\frac{h - h_{\max}}{H}\right\} \tag{3.21}$$

It is convenient to transform equation 3.13 by substituting for $\tau(h)$ from equation 3.21 to give

$$q(\chi,h) = \underbrace{\frac{\eta\phi_{\infty}^{photon}}{H\, e\, \sec(\chi)}}_{q_{\max}} \exp\left\{1 - \frac{h-h_{\max}}{H} - \exp\left\{-\frac{h-h_{\max}}{H}\right\}\right\} \tag{3.22}$$

where e is the base of the natural logarithm. Using equation 3.19, the height of maximum rate of ion production is calculated

$$h_{\max} = h_0 + H\, \ln\left(\sigma\, H\, n(h_0)\, \sec(\chi)\right) \tag{3.23}$$

Equation 3.22 is known as the *Chapman production function*. It gives the vertical profile of the rate of ion production in an idealized isothermal atmosphere with *planar* stratification in terms of the maximum rate of ion production q_{max} and the height of maximum rate of ion production h_{max}. It can be seen from equation 3.18 and 3.23 that both the maximum rate of ion production and the corresponding height h_{max} are dependent of the solar zenith angle. Furthermore, q_{max} depends on solar radiation intensity at top of the atmosphere ϕ_{∞}^{photon} and the ionization efficiency η while h_{max} depends on the absorption cross-section σ, but not on ϕ_{∞}^{photon} and on η. It is usually convenient to express the rate of ion production in terms of q_{max}^0 and h_{max}^0, the corresponding quantities for q_{max} and h_{max} when the sun is overhead $\chi = 0$. For this purpose, from equation 3.18, q_{max}^0 is

$$q_{\max}^0 = q_{\max}\ \sec(\chi) \tag{3.24}$$

The relation between h_{max}^0 and h_{max}, from equation 3.20, is

$$\exp\left\{\frac{h_{\max}-h_0}{H}\right\} = \sec(\chi)\ \exp\left\{\frac{h_{\max}^0-h_0}{H}\right\} \tag{3.25}$$

Substituting q_{max} and h_{max} from equations 3.24 and 3.25 into equation 3.22 gives

$$q(\chi,h) = \underbrace{\frac{\eta\ \phi_{\infty}^{photon}}{H\ e}}_{q_{\max}^0}\ \exp\left\{1 - \frac{h-h_{\max}^0}{H} - \sec(\chi)\ \exp\left\{-\frac{h-h_{\max}^0}{H}\right\}\right\} \tag{3.26}$$

Or with $z = (h - h_{max}^0)/H$ as a new parameter to measure heights, in terms of H as a scale unit, with height h_{max}^0,

$$q(\chi,z) = q_{\max}^0\ \exp\{1 - z - \sec(\chi)\ \exp\{-z\}\} \tag{3.27}$$

Figure 3.4 illustrates the height dependence of the Chapman production function for three different solar zenith angles. The maximum rate of ion production becomes smaller and shifts higher altitudes when the solar zenith angle is increased.

Since at large solar zenith angles the upper atmosphere can not be considered as planar (see figure 3.5), the Chapman production function must be modified by considering a *spherical* stratification in the idealized atmosphere. This is much more important in the upper atmosphere than in the lower atmosphere. Much more effort is required to account properly for the spherical shell form of the atmosphere. This can be carried out by introducing the

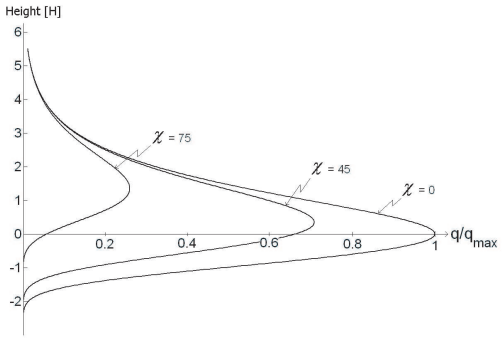

Figure 3.4: Chapman production function for three different solar zenith angles. The vertical axis shows normalized height $z = (h - h^0_{max})/H$ in terms of scale height as a unit.

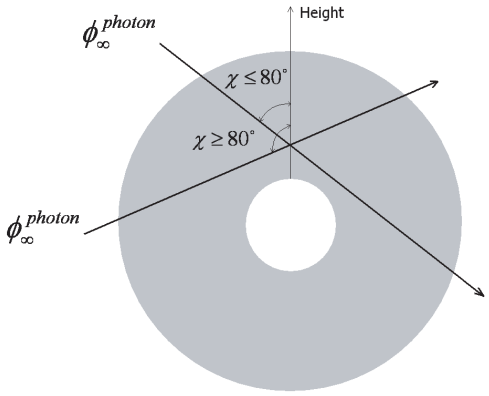

Figure 3.5: Modification of the absorption path at large zenith angle using the spherical shell for the upper atmosphere

so-called *Chapman grazing incidence function* $Ch(\chi, h)$ that its detailed derivation is given in (Chapman, 1931b). The Chapman grazing incidence function is a measure of the depth of atmosphere (or absorption of ionizing radiation by atmosphere) as a function of the solar zenith angle and height (Chapman, 1931c)

$$Ch(\chi, h) = \sec(\chi) - \int_0^{\chi} \exp\{r(1 - \sin(\chi)\cos ec(\theta)\}\sec(\theta)\tan(\theta)d\theta \tag{3.28}$$

with $r = (R_e + h)/H$ where R_e is the radius of the Earth. Note that $Ch(\chi, h)$ is a dimensionless function. It is also understandable that $Ch(\chi, h)$ must depend on the height.

It is sufficient to know that for $\chi \leq 80^\circ$ the effect of curvature of the Earth can be ignored and the Chapman grazing function is approximately equal to $\mathsf{sec}(\chi)$ (Davies, 1989). However, for large zenith angles, the effect of curvature of the Earth is important and around sunrise and sunset the values of $Ch(\chi, h)$ are very different from $\mathsf{sec}(\chi)$ (see figure 3.6).

Taking the spherical stratification into account in the idealized atmosphere, the optical depth in equation 3.7 is modified by replacing $\mathsf{sec}(\chi)$ by the Chapman grazing incidence function. The equation 3.26 becomes

$$q(\chi, h) = q^0_{\max}\ \exp\left\{1 - \frac{h - h^0_{\max}}{H} - Ch(\chi, h)\ \exp\left\{-\frac{h - h^0_{\max}}{H}\right\}\right\} \tag{3.29}$$

Figure 3.7 shows the Chapman production function using the Chapman grazing incidence function $Ch(\chi, h)$ and using secant function $\mathsf{sec}(\chi)$ for solar zenith angle $\chi = 88^\circ$. It is clear that production rate and the height of maximum rate of ion production are changed.

In order to determine the vertical profile of the rate of ion production in the real upper atmosphere, the Chapman production function must be modified. First, the upper atmosphere consists of several gases rather than a single gas. Therefore, $\sigma n(h)$ in equations 3.8 and 3.12 must be replaced by a sum of the products accounting for the contributions from the various gas particles. Furthermore, because the incident solar radiation is *polychromatic* rather than monochromatic, it is necessary to add up the different contributions

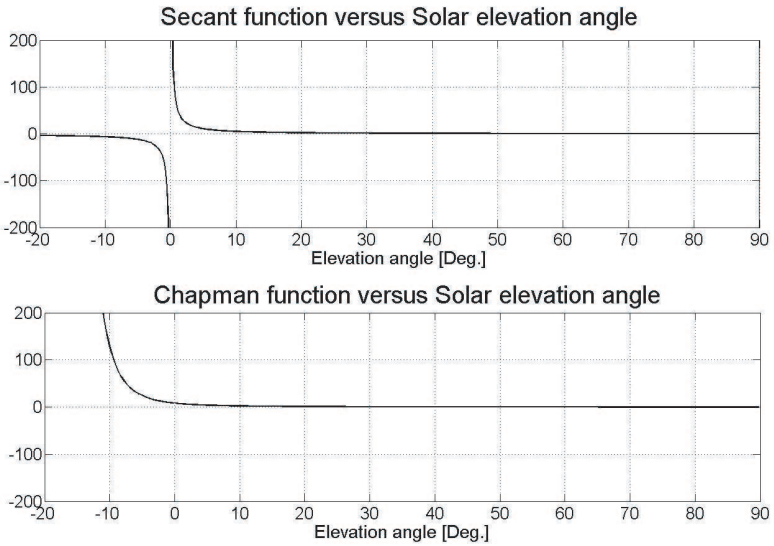

Figure 3.6: Illustration of the Chapman grazing incidence function versus solar elevation angle (H=50 km, h=350 km)

from the individual wavelengths. Accordingly, equation 3.12 becomes the following form for real upper atmosphere

$$q(\chi,h) \approx \sum_{\lambda} \sum_{i=O,N_2,O_2} \eta_i\,\sigma_i\,n_i(h)\,\phi_{\infty}^{photon}(\lambda)\,\exp\left\{-Ch(\chi,h)\sum_{j=O,N_2,O_2}\eta_j\,\sigma_j\,n_j(h)\,H_j\right\} \tag{3.30}$$

where $\sum_{\lambda}$ indicates a summation over the different wavelengths of interest and $\sum_{i=O,N_2,O_2}$ and $\sum_{j=O,N_2,O_2}$ are summations over the gas particles of the upper atmosphere (only the most important gases are considered). Note that the upper atmosphere is surely not isothermal, but considering non-isothermal condition is too complicated to be considered in the analytical calculations.

3.2.3 Ion and electron disappearance

Each ion in the ionosphere has a finite lifetime: either it is destroyed by ionic chemical reactions or after a certain time it is recombined with an electron, returning to the neutral state. Once electrons are produced by solar radiation in the upper atmosphere, there are three principal reactions in which electrons may disappear after some time:

- Radiative recombination (an electron combines with an atomic positive ion)
  $O^+ + e \rightarrow O + \{photon\}$
- Dissociative recombination (an electron combines with a molecular ion)
  $O_2^+ + e \rightarrow O + O$
  $N_2^+ + e \rightarrow N + N$

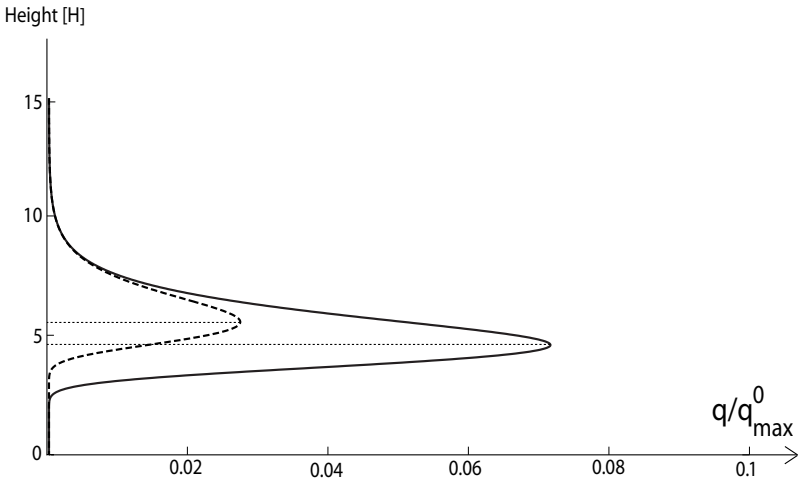

Figure 3.7: Chapman production function using the Chapman grazing function (solid) and using secant function (dashed) for solar zenith angle $\chi = 88°$. The vertical axis shows normalized height $z = (h - h^0_{\max})/H$ in terms of scale height as a unit.

$NO^+ + e \rightarrow N + O$
Note that neutral nitrogen oxide is only a minor atmospheric gas and is mostly produced as the final product of the charge exchange process due to ion-molecular reactions between primary ions (O^+, N_2^+ and O_2^+).

- Attachment (an electron attach to a neutral particles and produce negative ion)
  $O_2 + e \rightarrow O_2^-$

Since the radiative recombination is a much slower process than dissociative recombination the greater part of electrons tend to recombine with positive molecular ions (Gran, 1965). The attachment process only occurs in lower altitudes of the ionosphere, where more neutral particles are available, to form negative ions (D-region). The negative ions are in turn rapidly neutralized by further reactions, so that loss of electrons by the attachment can be neglected.

If only the dissociative recombination is considered the recombination rate of ions and electrons can be computed as follows. Let the number density of electrons and positive ions be denoted by $n^-(h)$ and $n^+(h)$, respectively. If we assume that there is no negative ion then $n(h) = n^-(h) = n^+(h)$. The rate of electron loss is considered to be proportional to the number of possible collision between electrons and ions $n^2(h) = n^-(h)n^+(h)$ in a unit volume of the ionosphere in a unit time and it is given by

$$L(h) = \frac{dn(h)}{dt} = -\alpha(h)\, n^2(h) \tag{3.31}$$

where $\alpha(h)$ is known as the quadratic recombination coefficient which varies only weakly with height via its temperature dependence (Prolss, 2004). Negative sign denotes that the electron density is decreasing by time. This quadratic form of the recombination rate is

only valid for lower part of the ionosphere (E-region). In the higher altitudes (F2 region) where atomic oxygen ion dominates, the radiative recombination between atomic oxygen ion and an electron is not the main process. Instead two charge exchange reactions first produce NO^+ and O_2^+ from O^+

$$O^+ + N_2 \rightarrow NO^+ + N \tag{3.32}$$

$$O^+ + O_2 \rightarrow O_2^+ + O \tag{3.33}$$

which are then recombined with an electron by the rapid process of dissociative recombination. This implies that the loss rate is proportional to the number of possible collision between O^+ and the neutral molecules (Oxygen and Nitrogen molecules). If $n^0(h)$ denotes the number density of neutral molecules then by assuming charge neutrality (number of O^+ is equal to number of electrons $n(h)$)

$$L(h) = \frac{dn(h)}{dt} = -\underbrace{k\, n^0(h)}_{\beta(h)}\, n(h) \tag{3.34}$$

where k is coefficient of proportionality and $\beta(h)$ the linear recombination coefficient. The loss rate depends now linearly on the electron density $n(h)$. The linear recombination coefficient $\beta(h)$ in equation 3.34 has the same height-dependence as the neutral molecules (Ratcliffe, 1972). The height at which the recombination coefficient changes from quadratic (α) to linear $\beta(h)$ is called the *transition level* h_t.

3.2.4 Chapman layer

Under an idealized atmosphere, a simple model for vertical electron density profile can be obtained by substitution of the Chapman production function (equation 3.29) into the plasma continuity equation 3.5. If transport term is neglected and the dissociative recombination rate is assumed to be proportional to the square of the electron density, the time variation in the electron density can be written as:

$$\frac{\partial n_e}{\partial t} = \frac{\eta \phi_\infty^{photon}}{eH} \exp\left\{1 - z - Ch(\chi)\exp(-z)\right\} - \alpha\, n_e^2 \tag{3.35}$$

$$z = \frac{h - h^0_{\max}}{H} \tag{3.36}$$

Solution of equation 3.35 under *photochemical equilibrium* (rate of ionization equal to recombination rate, $q = L$) $\frac{\partial n_e}{\partial t} = 0$ results in an idealized vertical profile of the electron density called the *Chapman layer* or *Chapman model*

$$n_e(z, \chi) = n^0_{\max} \exp\left\{\frac{1}{2}\left[1 - z - Ch(\chi)\exp(-z)\right]\right\} \tag{3.37}$$

In which n^0_{max} is the peak electron density with the Sun overhead ($\chi = 0$)

$$n^0_{\max} = \sqrt{\frac{\eta\ \phi_\infty^{photon}}{eH\alpha}} \tag{3.38}$$

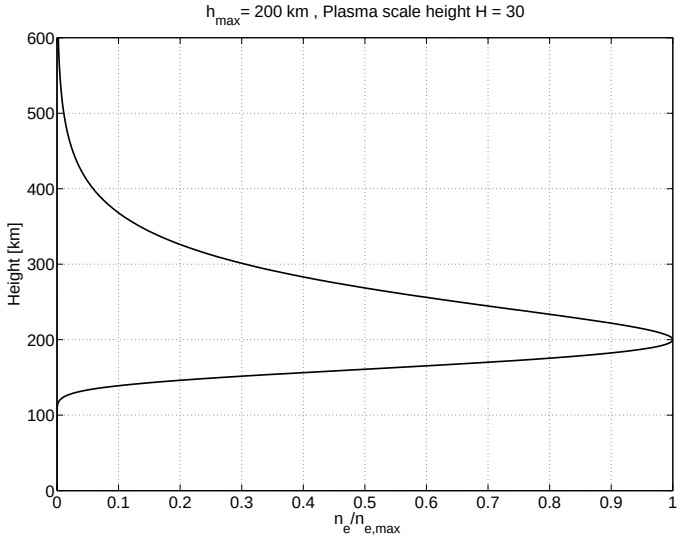

Figure 3.8: Vertical electron density profile in Chapman layer

The peak electron density for a given solar zenith angle is computed from equation 3.24

$$n_{\max} = \sqrt{\frac{\eta\ \phi_{\infty}^{photon} \cos(\chi)}{eH\alpha}} \tag{3.39}$$

which is independent of the ionization cross-section. The height of the peak electron density is obtained by equation 3.23 which it is independent of the solar radiation intensity, but depends on the absorption cross-section. Considering $\mathsf{sec}(\chi)$ as an approximation the Chapman grazing incidence function, the Chapman model can be re-written as follows with the reference height is replaced by the height of the peak electron density

$$n_e(z,\chi) = n_{\max} \exp\left\{\frac{1}{2}\left[1 - z - \exp(-z)\right]\right\} \tag{3.40}$$

$$z = \frac{h - h_{\max}}{H} \tag{3.41}$$

In practice, any ionized medium in which the recombination rate is proportional to square of the electron density can be reasonably modeled by the Chapman model (Tascione, 1994). As an example, a typical form of the electron density profile is illustrated in figure 3.8. The electron density is almost parabolic around the height of peak electron density.

3.3 Transport process in the ionosphere

The Earth's external magnetic field plays a key role in the charged particles transport in the ionosphere. This is because of the fact that the earth's magnetosphere and ionosphere are closely linked together via magnetic field lines and strong interactions between the field lines and the charged particles result in the capture of the particles and impose restrictions for their motion in the ionosphere. In the ionosphere, charged particles transport are mainly

related to the process of *diffusion* in a gravity field, and to *thermospheric* winds and *electric* fields (Davies, 1989).

3.3.1 Charged particle motion in a magnetic field

The motion of a charged particle in magnetic and electric fields is described by the *Lorentz* equation (Walt, 1994)

$$\vec{F} = m\frac{d\vec{v}}{dt} = q\left(\vec{v} \times \vec{B} + \vec{E}\right) \tag{3.42}$$

where $\vec{F}$ is the force in Newton, q is the charge in Coulomb, $\vec{E}$ is the electric field in $Volt/m$, $\vec{B}$ is the magnetic field in $[Tesla]$ and $\vec{v}$ is the velocity in m/s. Equation 3.42 can be separated into components parallel and perpendicular to the magnetic field giving

$$\vec{F}^{\parallel} = m\frac{d\vec{v}^{\parallel}}{dt} = q\vec{E}^{\parallel} \tag{3.43}$$

$$\vec{F}^{\perp} = m\frac{d\vec{v}^{\perp}}{dt} = q\left(\vec{v}^{\perp} \times \vec{B} + \vec{E}^{\perp}\right) \tag{3.44}$$

Equation 3.43 is the usual equation for describing the motion of a charged particle in an electric field. In equation 3.44, the magnitude and direction of the magnetic force is a function of the velocity, therefore the solution of equation 3.44 is complicated when the magnetic field is not uniform. For simplicity, we consider only a *uniform* magnetic field and the solution of equation 3.44 is described only for following two cases:

- In the absence of an external electric field $\vec{E} = 0$, the velocity $\vec{v}^{\parallel}$ along the magnetic filed lines is constant according to equation 3.43, the acceleration $\frac{d\vec{v}^{\perp}}{dt}$ in equation 3.44 must be perpendicular to $\vec{v}^{\perp}$, so that also $\left\|\vec{v}^{\perp}\right\| =$ constant. Therefore, the trajectory of the charged particle in a uniform magnetic field with no electric field is a *helix* (see figure 3.9). The projection of the helix on to a plane perpendicular to magnetic filed line is a circle with radius is $r = mv^{\perp}/qB$, which is the so called *Larmor radius*. The angular frequency of this gyration motion is $\omega = v^{\perp}/r$ in $\frac{radian}{s}$, which is called the *angular gyro-frequency*. The direction of the magnetic force dependence on the charge, therefore direction of circular motion for positively charged particles is opposite to that for negatively charged particles.

- The presence of an external electric field, $\vec{E} \neq 0$, results in a drift motion of the charged particles in the geomagnetic field. If $\vec{E}^{\parallel} \neq 0$ and constant, which the particle is accelerated along the magnetic field line. The force $q\vec{E}^{\perp}$ which is perpendicular to $\vec{B}$, results into a drift of charged particles in a direction is perpendicular to both $\vec{B}$ and $\vec{E}$. The drift velocity is obtained by

$$\vec{v}_{Drift} = \frac{\vec{E}^{\perp} \times \vec{B}}{\left\|\vec{B}\right\|^2} \tag{3.45}$$

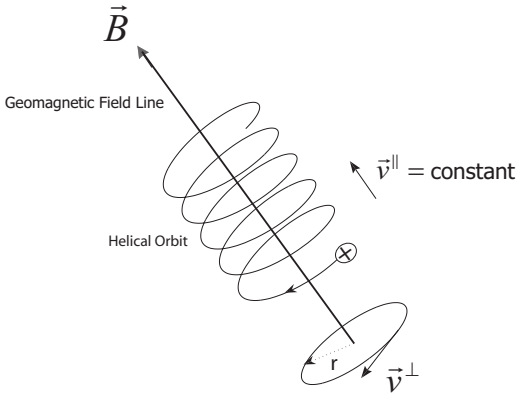

Figure 3.9: Charged particle motion in a uniform magnetic field ($\vec{E} = 0$)

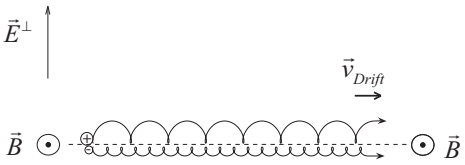

Figure 3.10: Drift motion in a uniform magnetic field ($\vec{E}^{\perp} \neq 0$)

The drift velocity is charge independent. Therefore, both ions and electrons move together (*ambipolar*) with the same velocity in the same direction perpendicular to both magnetic and electric fields (see figure 3.10).

3.3.2 Plasma diffusion

The Earth's gravitational force causes a downward motion of the ions and electrons in the ionospheric plasma. In the other direction the vertical pressure gradient force induces an upward motion of the ions and electrons. Since charged particles are only free to move parallel along the geomagnetic field, the forces must be mapped along the geomagnetic field lines in order to compute the total flux or *diffusion speed* of the ions and electrons (see figure 3.11). The expression for the diffusion speed of ions and electrons along geomagnetic filed line, under the ambipolar approximation that the ions and electrons move together with the same velocity, is given by (Davies, 1989)

$$v = -D(h)\sin I(\frac{1}{n_e}\frac{\partial n_e}{\partial h} + \frac{Mg}{2kT}) \tag{3.46}$$

where M and T are respectively the ion mass and temperature, $D(h)$ is plasma (or ambipolar) diffusion coefficient, n_e is electron density, I is the geomagnetic inclination and k is *Boltzmann's Constant*. Since $D(h)$ is inversely proportional to the neutral density, diffusion increases rapidly at higher altitude. In equation 3.46, the first term in the bracket relates to the pressure gradient force and second term is due to the gravitational force.

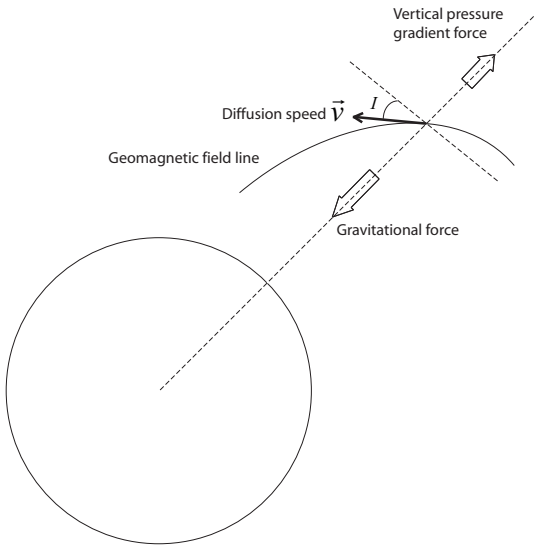

Figure 3.11: Plasma diffusion along geomagnetic field lines

Since diffusion across the magnetic field is inhibited, the vertical component of diffusion vanishes at the geomagnetic equator ($I = 0°$). Vertical diffusion has a maximum at the geomagnetic poles where $I = 90°$.

3.3.3 Thermospheric wind

A considerable temperature differences exist in the upper ionosphere between day and night side of the Earth. This creates a pressure difference that drives a horizontal wind in the upper atmosphere from the dayside toward the nightside, called the *neutral wind* or *thermospheric wind*. Since the horizontal neutral wind can not move the ions and electrons across the earth's magnetic field lines, they stick to the magnetic field lines like rings on a bar; they can slide along the field lines easily but can not move across it. If the horizontal wind is toward the equator, which occurs mainly at night in mid-latitude, the ionospheric plasma is lifted and tends to increase the ion density at higher altitudes of the ionosphere. If the wind is toward the pole, which occurs mainly by day at mid-latitude, the plasma is lowered and tends to reduce the ion density in the higher altitudes of the ionosphere (see figure 3.12). These effects depend on the geometry of the magnetic field and vary with latitude and the geomagnetic inclination. The plasma velocity along geomagnetic filed is given by

$$w = V \cos I \tag{3.47}$$

where V is speed of the neutral wind which follows from the equation of motion by taking the horizontal pressure gradient, the inertia and viscosity of the air, gravity, the Coriolis and centripetal accelerations due to the Earth's rotation, and ion-drag into account (Rishbeth, 1972). Note that at the geomagnetic poles ($I = 90°$), the thermospheric plasma velocity is non existent.

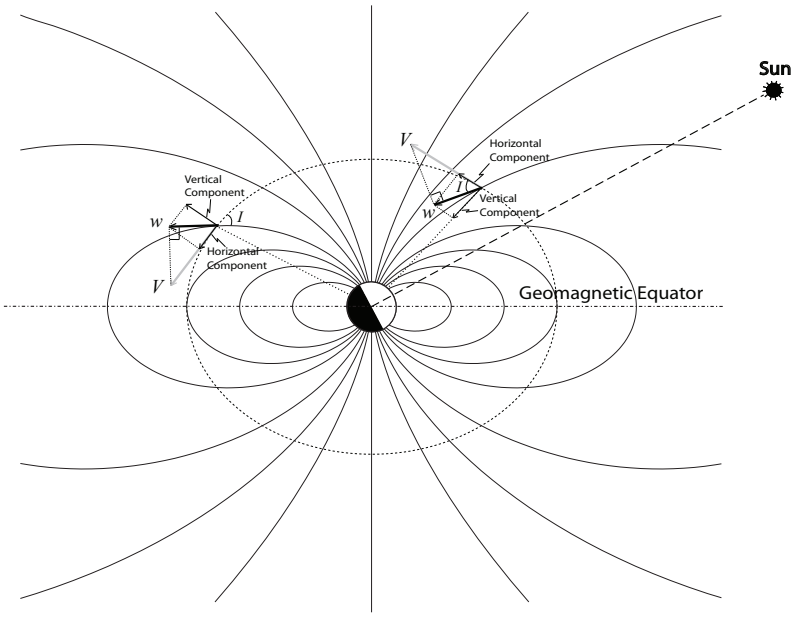

Figure 3.12: Illustration of the vertical plasma transport caused by the neutral wind in the mid-latitude region

3.3.4 Electromagnetic drift

The conductivity of the ionosphere varies with height and reaches a maximum about 100 km (E-region). This is known as the *central dynamo layer* or the *conducting plasma layer*. This is due to an increase of the collision frequency of the charged particles with respect to their angular gyro-frequencies. The conducting plasma layer is moved across the earth's magnetic field under the influence of mainly the sun and the moon. The solar EUV radiation, is not only responsible for the ionization at low-altitudes of the ionosphere, but also heats up the atmosphere and induces the thermospheric wind and resulting horizontal atmospheric movements with period of 12 hours. The gravitational attraction of the moon results in an atmospheric tide with a period of half a lunar day. The motion of the conducting plasma layer across the geomagnetic field generates an electric field ($\vec{E}$) and as a result a system of electric currents is produced in the layer. The pattern of the currents is fixed to the Sun, with the Earth rotating under it. The electric currents produced by the solar heating and lunar atmospheric tide are called S_q and L electric currents, respectively (Matsushita and Campbell, 1967). The idealized, current system is shown in figure 3.13.

At the dip equator, the geomagnetic field lines are horizontal ($I = 0^\circ$), and this special geometry leads to an intense current sheet, known as the *equatorial electrojet* (see figure 3.13) that flows along the dip equator and is concentrated in a strip only a few degrees wide in geomagnetic latitude (Tascione, 1994). During day time, the equatorial electrojet flows toward the east, and during night time toward the west, however the produced electric field in night time is small because of small electron concentration at night.

The presence of an eastward electric field results in an upward drift motion of the charged particles (see section 3.3.1) into the high altitudes (F-region). The drift velocity is charge independent, therefore both ion and electron move together in the direction perpendicular

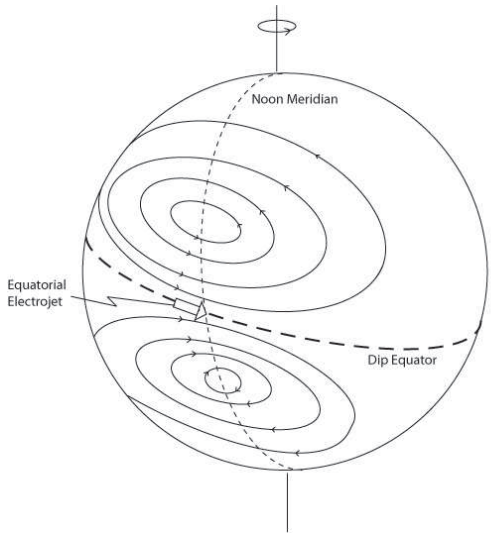

Figure 3.13: Idealized S_q electric current system induced by the solar heating (adapted from Tascione, 2004)

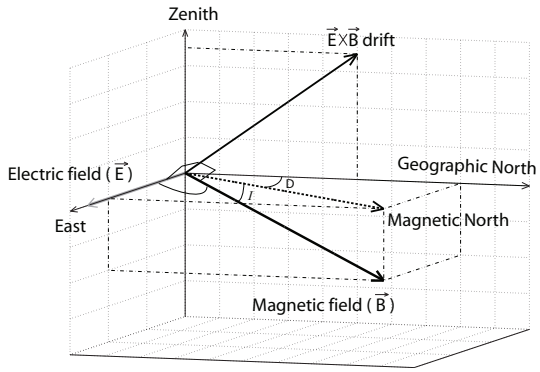

Figure 3.14: Electromagnetic drift caused by the west-to-East electric field

to the magnetic field $\vec{B}$ and the electric field $\vec{E}$. This ambipolar movement of the charged particles in the ionosphere is known the *electromagnetic drift* or $\vec{E} \times \vec{B}$ *drift*, with velocity

$$v_{drift} = \frac{E}{B} \cos(I) \tag{3.48}$$

The electromagnetic drift is the main mechanism that can cause plasma to move across magnetic field lines at high ionospheric altitudes (see figure 3.14).

3.4 Ionospheric stratification

The above the turbopause, which typically starts at 80 km, in the Earth's atmosphere density of each atmospheric component reduces exponentially with height. The reduction depends on the scale height H. Since scale height is inversely proportional to mass of the component and proportional to the temperature, the relative concentration of components is a function of height, and in high altitudes the lightest components become predominant

(Banks, 1969). Typically profiles are given in figure 2.4. From 80 km to 170 km, the predominant components are initially N_2 and O_2. Above 170 km atomic oxygen O, produced by photodissociation of O_2 ($O_2 + photon \rightarrow O + O$) becomes predominant. Finally, at very high altitudes, atomic hydrogen becomes predominant.

According to section 3.1, the photoionization and photodissociation of the atmospheric components are caused by various parts of the solar UV spectrum. The absorption of the solar incidence energy depends on the absorption cross-sections, or more accurately on the ionization cross-section (see equation 3.2). Consequently, because of the height dependency of the composition and the absorption characteristics of the atmosphere, the Earth's ionosphere has a tendency to be horizontally stratified at all latitudes, so that different ionospheric regions are formed. The term *regions* is preferred rather than *layers* because the boundary transitions are not distinct. The ionospheric regions not only differ from the predominant ion composition point of view, but also their formation mechanisms are not the same. From the composition point of view, the ionosphere can be divided into four main ionized regions: the D, E, F and *topside* regions in order of increasing altitude. In the following sections these ionospheric regions are explained. Note that the heights given in the following sections should only be considered as indication only.

3.4.1 The D-Region

The lowermost region of the ionosphere, bellow a height of roughly < 90 km, is called the D-region. It has comparatively weak electron density that develops shortly after sunrise and disappears shortly after sunset. In the D-region the primary source of photoionization is solar X-rays, causing ionization of the major neutral gases N_2 and O_2, and strong solar *Lyman*-α radiation (wavelength $\lambda = 1216\dot{A}$) causing ionization of the nitric oxide NO. As can be seen from figure 3.1, the *Lyman*-α wavelength can penetrate to heights of the D-region with little absorption at high altitudes. Since most of the X-rays radiation that could ionize gases in the D region are absorbed at greater heights this relatively strong radiation plays important role in the ion production in the D-region. However, the vertical profile of the electron density in the D-region does not have a typical maximum (Hunsucker and Hargreaves, 2003).

Since the neutral gas density is relatively large in the D-region collision with electrons and positive ions are more frequent resulting in negative ions (mostly O_2^-) due to attachment (see subsection 3.2.3). This O_2^- can react with other particles to produce other kind of negative ions. The distinguishing feature of the D-region is therefore the predominance of negative ions. The produced negative ions disappear either by recombination with a positive ions or by a process in which the electron is detached to become free again (Ratcliffe, 1972). The attachment and detachment of electrons occur very fast, and the electrons finally disappear mostly due to recombination with positive ions.

3.4.2 The E-Region

The E-region is an ionospheric region that extends between heights typically from 90 to 170 km. This region of the ionosphere is dominated by O_2^+ and NO^+ ions. O_2^+ is mainly

produced by photoionization of neutral diatomic oxygen and NO^+ produced by a rapid charge exchange process between primary ions O^+, N_2^+ and O_2^+ (see subsection 3.2.3). Unlike the D-region, the ionization of the E-region remains at night due to the solar particle (or proton) radiation, though it is considerably diminished.

The collision frequency of the charge particles increased with respect to their angular gyro-frequencies in heights of the E-region (Prolss, 2004) and results in a very short lifetimes of the principal ions NO^+ and O_2^+ (about 10 second). Because of the high collision frequency, the transport of the charged particles also becomes quiet difficult, therefore the transport process is negligible in the E-region as compared with the other two photochemical processes (Tascione, 1994; Ivanov-Kholodny and Mikhailov, 1986; Schunk and Nagy, 2000; Titheridge, 2000). The main consequence of this condition is that the *conductivity* of the ionosphere is increased in this region and reaches to maximum value in height about 100 km, which leads to possibility of the electric currents and the formation of the central *dynamo layer* or the *conducting plasma layer*.

In the E-region, NO^+ and O_2^+ disappear principally by dissociative recombination, which first forms unstable intermediate molecules, then separate into the individual neutral atomic species (see subsection 3.2.3). In the E-region, the ion loss rate is proportional to squared electron density n_e^2. Because of this, under geomagnetically quiet condition, the vertical electron density profile of the E-region is reasonably well explained by the Chapman model. It can be said that the E-region of ionosphere is formed under *photochemical equilibrium* (ion production rate is balanced by the ion disappearance rate). The electron density has a maximum at a height about $110\ km$ and the maximum electron density is typically 10^5 electrons per cubic centimeters during the day in the mid-latitudes. Because the neutral density is greater than $10^{11}\ cm^{-3}$ the E-region is only a weakly ionized region with respect to higher regions (Schunk and Nagy, 2000). The mean coefficient for the dissociative recombination of electrons and molecular ions (O_2^+ and NO^+) is $\alpha \approx 10^{-7}\ cm^3/s$ (Titheridge, 2000).

3.4.3 The F-Region

The F-region is located between 170 km and 600 km. It is divided into lower F1 (170-200 km) and upper F2 (200-600 km) sub-regions.

The F1-Region In the F1-region, the main primary ions which are produced directly by photoionization are O^+ and N_2^+. The most numerous is the atomic oxygen ion O^+. Since atomic oxygen ion rarely losses its charge by the radiative recombination; it is always first converted to molecular ion and then the molecular ion losses its charge by the dissociative recombination process (see section 3.2.3). In other words, indirect dissociative recombination is carried out in case of the F1-region. While in case of the E-region, dissociative recombination is carried out directly. This is the distinguishing feature between the E- and F1-regions.

In the F1-region due to the relatively high density of the neutral molecules (Oxygen and Nitrogen molecules) the loss rate is proportional to squared electron density n_e^2. Thus, vertical electron density profile of this region is also explained reasonably well by the Chapman model. It implies that the formation mechanism of the F1-region is governed by the

photochemical equilibrium.

The maximum electron density of the F1-region appears as a bulge below the main peak ($F2$-*peak*) of vertical electron density profile of ionosphere when the height of maximum electron density of the F1-region ($h^0_{max,F1}$) is smaller than the transition height (h_t) (Ratcliffe, 1972; Hunsucker and Hargreaves, 2003). The bulge is called the $F1$-*peak* and its appearance depends on the solar cycle and the solar zenith angle. Since $h_t - h^0_{max,F1}$ is increased for small solar zenith angles the F1-peak is more common near midday and in summer time (Ratcliffe, 1972).

The F2-Region This is the region between roughly 200 km to 600 km in which has the greatest concentration of electrons of any region. The F2-region is the most important region from navigation and space communication point of view. It is also the region which is the most variable and the most irregular from prediction point of view. The electron concentration reaches its maximum value in the F2-region (particularly in the F2-peak) and persists during day and night times. F2-region is dominated by the atomic oxygen ion O^+. The maximum electron density is typically 10^6 electrons per cubic centimeters, which is roughly a factor of 10 greater than that of the E-region. The neutral density ($10^8\ cm^{-3}$) is still orders of magnitude greater than the ion density (Schunk and Nagy, 2000). Above the F2-peak, the electron density decreases exponentially with altitude.

In this region, the rate of ion production is maximized. Due to a smaller number of the neutral particles the recombination rate is proportional to the electron density n_e (see subsection 3.2.3). Furthermore, the collision frequency of the charge particles is less than their angular gyro-frequencies, which results in an increased lifetime of the charged particles. Therefore the transport process can not be neglected. In the F2-region, all three photochemical processes contribute simultaneously to its formation. In the F2-region a transition from photochemical equilibrium to diffusive equilibrium takes place (Ivanov-Kholodny and Mikhailov, 1986; Schunk and Nagy, 2000).

3.4.4 The topside region and the protonosphere

The topside ionosphere is the region above the F2-peak extended from about 600 to 1000 km where atomic Oxygen ion is still dominant. In this region, the decrease in the density of neutral particles does increase the coefficient of diffusion, but also decreases the production and loss rates. Consequently, the transport process becomes the principal process, while the processes of ionization and recombination are insignificant in the formation mechanism of the topside region. In other words, the diffusive equilibrium is dominant. The electron density decreases exponentially with altitude in the topside region.

Above the topside region, there is a region where the lighter ions (Hydrogen ion H^+ and Helium ion He^+) dominate. This region of ionosphere is effectively is fully ionized, is called the *plasmasphere* or the *protonosphere*. The boundary between the topside and protonosphere regions is defined as the transition from atomic oxygen to atomic hydrogen as the primary ion constituent. This transition occurs at an altitude roughly 600 to 2000 km. The protonosphere is often considered as part of the ionosphere for radio and navigation applications.

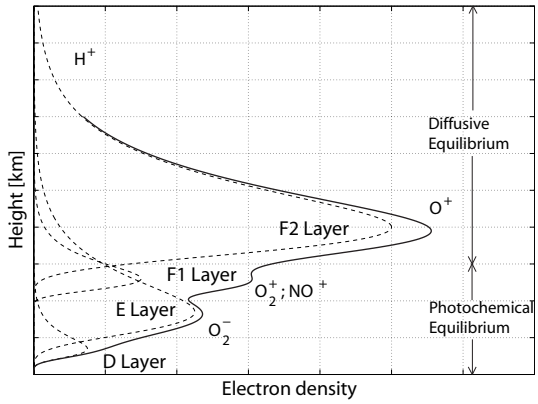

Figure 3.15: A typical height profile of the ionospheric electron density during day time at the mid-latitude region (as the summation of the E, F1, and F2)

3.4.5 Vertical electron density profile of the ionosphere

In table 3.1, the main characteristics of the ionospheric regions are outlined. A typical profile of the vertical electron density during day time at the mid-latitude region is shown in figure 3.15. The solid curve gives the electron density and dashed curves indicate the contributions from each regions (D, E, F1, F2, and topside). Vertical structure of the ionosphere governed by two equilibrium regimes, photochemical and diffusive, which play an important role in the formation of the ionospheric regions at different heights. The E- and F1-regions are formed in photochemical equilibrium, the *topside* region is formed in diffusive equilibrium, and the F2-region is a transition between these two equilibriums. In this thesis we will assume that also in the F2-region the diffusive equilibrium regime is the dominant regime.

Table 3.1: The characteristics of the ionospheric regions

Region	D	E	F1	F2	Topside	Protonosphere
Neutral Particles	N_2, O_2, NO	N_2, O_2, NO	N_2, O_2, O	N_2, O_2, O	O, H, He	H, He
Height range $[km]$	< 90	90-170	170-200	200-600	600-1000	$1000 <$
Nominal neutral density $[m^{-3}]$	$10^{11} <$	10^{11}	$10^{8} <$	10^{8}	$< 10^{8}$	-
Dominant ions	N_2^+, O_2^+, NO^+	O_2^+, NO^+	O_2^+, NO^+, O^+	O^+	O^+	H^+, He^+
Source of ionization	X rays, Layman-α	X rays, charge particle radiation	X rays	X rays	X rays	X rays
Dominant photochemical process	Recombination, Photoionization, Attachment	Recombination, Photoionization	Recombination, Photoionization	Recombination, Photoionization, Transport	Transport	Transport
Type of recombination	-	Dissociative recombination	Dissociative recombination	Dissociative recombination	Radiative recombination	-
Recombination coefficient	-	Quadratic	Quadratic	Linear	Linear	-
Chapman Layer	No	Yes	Yes	No	No	No
Appearance duration	Daytime	Daytime Nighttime	Daytime	Daytime Nighttime	Daytime Nighttime	Daytime Nighttime
Nominal daytime electron density $[m^{-3}]$	$< 10^{9}$	10^{11}	2×10^{11}	10^{12}	10^{10}	10^{9}
Nominal height of Maximum electron density	-	$110\ km$	h_t	$300\ km$	-	-
Equilibrium regime	-	Photochemical equilibrium	Photochemical equilibrium	-	Diffusive equilibrium	Diffusive equilibrium
Special features	Negative ions (mostly O_2^-)	Maximum conductivity	-	Plasma fountain	-	Exosphere

3.4.6 Characteristic parameters of the ionospheric regions

Every electromagnetic wave is modified during its passage through the ionosphere and its propagation direction, amplitude and velocity are changed. This is not only needs to be considered in radio communication, it can be also used to extract information about the properties of the ionospheric regions for ionospheric research. The interaction between an electromagnetic wave and the ionosphere results in the oscillation of the charge particles in the ionosphere, regardless the geomagnetic field. The angular frequency of the oscillation is given by (Dendy, 1993)

$$\omega = \sqrt{\frac{ne^2}{\varepsilon_0 m}} \tag{3.49}$$

which n is charged particle density per cubic meter, e and m are charge and mass of the particle and ϵ_0 is the electric permittivity of vacuum ($8.85 \times 10^{-12}\ Farad/m$). Since the mass of an ion is much larger than that of an electron, the ions oscillate at much lower amplitude than the electrons, and so to a very good approximation the motion of the ions can be ignored. In other words, only the electrons respond to the wave. Therefore the angular frequency of electrons is generally known as the plasma frequency. As a handy rule-of-thumb $\omega_p \simeq 56.4\sqrt{n}$ or $f_p \simeq 8.98\sqrt{n}$ where ω_p is in $1/s$ and f_p is in Hz (Prolss, 2004).

It is well-known that the ionosphere is a *dispersive medium* and the type of interaction between a radio wave (from KHz to GHz) and the ionosphere depends on the frequency of the wave. When frequency of a radio wave is less than the plasma frequency, the ionosphere behaves like a metallic mirror, if the frequency is larger than the plasma frequency the wave penetrates into the ionosphere without being reflected. Due to height dependency of the electron density in the ionosphere a transmitted radio wave from the ground in vertical direction penetrates into the ionosphere up to that height where the local plasma frequency reaches the wave frequency. At this point the wave is reflected back toward the Earth (see figure 3.16). Radio sounding using a *ionosonde* take advantage of this property of the ionosphere to measure maximum frequency at which a wave reflects from each ionospheric region. This is known as the *critical frequency* or *peak plasma frequency* or *maximum plasma frequency* of the region.

The ionosphere does not reflect radio waves on frequencies above 30 MHz, because the electrons can not respond fast enough to changes in the electric field of those waves to reflect them back to the Earth for these high frequencies. Consequently, frequencies above 30 MHz usually penetrate the ionosphere without any reflection and are very useful for ground to space communications.

Typically, an ionospheric region (E, F1, and F2) is generally described by the peak plasma frequency (foE, $foF1$, and $foF2$) and peak height (hmE, $hmF1$, and $hmF2$) for which the electron density decreases with altitude on both sides of peak height. The maximum or peak electron density of the ionospheric regions, which is proportional to the squared peak plasma frequencies, are denoted by NmE, $NmF1$, and $NmF2$. Another ionospheric characteristic parameter, which is usually used in modeling of the ionosphere, is a propagation factor defined as $M(3000)F2 = MUF(3000)/foF2$ with $MUF(3000)$ the highest

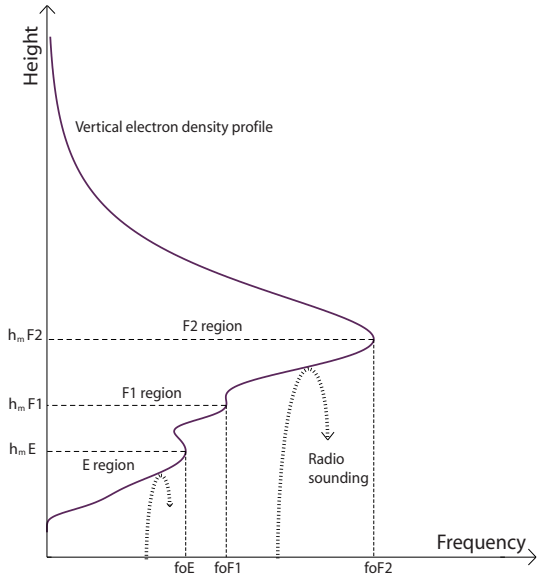

Figure 3.16: Illustration of the peak plasma frequency and the peak height

frequency refracted by the ionosphere which can be received at a distance of $3000\ km$ from the transmitter. $M(3000)F2$ is called the F2 region *transfer* parameter and is closely correlated with the peak height of the F2 region (Dieminger et al., 1996). These parameters, which are known as *ionosonde parameters*, are time and space dependent.

3.5 Spatial and temporal variability of the ionosphere

The ionosphere shows a wide range of spatial and temporal variability of the electron density, ranging from small fluctuations to large changes. In this section we describe the regular variations of the ionosphere typically under geomagnetically quiet condition.

3.5.1 Regular variations

The most important physical property of ionosphere is the existence of free electrons. According to equation 3.30, the ion production rate depends on the solar zenith angle and ionizing radiation intensity. Because of the temporal variation of the Earth-Sun geometry, which affect the solar zenith angle, and the changing in solar ionizing radiation intensity during the solar cycle, the number of free electrons per unit volume at a certain point in the ionosphere depends on time. The temporal variation of the Earth-Sun geometry is related to the Earth's daily rotation and yearly revolution. Accordingly, the ionosonde parameters in the ionospheric regions vary with altitude, latitude, longitude, universal time, season, solar cycle, and geomagnetic activities.

The spatial dependency of the electron density with respect to height has been explained in the previous section. The variation of the electron density with geomagnetic longitude is mainly related to changing in the solar zenith angle. The variation of the electron density

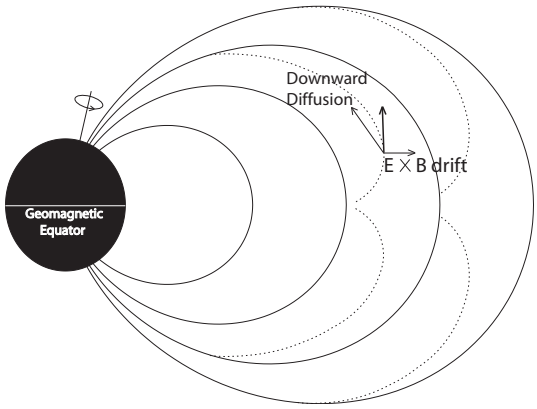

Figure 3.17: Plasma fountain from the equatorial region to higher latitudes

versus geomagnetic latitude is not only affected by the solar zenith angle, but also due to effects of the geomagnetic field which plays an important role in the transport process. Hence, considerable differences exist between the ionospheric conditions at low, middle, and high latitudes.

Specific features of the ionosphere that departure from expected behavior of the Chapman model are described as *anomalies*. The F2-region has the greatest electron density compared to other regions and it is also the most variable region. From point of view of the Chapman theory the F2-region's behavior is anomalous in several ways. These are sometimes called the *classical anomalies* of the F2-region.

Equatorial anomaly During daytime, the eastward electric field near the dip equator produces a upward $\vec{E} \times \vec{B}$ drift which in conjunction with plasma downward diffusion along the magnetic field lines due to the influence of gravity and pressure gradient forces, result in the formation of a *plasma fountain* (see figure 3.17). The plasma fountain is centered at the dip equator and transports plasma from the equatorial region to higher latitudes (Bailey et al., 1997). Because of this, when maximum electron density of F2-region ($NmF2$) at noon-time is plotted with respect to dip latitude, the electron density curve has two peaks north and south of dip equator (Anderson, 1981), as shown in figure 3.18. The peaks are located in the equatorial region on either side of the dip equator at $\pm 16^\circ$ to $\pm 18^\circ$ dip latitude (Anderson, 1973). This feature of ionosphere is the most distinctive feature of the F2-region in low latitude ionosphere and it is called the *geomagnetic anomaly*, or *Appleton anomaly* or the *equatorial anomaly*.

Diurnal anomaly The ion production rate is mainly governed by the intensity of solar radiation and the density of atmospheric gases. From theory, it is expected that ion production reaches its maximum when the solar zenith angle reaches its maximum value at local noon time. Since maximum temperature of the upper atmosphere occurs at around 14:00 local time (Rishbeth, 1972), and the atmospheric density reaches its maximum value at the same time, the maximum ion production occurs after the local noon time. This departure from expected Chapman behavior is termed *diurnal anomaly*.

Seasonal anomaly The noon values of the peak electron density in the F2-region are usually greater in winter than they are in summer, whereas the Chapman theory leads us to

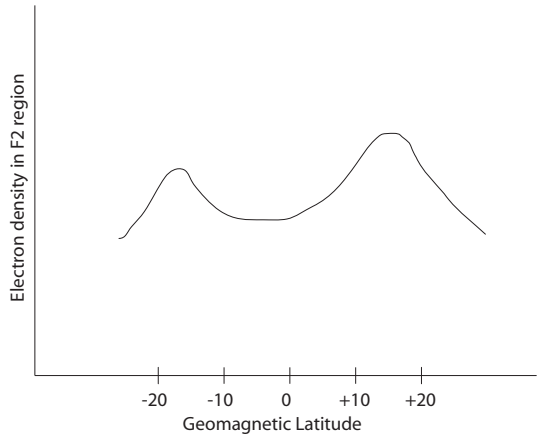

Figure 3.18: Typical equatorial anomaly in midday

expect the opposite. This is known as the *seasonal anomaly* in the F2-region. The amount of seasonal anomaly is spatial dependent and also depends on the solar cycle (Ivanov-Kholodny and Mikhailov, 1986). It is more evident in high middle latitudes than in low and high latitudes. The main reason for the seasonal anomaly is that the ratio between the atomic and the molecular components at the altitudes of the F2-region is changed and results in changes in the rates of ion production and recombination of electrons. This mechanism leads to increase loss rate of the electrons in the summer hemisphere (Dieminger et al., 1996).

3.5.2 Geomagnetic regions

The ionospheric electron density has a large dependence on the geomagnetic latitude. The ionosphere may be divided broadly into three regions that have rather different properties and features according to their geomagnetic latitudes: the *Equatorial*, *Mid-latitude*, and *Polar* (or *Auroral*) regions, see figure 3.19.

Equatorial region This the region spanning about 20° either side of the geomagnetic equator. Due to horizontal geomagnetic field geometry ($I = 0^\circ$) at the dip equator, the electromagnetic drift and associated plasma fountain, results in a large variability of the electron density in low latitudes which makes it complication to model.

Mid-Latitude region The mid latitude region of the ionosphere includes geomagnetic latitudes from about 20° to about 60°. It is out of the direct influence from phenomena associated with the plasma fountain. In this region, only solar photon radiation is responsible for the ionization process and the electron density is generally not affected by the particle radiation. The mid latitude ionosphere is the best understood of all regions. The day to day changes in the E, F1, and F2-regions show regular variations and little irregular variations associated to changes in the neutral atmosphere density and winds. The peak electron density occurs at the height where three photochemical processes are of comparable importance.

Polar region The polar region consists of the high latitudes, which are divided in the

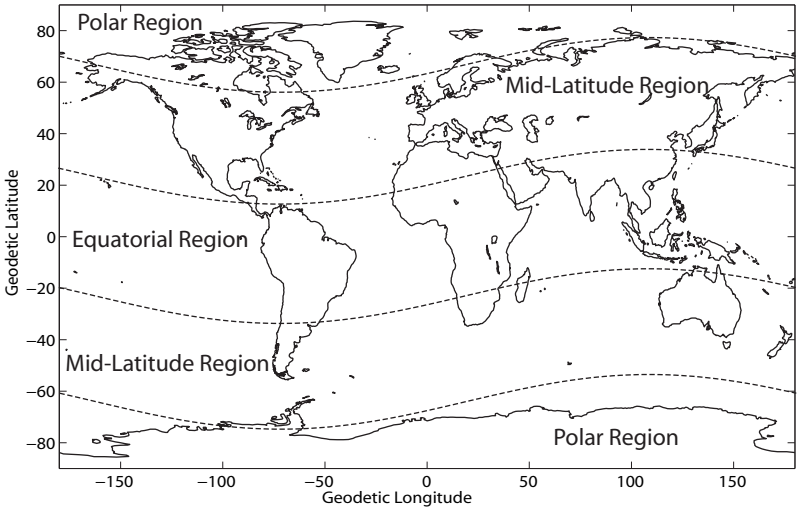

Figure 3.19: Division of the Earth into geomagnetic regions. Plotted are the geomagnetic parallel lines in the geodetic reference frame.

auroral zone (approximately $60°$-$70°$ geomagnetic latitudes) and the polar cap (poleward of the auroral zone). At high latitudes the geomagnetic field runs nearly vertical, and this leads to the existence of an ionosphere that is considerably more complex than that in the middle and low latitudes. This complexity is also because the geomagnetic field lines connect the high latitudes to the outer part of the *magnetosphere* which is driven by the solar wind, whereas the ionosphere at mid latitude is connected to the inner magnetosphere, which essentially rotates with the Earth and so is less sensitive to external influence. In the mid and low latitude regions, the primary source of ionization is the solar photon radiation, but at high latitudes, solar proton radiation produces additional ionization particularly in the E-region.

3.6 Solar disturbances

A sudden increase in the solar wind energy (the solar disturbance), accompanied by an increase in the solar photon radiation or an increase in the velocity and concentration of the solar charge particle radiation in the polar upper atmosphere, leads to not only electric currents but also heats the polar upper atmosphere. Solar disturbances includes geomagnetic activity (*geomagnetic storm*) and ionospheric disturbances (*ionospheric storm*), which increase in intensity of the Aurora. Ionospheric disturbances are *non-normal* and irregular variation of the ionosphere that are usually observed during geomagnetic activity. Although all ionospheric regions are affected by the solar disturbances, the most significant perturbations occur in the F-region particularly near the peak electron density. Therefore, subsequent sections discuss only disturbances in the F2-region.

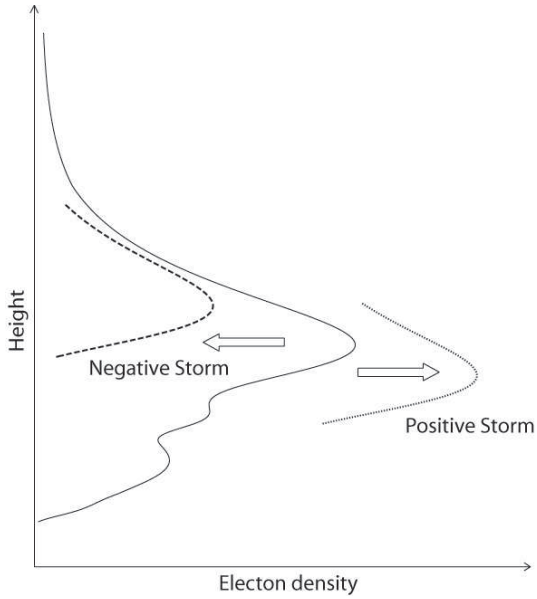

Figure 3.20: Changes in vertical electron density profile during a positive (dotted) and a negative (dashed) ionospheric storms in the F2 region, solid curve is the profile during normal ionosphere

3.6.1 Ionospheric disturbances

Observations show that during geomagnetic activity the density can both increase or decrease at mid latitudes. The term of *ionospheric storm* is used to describe several disturbances that appear in the ionosphere following a geomagnetic storm. The ionosphere reacts to the geomagnetic activity in a complex way and its behavior varies with latitude, longitude, time of day and season (Dieminger et al., 1996). The mid latitude F2-region responses to geomagnetic storm in three phases: 1) an initial or positive phase in which the peak electron density increases with respect to pre-storm conditions and lasts for a few hours on the first day of a storm (Hunsucker and Hargreaves, 2003), known as *positive* ionospheric storm, 2) a negative phase in which the peak electron density decreases relative to pre-storm conditions, known as *negative* ionospheric storm, which can last several days. 3) Finally, the ionosphere gradually returns to normal conditions over a period of one to several days in the recovery phase. The changes in the vertical electron density profile are shown in figure 3.20.

A geomagnetic storm can produce thermospheric winds, generate electric currents and change the composition in the upper atmosphere. These three effects influence the normal ionosphere structure. Every one of them can make non-normal variation in the photoionization, recombination, and transport processes and cause the irregular changes in electron concentration of the ionosphere. It should be noted that the morphology and the physics of the ionospheric storms are not completely understood and a number of open issues concerning this phenomenon exist.

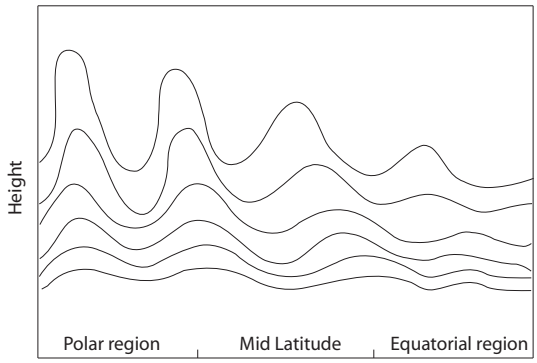

Figure 3.21: Typical propagation of the atmospheric gravity waves from polar to equatorial regions

3.6.2 Atmospheric gravity waves

In general, atmospheric waves are classified into three main groups that differ in the scale and origins (Schunk and Nagy, 2000). The largest scale waves which are propagating on a global scale, include the atmospheric tides. The smallest scale waves are related to acoustic waves, which do not play a prominent role in the dynamics of the atmosphere. The third group of atmospheric waves, which are produced by the *buoyancy forces* in the atmosphere, is known as the *atmospheric gravity waves* (AGWs).

AGWs have typically a localized source and propagate vertically as well as horizontally with a limited range of wavelengths. The amplitude of the AGWs grows exponentially with height in order to maintain a constant energy flux through an atmosphere whose density decreases with height (see figure 3.21) (Clark et al., 1971; Matsushita and Campbell, 1967). They can be either generated in the heights below the turbopause (stratosphere and mesosphere) and then propagate to the ionosphere or generated in the lower ionospheric heights (D- and E-regions). Presently it is not clear how AGWs with sources in the lower atmosphere reach the ionosphere (Rieger and Leitinger, 2002). Several sources of AGWs are known: below the turbopause by the flow of air over mountains, volcanoes, earthquakes, and above turbopause by ionospheric disturbances mostly due to aurora at the polar region (Afraimovich et al., 2002).

Typically, AGWs are divided into three groups based on period and wavelength (Hunsucker and Hargreaves, 2003). The *large-scale* AGWs have horizontal wavelengths of about 1000 km that wave periods are more than an hour and horizontal velocities of 250-1000 m/s. The *medium-scale* AGWs have horizontal wavelengths of several hundred kilometers, wave periods of about 15-70 minutes, and horizontal velocities of about 90-250 m/s. The *small-scale* AGWs have periods of 2-5 minutes with velocities less than 300 m/s and their wavelengths are less than those of the medium-scale AGWs.

AGWs play an important role in the dynamics of the upper atmosphere because they interact with the ionospheric plasma. AGWs passing through the upper atmosphere leads to changes in the ionospheric plasma and causes wavelike fluctuations of electron density (especially in the F2-region). These wavelike disturbances in the ionosphere are known as *traveling ionospheric disturbances* (TIDs).

3.6.3 Traveling ionospheric disturbances

Usually, AGWs manifest themselves in the ionosphere as traveling ionospheric disturbances which are due to collisional coupling between neutral and ionized particles. This force acts in the direction of motion of the neutral particles, but there are some complications in the ionosphere, as this effect is strongly modified by the geomagnetic field, which limits charged particles motion only along the magnetic field lines.

TIDs are usually divided into three classes (Rieger and Leitinger, 2002; Velthoven, 1990):

- *Large Scale* TIDs (LSTIDs) are the signature of AGWs generated in the auroral zone or are generated from specific geophysical events like geomagnetic sub-storms at heights of around $100\ km$, with horizontal phase velocities from 100 to $300\ m/s$ (exceeding the velocity of sound). The horizontal wavelengths are greater than $1000\ km$, with periods ranging from 30 minutes to 3 hours. LSTIDs are mostly propagating in the mid latitude region inthe direction of the equator.
- *Medium Scale* TIDs (MSTIDs) are generated mostly by AGWs originated in the lower atmosphere, but some MSTIDs with propagation toward equator in mid latitude could be signature of AGWs originated in the auroral zone. The wavelengths are between 100 to $1000\ km$, with periods from about 12 minutes to 1 hour, and horizontal phase velocities from 100 to $300\ m/s$.
- *Small Scale* TIDs (SSTIDs) have their origin only in the lower atmosphere. The wavelengths, velocities and periods are smaller than those of MSTIDs and nearly always show interference patterns.

The appearance of LSTIDs are a comparatively rare and correlate clearly with geomagnetic activity, while MSTIDs are regularly observed at the F2-region heights during the daytime both in the quiet and in disturbed ionosphere (Kalikhman, 1980). The LSTIDs move predominantly equatorwards but the movements of MSTIDs, and especially SSTIDs, are more complex.

4 Ionospheric delay measured from GNSS

In this chapter the GNSS observation equations are reviewed and the propagation of GNSS signals through the ionosphere is discussed. The structure of this chapter is the following. We will start with overview on different GNSS constellations in section 4.1. The fundamental GNSS observation equations (for phase and code) are given in section 4.2. The underlying simplifications will be described. Section 4.3 first describes the geometric effects of the ionospheric refraction, then reviews the ionospheric error on the GPS signals that approximated as a sum of error components, referred to as first-order, higher-order and signal bending terms. In section 4.4 a simple single-layer representation of the ionosphere is introduced to approximate the higher-order and bending effects. Many models for the TEC have been developed in the past, among others, GIM maps and well-known Klobuchar model. A classification of the ionospheric models is described in section 4.5. In section 4.6, the mathematical model for estimating the ionospheric effect on the GPS data using the geometry-free linear combination of the code and phase observables is derived. This model will be used for measurement of slant total electron content (TEC) in the next chapters. In section 4.7 the concluding remarks of this chapter are summarized.

4.1 Global Navigation Satellite Systems (GNSS)

The Global Positioning System (GPS) as a space-based navigation system has been developed by the US military to provide real-time and world wide absolute positioning under all weather conditions. It is the first Global Navigation Satellite System (GNSS) offering the accuracy nowadays for e.g. surveying, geodesy, navigation, and geophysics. More or less synchronously to GPS, the former Soviet Union has been developing a similar system under the name GLONASS. Both of these systems are now undergoing extensive modernization. Other systems such as the European Galileo system and the Compass from China, that are still under development, are joining the GNSS club (Hein and Eissfeller, 2007).

GPS The nominal GPS constellation consists of 24 satellites that orbiting over six orbital planes with 55° inclination and with an orbital radius about $26500\ km$. The present constellation exceeds the nominal constellation with more than 30 satellites. Since the first GPS satellite was launched in 1978, there have been 5 generations of GPS satellites, the so-called Block I, Block II/IIA, Block IIR (Replenishment), Block IIR-M (Replenishment-Modernization) and Block IIF (Follow-on). The last successful launch was on March 24, 2009 from Block IIR-M generation.

The Block I, Block II/IIA, Block IIR satellites, were launched before December 2005,

transmit their signals at two frequencies (L1=1575.42 MHz and L2=1227.6 MHz) in the L-band. The GPS carrier frequencies are derived from the same (atomic) clock driven oscillator with a frequency of 10.23 MHz. The carriers are modulated by so-called PRN (*Pseudo Random Noise*) codes, which are unique for each satellite. In this thesis the GPS satellites will be referred to by their PRN code. In the Block I, Block II/IIA, Block IIR satellites, two types of PRN codes: *Precision* (P) code and Coarse/Acquisition (C/A) code, are modulated on the carriers. The carrier waves on L1 and L2 are modulated by P-code and only the L1 carrier is modulated by the C/A-code. The P-code has a 10 times higher resolution than the C/A-code and thereby the determination of the pseudoranges can be done more precisely. Although the P-codes are much more precise than C/A-code, the P-codes are encrypted to the secret Y-code which are only available for military users. The civilian users can just access the encrypted P-codes, referred to as P(Y)-codes. This is called *Anti-Spoofing* (A-S). On the contrary, the C/A-code is available for all GPS users. The C/A-code defines *Standard Positioning Service* (SPS) and the P(Y)-code defines the *Precise Positioning Service* (PPS). Although these services are of relatively good quality, the United States planed for modernizing the signals in order to improve the quality and protection of the both civil and military users.

According to the GPS modernization plan, from December 2005 the L2 carrier wave, like the L1, was modulated with a civil C/A like code in the GPS satellites. Furthermore, for military users a Military (M) code were modulated on L1 and L2 carriers. These satellites are known as the Block IIR-M generation. Block IIF satellites are the latest generation which will transmit a third frequency (L5=1176.45 MHz) additional to L1 and L2. The L5 wave will be modulated by two I and Q codes. Table 4.1 gives an overview of the GPS signals.

Table 4.1: Overview of current and future GPS signals

carrier	frequency [MHz]	wavelength [cm]	civil	precise	military
L1	$154 \times 10.23 = 1575.42$	19.03	C/A	P	M
L2	$120 \times 10.23 = 1227.60$	24.42	C/A	P	M
L5	$115 \times 10.23 = 1176.45$	25.48	I+Q		

GLONASS The GLObal NAvigation Satellite System (GLONASS) is the Russian navigation satellite system equivalent to GPS. Its nominal constellation is composed of 24 satellites with an orbital inclination of $64.8°$ and an orbital radius of $25500\ km$. GLONASS also transmit on the L1 and L2 frequency band, but each satellite transmits its signal with a different frequency offset. The GLONASS carriers are often denoted by G1 and G2 to distinguish them from the GPS carriers. Like GPS, GLONASS is also a dual-use (military and civilian) system that provides both the SPS and PPS services. The complete GLONASS constellation of 24 operational satellites has only been available for a short time in 1996. The current status is far away from its nominal numbers and as of January 18, 2007 only 9 active GLONASS satellites were transmitting from space. It is planned the GLONASS should become full-operational by the end of 2009 (Hein and Eissfeller, 2007).

Galileo The Galileo is the European global navigation satellite system. Unlike GPS and GLONASS, Galileo is under civilian control that have international participation and investment. The Galileo satellite constellation will consist of 30 satellites (27 operational +

3 non-active spares). The satellites are divided over three orbital planes with orbital radius $29600\ km$ at an inclination of $56°$. Galileo is not yet in operation but two Galileo's satellite GIOVE-A and GIOVE-B were successfully launched respectively in December 2005 and in April 26 2008. Galileo's satellites transmit on four carrier frequencies that are given in table 4.2.

Table 4.2: Galileo's signal plan (+: with integrity message, ++: with commercial data) (adopted from (Verhagen, 2005))

carrier	frequency [MHz]	wavelength [cm]	OS	SOL	CS	PRS
E1	$154 \times 10.23 = 1575.42$	19.0	I/N	I/N+	I/N++	G/N
E6	$125 \times 10.23 = 1278.75$	23.4			C/N++	G/N
E5b	$118 \times 10.23 = 1207.14$	24.8	I/N	I/N+	I/N	
E5	$116.5 \times 10.23 = 1191.79$	25.2				
E5a	$115 \times 10.23 = 1176.45$	25.5	F/N	F/N	F/N	

Galileo plans to provide four types of navigation services:

- The *Open Service* (OS) provides basic navigation service, like GPS SPS, using freely accessible signals and which are the Open Access Navigation Signals (denoted as F/N) on E5a and the Integrity Navigation Signals (denoted as I/N) on E5b and E1 (also denoted as L1 or E2-L1-E1), but without the commercial data or integrity messages.
- *Safety-Of-Life* (SOL) provides access to the same signals as OS users, but additionally SOL users have access to the I/N including the integrity messages transmitted on the E5b and E1 carriers.
- The *Commercial Service* (CS) is also available at the same signals at the OS, but users will have additional access to commercial data (for which they have to pay) transmitted on the Integrity Navigation Signals on E5b and E1, and to the Controlled Access Navigation Signals (denoted as C/N) on E6.
- The *Public Regulated Service* (PRS) will include navigation and timing services through Restricted Access Navigation Signals (denoted as G/N) on E6 and E1.

Compass A Chinese satellite navigation system called Compass (or Beidou-2) which is presently under development. As with GPS, GLONASS, and Galileo, the system is a dual-use system which will provide two navigation services: an open service for (commercial) users and an authorized positioning, velocity, and timing communications service (Hein and Eissfeller, 2007). Compass consists of a constellation of 30 *non-stationary* satellites, five *geostationary* (GEO) satellites, and 3 more satellites in geosynchronous orbit (Hofmann-Wellenhof et al., 2008). The non-stationary satellites which will be in 3 orbital planes with an inclination of $55°$ and with orbital radius of $27900\ km$. The geosynchronous orbit satellites have an altitude of $35785\ km$ with an inclination angle equal to $55°$. China sent the first non-stationary Compass satellite into orbit on April 13, 2007 and the first GEO satellite of Compass has been launched on February 3, 2007. Each Compass's satellite transmits navigational signals the same for carrier frequencies. Little official information is publicly available about structure of the Compass's signals.

4.2 GNSS observation equations

The GNSS signals, receiver technology and observation equations have been described in many text books, such as for example (Hofmann-Wellenhof et al., 2001; Parkinson and Spilker, 1996; Teunissen and Kleusberg, 1998; Seeber, 2003; Leick, 2003). Here, we will give the GPS observation equation for both code and carrier phase observables.

4.2.1 Code or pseudo-range observation equation

A pseudo-range from code measurements is a measure of the geometric distance between receiver r and satellite s. The code pseudo-range observable $\vartheta^s_{r,j}$ on the carrier with frequency j is

$$\vartheta^s_{r,j} = c\,[t_r - t^s] + \varepsilon_{p^s_{r,j}} \tag{4.1}$$

with c the velocity of light, $t_r - t^s$ the time difference between signal reception at receiver t_r and signal transmission at satellite t^s, and $\varepsilon_{p^s_{r,j}}$ the measurement error or code observation noise. It is essential to describe the receiver and satellite times in a reference time frame. Using the GPS time as a common reference time frame, the reception and transmission times of signal can be expressed by the following equations

$$t_r(t_i) = t_i + dt_r(t_i) \tag{4.2}$$

$$t^s(t_i - \tau^s_r(t_i)) = t_i - \tau^s_r(t_i) + dt^s(t_i - \tau^s_r(t_i)) \tag{4.3}$$

where $(t_i - \tau^s_r(t_i))$ and t_i are respectively time of transmission and time of observation in the GPS time frame (in seconds). Note that the time of transmission is obtained by reducing the time of observation with the signal travel time $\tau^s_r(t_i)$. The receiver and satellite clocks errors are respectively $dt_r(t_i)$ and $dt^s(t_i - \tau^s_r(t_i))$, which are time dependent. Inserting equations 4.2 and 4.3 into 4.1, yields

$$\vartheta^s_{r,j}(t_i) = c\tau^s_r(t_i) + c\,[dt_r(t_i) - dt^s(t_i - \tau^s_r(t_i))] + \varepsilon_{p^s_{r,j}}(t_i) \tag{4.4}$$

To obtain the geometric range, the signal travel time must be corrected for systematic effect such as atmospheric refraction (Ionospheric and Tropospheric delays), instrumental delays (in the receiver and satellite) and other non-random effects such as multipath. The signal travel time for code observation $\tau^s_{r,j,g}(t_i)$ is

$$\begin{aligned}\tau^s_{r,j,g}(t_i) = \underbrace{\frac{1}{c}\rho^s_r(t_i - \tau^s_{r,j,g}(t_i))}_{\tau^s_r(t_i)} + \tfrac{1}{c}I^s_{r,j,g}(t_i) + \tfrac{1}{c}T^s_r(t_i) + \tfrac{1}{c}dm^s_{r,j}(t_i) \\ + d^s_{,j}(t_i - \tau^s_{r,j,g}(t_i)) + d_{r,j}(t_i) + \varepsilon_{p^s_{r,j}}(t_i)\end{aligned} \tag{4.5}$$

with:

$\rho^s_r(t_i - \tau^s_{r,j,g}(t_i))$	: Geometric distance between satellite and receiver at the time of transmission [m]
$I^s_{r,j,g}(t_i)$	: Ionospheric delay (frequency dependence) [m]
$T^s_r(t_i)$	: Tropospheric delay [m]
$dm^s_{r,j}(t_i)$	: Sum of other (non-random) code errors [m]
$d^s_{,j}(t_i - \tau^s_{r,j,g}(t_i))$	: Instrumental code delay of the satellite at the time of transmission [s]
$d_{r,j}(t_i)$	: Instrumental code delay of the receiver [s]

The travel time for the code is the group travel time of the signal, is different from the travel time for the carrier, which is why a subscript g is used. Substituting of equation 4.5 for the travel time in equation 4.4, the full code observation equation is obtained:

$$\begin{aligned}\vartheta^s_{r,j}(t_i) = \rho^s_r(t_i - \tau^s_{r,j,g}(t_i)) + I^s_{r,j,g}(t_i) + T^s_r(t_i) + dm^s_{r,j}(t_i) \\ +c\left[dt_r(t_i) - dt^s(t_i - \tau^s_{r,j,g}(t_i)) + d_{r,j}(t_i) + d^s_{,j}(t_i - \tau^s_{r,j,g}(t_i))\right] + \varepsilon_{p^s_{r,j}}(t_i)\end{aligned} \tag{4.6}$$

The mathematical expectation of the random error should be zero, i.e. $E\left\{\varepsilon_{p^s_{r,j}}(t_i)\right\} = 0$.

4.2.2 Carrier beat phase observation equation

The GPS carrier beat phase or carrier phase observable is obtained by trading the phase difference between the received signal and a copy in the receiver. The carrier phase observable is more precise than the code observable, but it is an ambiguous measure of the geometric distance between a satellite and the receiver. The carrier phase observable [in cycles] with frequency j can be mathematically written as

$$\varphi^s_{r,j}(t_i) = \varphi_{r,j}(t_i) - \varphi^s_{,j}(t_i - \tau^s_{r,j}(t_i)) + N^s_{r,j} + \varepsilon_{\varphi^s_{r,j}} \tag{4.7}$$

with φ^s the phase of the carrier signal in satellite at transmission time $(t_i - \tau^s_r(t_i))$, and φ_r the phase of the carrier signal in the receiver at reception time t_i, an integer number of full cycle N^s_r is unknown, and $\varepsilon_{\varphi^s_{r,j}}$ measurement error or phase observation noise. The phase ambiguity is time-constant as long as there is no interruption when tracking the satellite (no cycle slips are present).

The phase of the receiver and satellite can be written as

$$\varphi_{r,j}(t_i) = f_j t_i + \varphi_{r,j}(t_0) \tag{4.8}$$

$$\varphi^s_{,j}(t_i - \tau^s_{r,j}(t_i)) = f_j t^s(t_i - \tau^s_{r,j}(t_i)) + \varphi^s_{,j}(t_0) \tag{4.9}$$

with

f_j	: carrier frequency $[s^{-1}]$
$\varphi_{r,j}(t_0)$	: initial phase in the receiver [cycles]
$\varphi^s_{,j}(t_0)$	: initial phase in the satellite [cycles]

Substituting equations 4.8 and 4.9 into 4.7 and using equations 4.2 and 4.3 for the reception and transmission times in the GPS time frame, yields

$$\varphi^s_{r,j}(t_i) = f_j\left[dt_r(t_i) + \tau^s_{r,j}(t_i) - dt^s(t_i - \tau^s_{r,j}(t_i))\right] + \left[\varphi_{r,j}(t_0) - \varphi^s_{,j}(t_0) + N^s_{r,j}\right] + \varepsilon_{\varphi^s_{r,j}}(t_i) \tag{4.10}$$

Similar to the travel time of the code observable in equation 4.5, the signal travel time of the phase can be re-written as

$$\tau^s_{r,j,\varphi}(t_i) = \underbrace{\frac{1}{c}\rho^s_r(t_i - \tau^s_{r,j,\varphi}(t_i))}_{\tau^s_r(t_i)} - \frac{1}{c}I^s_{r,j,\varphi}(t_i) + \frac{1}{c}T^s_r(t_i) + \frac{1}{c}\delta m^s_{r,j}(t_i) + \delta^s_{,j}(t_i - \tau^s_{r,j,g}(t_i)) + \delta_{r,j}(t_i)$$

(4.11)

where

$\rho_r^s(t_i - \tau_{r,j,\varphi}^s(t_i))$	: Geometric distance between satellite and receiver at the time of transmission [m]
$I_{r,j,\varphi}^s(t_i)$	: Ionospheric delay (frequency dependence) [m]
$T_r^s(t_i)$	: Tropospheric delay [m]
$\delta m_{r,j}^s(t_i)$	: Sum of other (non-random) phase error [m]
$\delta_{,j}^s(t_i - \tau_{r,j,\varphi}^s(t_i))$	: Instrumental phase delay in satellite at the time of transmission [s]
$\delta_{r,j}(t_i)$	: Instrumental phase delay in receiver [s]

The tropospheric delay is independent of the frequency and equal for both phase and code observables. Substituting equation (4.11) in equation (4.10) gives the carrier phase observation equation [in cycles] as follows

$$\begin{aligned}\varphi_{r,j}^s(t_i) = \tfrac{1}{\lambda_j}\left[\rho_r^s(t_i - \tau_{r,j,\varphi}^s(t_i)) - I_{r,j,\varphi}^s(t_i) + T_r^s(t_i) + \delta m_{r,j}^s(t_i)\right] \\ + f_j\left[dt_r(t_i) + \delta_{r,j}(t_i) - dt^s(t_i - \tau_{r,j,\varphi}^s(t_i)) + \delta_{,j}^s(t_i - \tau_{r,j,\varphi}^s(t_i))\right] \\ + \left[\varphi_{r,j}(t_0) - \varphi_{,j}^s(t_0) + N_{r,j}^s\right] + \varepsilon_{\varphi_{r,j}^s}(t_i)\end{aligned} \tag{4.12}$$

where, $\lambda_j = f_j/c$ is wavelength of the carrier frequency j in vacuum. The mathematical expectation of the measurement error is zero, i.e. $E\left\{\varepsilon_{\varphi_{r,j}^s}(t_i)\right\} = 0$.

4.2.3 Simplifications of the observation equations

Although the signal travel time is different for different observation types and different frequencies, in computation of the transmission time of the signal it is allowed to assume that the signal travel times of code and phase for all frequencies are equal:

$$t_i - \tau_{r,j,g}^s(t_i) = t_i - \tau_{r,j,\varphi}^s(t_i) = t_i - \tau_r^s(t_i) \tag{4.13}$$

The difference between the travel times of different observation types and frequencies is less than 10^{-7} seconds which corresponds to sub-millimeter satellite position differences. Furthermore, the clocks errors and instrumental delays are stable during the small sub microsecond.

The ionospheric delay consist of the first, second, higher orders and bending effects which will be described in the next section. The first order effect, which contains the main part of the ionospheric delay, has opposite sign for code and phase observables. By ignoring very small higher order and bending effects, taking only first-order ionospheric effects into account, we may write $I_{r,j}^s(t_i) = I_{r,j,g}^s(t_i) = -I_{r,j,\varphi}^s(t_i)$ (without specifying subscript of observable type). Since the ionosphere is a dispersive medium, the ionospheric delay for the other frequencies can be computed by the following equation

$$I_{r,j}^s(t_i) = \mu_j I_{r,1}^s(t_i); \qquad \mu_j = \lambda_j^2/\lambda_1^2 = f_1^2/f_j^2 \tag{4.14}$$

with $I_{r,1}^s$ the ionospheric delay on L1. Hence, the ionospheric delay can be parameterized as the first-order ionospheric slant delay of code observable on the L1 frequency $I_r^s(t_i) = I_{r,1}^s(t_i)$ (the frequency subscript 1 for L1 will be omitted) in all observation equations.

Non-random errors δm and dm respectively for phase and code observations are included antenna phase center variations and offsets and phase wind-up. The GPS observations have to be corrected for these errors. The corrected observations are shown in unit of length as

$$\begin{aligned} \phi^s_{r,j}(t_i) &= \lambda_j \varphi^s_{r,j}(t_i) - \delta m^s_{r,j}(t_i) \\ p^s_{r,j}(t_i) &= \vartheta^s_{r,j}(t_i) - dm^s_{r,j}(t_i) \end{aligned} \tag{4.15}$$

Other non-random errors like multipath can not be modeled a priori. For simplicity these are taken into account in $\epsilon^s_r(t_i)$.

Since the initial phase of the signal and integer phase ambiguity are not computed separately therefore they are lumped together which results in a non-integer ambiguity term in unit of length $M^s_{r,j}$ as follows

$$M^s_{r,j} = \lambda_j \left(\varphi_{r,j}(t_0) - \varphi^s_{,j}(t_0) + N^s_{r,j}\right) \tag{4.16}$$

When the GPS observables are not used for the positioning purpose a geometry-dependent term $g^s_{r,j}$ can be introduced by summing all frequency-independent terms as follows (in unit of length)

$$g^s_r(t_i) = \rho^s_r(t_i - \tau^s_r(t_i)) + T^s_r(t_i) + c\left[dt_r(t_i) - dt^s(t_i - \tau^s_r(t_i))\right] \tag{4.17}$$

Taking all above mentioned assumptions and simplifications into account and considering the instrumental code delays in unit of length, the modified observation equations (for code and phase) are given as follows for epoch t_i :

$$\begin{aligned} \phi^s_{r,j}(t_i) &= g^s_r(t_i) - \mu_j I^s_r(t_i) + \delta_{r,j}(t_i) + \delta^s_{,j}(t_i) + M^s_{r,j} + \varepsilon_{\varphi^s_{r,j}}(t_i) \\ p^s_{r,j}(t_i) &= g^s_r(t_i) + \mu_j I^s_r(t_i) + d_{r,j}(t_i) + d^s_{,j}(t_i) + \varepsilon_{p^s_{r,j}}(t_i) \end{aligned} \tag{4.18}$$

where the instrumental delays are in meter. Note that for notational convenience we used $d^s_{,j}(t_i)$ instead of $d^s_{,j}(t_i - \tau^s_r(t_i))$ and $\delta^s_{,j}(t_i)$ instead of $\delta^s_{,j}(t_i - \tau^s_r(t_i))$.

4.2.4 Tropospheric effects

The troposphere, the lower part of the atmosphere, is a *non-dispersive* medium whereas the ionosphere is a *dispersive* medium (the propagation velocity is frequency dependent) in the upper part of the atmosphere. In contrast to the ionosphere, in the non-ionized troposphere the phase and group velocities of the electromagnetic wave are exactly equal, and also independent of the frequency. The tropospheric delay is not constant, but variable in both space and time. More details on the tropospheric refraction effects can be found in e.g. (Kleijer, 2004).

4.3 Ionospheric propagation of GNSS signals

When an electromagnetic wave propagates in free space, its velocity is known to be equal to the *velocity of light*. When the wave propagates in a medium, its velocity changes due

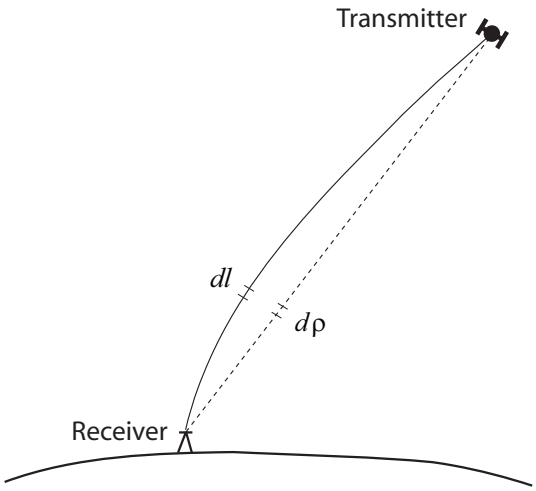

Figure 4.1: Bending of a wave through inhomogeneous refractive medium.

to interaction with the particles present in that medium. This is known *wave refraction* and the amount of refraction is described by the medium specific *refractive index*.

Refractive index Refractive index of a medium n is defined as the ratio of the velocity of light in free space c and the velocity of the wave in a medium v,

$$n = c/v \tag{4.19}$$

Since the refractive index is a ratio of two velocities, it is a dimensionless quantity. In free space the refractive index equals 1. When the refractive index is smaller than 1 we say that the wave is *advanced* and when it is larger than 1 it is *delayed*.

The ionospheric refractive index is unfortunately not a constant. This is because the ionosphere is an *inhomogeneous*, *anisotropic* and *dispersive* medium. In the following, the propagation properties of the ionosphere are described in detail.

4.3.1 Inhomogeneity of the ionosphere

When a property of a medium is the same at different points of the medium, the medium is known as a *homogeneous* or *uniform* in that property. The medium is *inhomogeneous* when the property changes at different points. A medium can be homogeneous in one property and inhomogeneous in another. Since electron density within the ionosphere differs for different locations, the ionosphere is an inhomogeneous medium. As a consequence, the ionospheric refractive index varies significantly in the spatial domain. Changing the ionospheric refractive index along propagation path results in bending of the path of a signal ray, making the path longer than the geometrical straight line path ρ, see figure 4.1.

Geometric refraction effects According to *Fermat's principle*, among all possible paths that the wave might take in an inhomogeneous refractive medium, it takes the path which requires the shortest time. In figure 4.1, the path length between transmitter and receiver

denoted as l, is computed as follows

$$l = c\int dt = c\int dl/v = \int ndl \quad = \rho + \underbrace{\int (n-1)d\rho}_{\iota} + \underbrace{(\int ndl - \int nd\rho)}_{\kappa} \tag{4.20}$$

where ρ denotes the geometric distance between transmitter and receiver. The path length due to refraction consists of two effects: A *propagation effect*, denoted as ι, and a *bending effect*, denoted as κ. Therefore, the ionospheric delay between a GPS satellite s and a receiver r can be written as

$$I_r^s = \iota_r^s + \kappa_r^s \tag{4.21}$$

4.3.2 Dispersivity of the ionosphere

If the refractive index of a medium depends upon the frequency of the wave, the medium is said to be *dispersive medium*. To describe the propagation of a modulated electromagnetic wave through a dispersive medium, we should distinguish between the *phase velocity*, denoted as v_ϕ, and the *group velocity* of the signal, denoted as v_g. Hence, refractive indexes for phase and group of the signal differ and are

$$\begin{aligned} n_\phi &= c/v_\phi \\ n_g &= c/v_g \end{aligned} \tag{4.22}$$

with $v_g \leq v_\phi$. The relation between v_g (velocity of the modulation of a wave) and v_ϕ (velocity of the carrier wave) is given by the *Rayleigh equation*

$$v_g = v_\phi + f\frac{\partial v_\phi}{\partial f} \tag{4.23}$$

where f is frequency of the carrier wave. Using this equation, the phase and group refractive indexes are connected as follows

$$n_g = n_\phi + f\frac{\partial n_\phi}{\partial f} \tag{4.24}$$

For electromagnetic waves in the *L-band* (a portion of the electromagnetic spectrum ranging roughly from one to two GHz), the ionosphere is dispersive (but the troposphere is not). Therefore, the ionosphere acts on the GPS carrier frequencies as a dispersive medium. It means that the phase refractive indexes of the GPS carriers frequencies are different. And also, the group and phase refractive indexes are different. Since $v_g < v_\phi$, the time of the code, denoted as $\tau_{r,j,g}^s$, is larger than the corresponding travel time of phase, denoted as $\tau_{r,j,\varphi}^s$ where j stands for frequency. This phenomenon is referred to as *code-carrier divergence* (Misra and Enge, 2006).

So the ionospheric delay for the GNSS phase and code observations at frequency j between satellite s and receiver r are written as follows

$$\begin{aligned} I_{r,j,\varphi}^s &= \iota_{r,j,\varphi}^s + \kappa_{r,j}^s \\ I_{r,j,g}^s &= \iota_{r,j,g}^s + \kappa_{r,j}^s \end{aligned} \tag{4.25}$$

The bending effects for the phase and code observations are the same.

4.3.3 Anisotropy of the ionosphere

A medium is said to be *isotropic* if the phase refractive index (or the phase velocity) of a wave is independent of the direction. Since the Earth's ionosphere is coupled with the geomagnetic field, the ionospheric refractive index depends on the direction of wave propagation relative to the geomagnetic field ($\vec{B}$). Therefore, the ionosphere is not an *isotropic* medium for electromagnetic wave propagation. That is why the GNSS waves are *circularly* polarized. With a *linear* polarization, the free electrons in the ionosphere would interact with the wave and cause its polarization to rotate which is known as *Faraday rotation*, which would cause signal fluctuations. However, the Faraday rotation does not or hardly affect the intensity of the received signal when the waves are circularly polarized (Hall et al., 1996).

There is also another effect. Under influence of the geomagnetic field a GPS wave is split up into *two* approximately parallel wavefronts, which each have an opposite (circular) polarization. One wave, the *ordinary* wave, has a right-handed polarization, but the other one, the *extraordinary* wave has a left-handed polarization. Both waves show a small difference in propagation velocity and consequently in refractive index. This effect is known as *double refraction* or *birefringence*. Despite this double refraction, only the ordinary wave needs to be considered in case of GNSS, because the extraordinary wave contains less than 0.35% of the power (for $L1$) (Bassiri and Hajj, 1993).

4.3.4 Ionospheric refractive index

The phase refractive index for the electrically neutral ionosphere, with a uniform magnetic field and neglecting the effect of the positive ions on the wave, is expressed by the complex *Appleton-Hartee* (or *Appleton-Lassen*) formula. Ignoring absorption effects due to collisions between the electrons, the formula reads, e.g. (Giraud and Petit, 1978)

$$n_{j,\phi,\pm} = \sqrt{1 - \frac{X_j}{1 - \frac{Y_{T,j}^2}{2(1-X_j)} \pm \sqrt{\frac{Y_{T,j}^4}{4(1-X_j)^2} + Y_{L,j}^2}}} \tag{4.26}$$

with $X_j = f_p^2/f_j^2$ and $Y_{T,j}$ and $Y_{L,j}$ the transversal respectively longitudinal components of $Y_j = f_g/f_j$. So $Y_{T,j} = Y_j|\sin\theta|$ and $Y_{L,j} = Y_j|\cos\theta|$, where θ is the angle between Y_j and $Y_{L,j}$. The frequency f_p is the *electron plasma frequency*, which according to equation 3.49, is computed as $f_p = \sqrt{An_e}$ with $A \approx 80.6\, m^3/s^2$ and n_e is the plasma electron density in m^{-3}. The frequency f_g is known as the *electron gyro-frequency*, which according to subsection 3.3.1, it is computed as $f_g = \frac{\omega}{2\pi} = \frac{e}{2\pi m_e}B$ where $e = 1.60218 \cdot 10^{-19}$ is electron charge in Coulomb and $m_e = 9.10939 \cdot 10^{-31}$ is electron mass in kg.

Note that the inhomogeneous, anisotropic and dispersive properties of the ionosphere are all present in the Appleton-Hartree formula. The *inhomogeneity* of the ionosphere is reflected in the free electron density n_e, which is not a constant, but a function of place and time. Moreover, the *dispersive* ionosphere can be recognized in the dependence on the frequency of the wave. The *anisotropic* ionosphere is expressed in the terms depending on B. The *double-refraction* is reflected by the $\pm$ sign, which means that either a plus or minus sign

can be used, depending on the polarization of the wave: A "+" corresponds to the left-handed circularly polarized wave (the extraordinary wave) and a "-" to the right-handed circularly polarized wave (the ordinary wave). Since for GNSS only the ordinary wave is significant, from now on only the refractive index using the *minus sign* is considered.

Using some simplifications for the Appleton-Hartree formula, an approximated ionospheric phase refractive index of the GNSS signals can be given (Hartmann and Leitinger, 1984)

$$n_{j,\phi} = 1 \; - \; \frac{1}{2}\frac{f_p^2}{f_j^2} \; - \; \frac{1}{2}\frac{f_p^2 f_g |\cos\theta|}{f_j^3} \; - \; \frac{1}{8}\frac{f_p^4}{f_j^4} \tag{4.27}$$

and by using equation 4.24, the ionospheric group refractive index reads

$$n_{j,g} = 1 \; + \; \frac{1}{2}\frac{f_p^2}{f_j^2} \; + \frac{f_p^2 f_g |\cos\theta|}{f_j^3} \; + \; \frac{3}{8}\frac{f_p^4}{f_j^4} \tag{4.28}$$

It is seen that the phase refractive index is always smaller than 1, while the group refractive index is larger than 1. This implies that in the ionosphere the phase of the wave is *advanced*, while at the same time its group is *delayed*.

4.3.5 Ionospheric first-, higher-order and bending effects

Substituting ionospheric refractive index in equation 4.20 by the equations 4.27 and 4.28 gives respectively the phase and group propagation effects

$$\begin{array}{rcccccccc} \iota_{j,\phi} & = & -\frac{1}{2f_j^2}\int f_p^2 d\rho & - & \frac{1}{2f_j^3}\int f_p^2 f_g |\cos\theta| d\rho & - & \frac{1}{8f_j^4}\int f_p^4 d\rho \\ \iota_{j,g} & = & \frac{1}{2f_j^2}\int f_p^2 d\rho & + & \frac{1}{f_j^3}\int f_p^2 f_g |\cos\theta| d\rho & + & \frac{3}{8f_j^4}\int f_p^4 d\rho \end{array} \tag{4.29}$$

The electron plasma frequency and the electron gyro frequency are not constant along the path and therefore remain within the integrals. Equation 4.29 shows that both the ionospheric phase advance and group delay can be expanded as a function of the *same* three effects. These three effects are better known as the ionospheric *first-*, *second-* and *third-order* delays, denoted as $\iota_{j,g}^{(1)}$, $\iota_{j,g}^{(2)}$ and $\iota_{j,g}^{(3)}$ respectively. Substituting $f_p = \sqrt{An_e}$ and $f_g = \frac{e}{2\pi m_e}B$ into equation 4.29, they are given as

$$\begin{array}{rcccc} \iota_{j,g}^{(1)} & \doteq & \frac{1}{2f_j^2}\int f_p^2 d\rho & = & \frac{A}{2f_j^2}\int n_e d\rho \\ \iota_{j,g}^{(2)} & \doteq & \frac{1}{f_j^3}\int f_p^2 f_g |\cos\theta| d\rho & = & \frac{eA}{f_j^3 2\pi m_e}\int B|\cos\theta| n_e d\rho \\ \iota_{j,g}^{(3)} & \doteq & \frac{3}{8f_j^4}\int f_p^4 d\rho & = & \frac{3A^2}{8f_j^4}\int n_e^2 d\rho \end{array} \tag{4.30}$$

The second- and third-order delays are often referred to as the ionospheric *higher-order* terms. Adding the effects of signal *bending*, the ionospheric *phase advance*, denoted by $I_{j,\phi}$, and the ionospheric *group delay*, denoted by $I_{j,g}$, can be written as

$$\begin{array}{rcccccccc} I_{j,\phi} & = & -\iota_{j,g}^{(1)} & - & \frac{1}{2}\iota_{j,g}^{(2)} & - & \frac{1}{3}\iota_{j,g}^{(3)} & + & \kappa_j \\ I_{j,g} & = & \iota_{j,g}^{(1)} & + & \iota_{j,g}^{(2)} & + & \iota_{j,g}^{(3)} & + & \kappa_j \end{array} \tag{4.31}$$

From these expressions it can be seen that the first-order phase effect is equal but opposite of sign to the first-order group effect. The second- and third-order phase effects are also opposite of sign to their corresponding group counterparts, but they do not have the same magnitude. The second-order phase effect is half the second-order group effect, while the third-order phase effect is one third of the third-order group effect.

4.4 Ionospheric Total Electron Content (TEC)

The first-order ionospheric delay is a function of the integral term $\int n_e d\rho$ of equation 4.30. This term is well known as the *Total Electron Content* (*TEC*) along the geometric line from receiver to satellite. TEC is the number of electrons in a slant column with unit-squared cross-section in the ionosphere along the signal path. It is expressed in *TECU* (TEC-Unit) with 1 *TECU*$= 10^{16}$ electrons per m^2. Using this *TECU*, the first-order delay may be rewritten as

$$\iota_{j,g}^{(1)} = \frac{A}{2f_j^2} TEC \tag{4.32}$$

with $A \approx 80.6$ m^3/s^2. TEC is highly variable in time and in space due to the variable ionospheric electron density. Moreover, it is important to realize that TEC strongly depends on the elevation angle of the satellite (receiver-satellite geometry). Since the signal path length through the ionosphere varies with the satellite position in the sky, as with a lower elevation, the signal path length gets longer and results in higher TEC.

4.4.1 A single-layer ionosphere approximation

In order to make an (rough) assessment for elevation dependency of the TEC and for the purpose of simple ionospheric modeling, the ionosphere may be considered as a *thin single-layer* surrounding the earth at a fixed height from the earth for which all free electrons in the ionosphere are assumed to be concentrated in this single-layer (see figure 4.2). If we assume there is no *lateral* or *horizontal* electron density gradients, the *vertical* TEC can be simply mapped to the slant TEC (and vice versa)

$$TEC = \frac{1}{\cos\xi} VTEC \tag{4.33}$$

where ξ is satellite zenith angle at the point of intersection of the line of sight with the spherical single-layer ionosphere and VTEC stands for vertical TEC. This point is referred to the *ionospheric (piercing) point*, denoted as IP, and the multiplier $\frac{1}{\cos\xi}$ is called *obliquity factor*. The vertical projection of IP on the ground is referred to the *sub-ionospheric point* and denoted as SIP in figure 4.2. Geometrically, relation between the satellite zenith angles at the receiver location ξ' and at the ionospheric pierce point ξ is

$$\sin\xi = \frac{R_e}{R_e + h_i} \sin\xi' \tag{4.34}$$

where h_i stands for height of the ionospheric single-layer from the Earth. Since this height is not exactly known, it is often assumed at 350 km. Note that the *co-secant mapping*

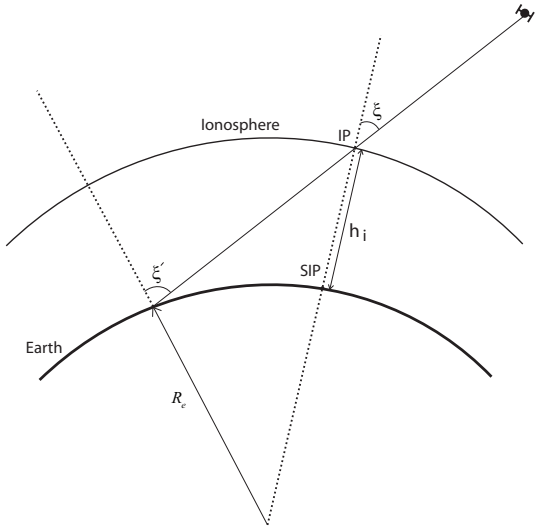

Figure 4.2: Geometry of the ionospheric single-layer approximation.

function depends on the height of the single-layer. It is shown in (van der Marel, 1993) that only for large zenith angles (about $\xi' > 70^\circ$) the mapping function is sensitive to h_i. Moreover, It is also shown that for large zenith angles TEC can reach at most *three* times the value of VTEC.

It should be emphasized that the sub-ionospheric point can be far away from the receiver location a few thousand kilometer (for large satellite zenith angles). This means that in the presence of horizontal electron density gradients, the use of a simple co-secant mapping function to convert slant TEC to vertical TEC will lead to significant TEC conversion errors (see (Nava et al., 2007)).

TEC values can range from 1 to several hundreds TECU along the signal path depend on elevation angle and time and space. In fact, variations of the TEC in time and space are due to temporal and spatial variation of the ionospheric electron density. To study the diurnal, seasonal and solar cycle behavior of TEC, it is normal to use single-layer ionosphere approximation and express slant TEC as an equivalent vertical TEC by the equation 4.33. Vertical electron conten (VTEC) can more easily be compared or modeled than slant TEC at various elevation angles.

As example, in figure 4.3 the daily mean global map of VTEC is plotted for day 130 of 2006 at UT time 14:30. The figure clearly shows spatial variation of the VTEC over the world. Note that the regions of highest VTEC are located, on average, approximately $\pm 15^\circ$ to $\pm 18^\circ$ either side of the Earth's magnetic equator. To show temporal variation of VTEC, the daily VTEC time series provided from the GIM maps for 130^{th} day of 2006 are shown in figure 4.4 for different latitudes in the Greenwich meridian. It is clearly seen that the VTEC pattern depends on latitudes.

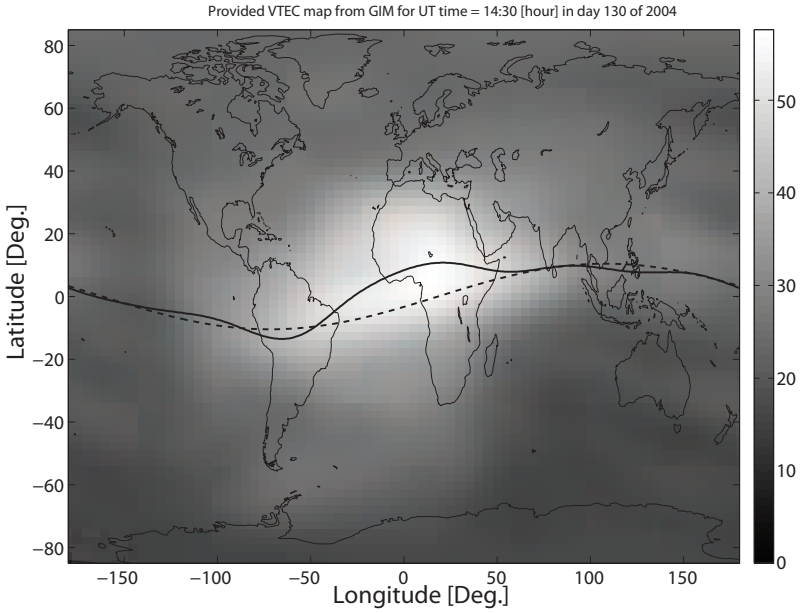

Figure 4.3: Global map of VTEC at 14:30 UT provided from GIM maps for day 130 of 2004.

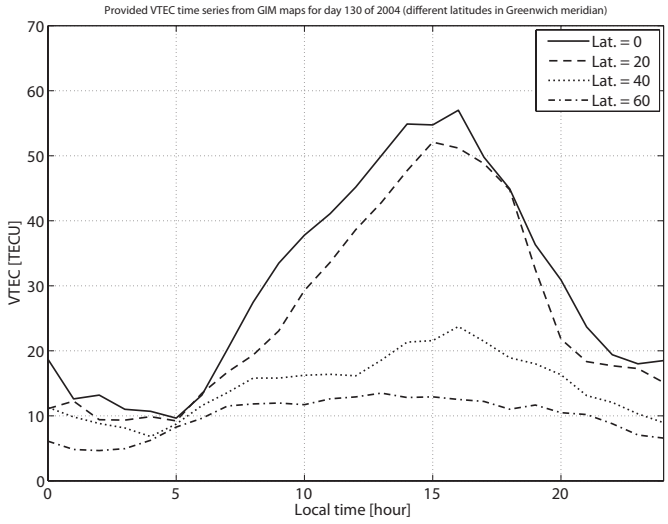

Figure 4.4: Daily VTEC time series (1 hour time interval) provided from GIM maps in day 130 of 2004 for different latitudes in the Greenwich meridian.

4.4.2 Approximation of the higher-order and bending effects

Using the ionospheric single-layer approximation, we may assumed that the product $B|\cos\theta|$ is *constant* along the signal path. As a result, it is allowed to take it *outside* the integral and write the second-order group delay in terms of TEC (Brunner and Gu, 1991)

$$\iota_{j,g}^{(2)} \approx \frac{eA}{f_j^3 2\pi m_e} B|\cos\theta|\, TEC \tag{4.35}$$

In equation 4.30, the third-order group delay is proportional to the integral of the squared electron density n_e^2 along signal path. Since the evaluation of this integral is difficult the integral is written in terms of TEC as follows (Hartmann and Leitinger, 1984),

$$\int n_e^2 \, d\rho = \Omega \; n_{e,max} \int n_e d\rho = \Omega \; n_{e,max} \; TEC \tag{4.36}$$

where $n_{e,max}$ denotes the maximum value of electron density along the signal path and Ω the so called *shape factor*. The shape factor Ω is as

$$\Omega = \frac{\int n_e^2 d\rho}{n_{e,max} \int n_e d\rho} \tag{4.37}$$

Using equation 4.36, the third-order group delay is obtained in terms of the TEC and $n_{e,max}$ and the shape factor,

$$\iota_{j,g}^{(3)} \; = \; \frac{3A^2}{8f_j^4} \Omega \; n_{e,max} \; TEC \tag{4.38}$$

To evaluate the ionospheric bending effect, the following simplified approximated formula (in terms of the TEC and the same shape factor) is used for the GPS signals

$$\kappa_j \; \approx \; \frac{A^2}{8f_j^4} (\tan \xi)^2 \Omega \; n_{e,max} \; TEC \tag{4.39}$$

Now the approximated second-, third- and bending effects have all been written as functions of the TEC. We can compute the magnitude of these effects. For this, a constant value $\Omega = 0.66$ can be used as an approximation for the shape factor.

The TEC heavily depends on the satellite zenith angle from receiver to the satellite. In (Odijk, 2002), the ionospheric effects were evaluated as function of the zenith angle for a worst case ionospheric condition on different frequencies. It has been shown that the first-order effect even on L1 frequency reaches several tens of meters. The bending and the second-order effects could not exceed a few centimeter and the third-order effect is in a few millimeter level. Furthermore, it is emphasized that the first-order ionospheric delay is of more importance for relative GNSS positioning than the higher-order and bending effects. According to (Odijk, 2002) the higher-order effects may be neglected in the case of relative GPS positioning with inter-receiver distances up to 400 km.

Note that the first-order ionospheric delay is inversely proportional to the square of the signal frequency. This allows to measure the first-order ionospheric effects using GNSS observations on different frequencies and accordingly the TEC can be provided from GNSS data. This is the subject of subsection 4.6.

4.5 Ionospheric models

Because of the complicated nature of the ionosphere, there have been numerous approaches for ionospheric modeling. In an effort to achieve simplicity, some of the models have been

restricted to certain altitude or latitude domains, while others have been restricted to certain ionospheric parameters, such as $NmF2$, $M(3000)F2$, and $VTEC$. These models describe the characteristic ionospheric features and their variations with time, season, solar cycle, and geomagnetic activity, as represented by Kp-index or Ap-index. The different types of the ionospheric models can be categorized as follows

- *Empirical* models are based on ionospheric measurements. Typically, the data are collected over an extended period of time and then fitted with simple analytical and empirical functions. The Klobuchar model is a simple ionospheric model that gives the vertical ionospheric delay at a given time and location for the GPS single-frequency users in real-time. Another example is the NeQuick model that proposed to use as ionospheric correction model for the future European Galileo positioning system. This model will be discussed in detail in the chapter 5.

- *Numerical* maps are in the global and regional scales that provide values of a ionospheric parameter in a grid. As an example, we can refer to the GIM that will be discussed in the following.

- *Analytical* models are based on orthogonal function fits to the output obtained from numerical models. As example, maps of $foF2$ and $M(3000)F2$ are presented by the CCIR (*Comite Consultatif International des Radiocommunications*) maps using the orthonormal and spherical forms of the Legendre functions. These maps will be discussed in chapter 5.

- *Physical* models are developed based on typically solution of the continuity equation or the momentum and energy equations for the electrons and ions. The solutions of equations provide for instance the electron density either as a function of altitude or along geomagnetic field lines. Such a physical modeling for the VTEC will be developed in chapter 6.

4.5.1 Klobuchar model

The Klobuchar model is one of the most widely used models due to its computational simplicity. The model is built on a simple *cosine* representation of the vertical ionospheric delay during daytime and a *constant offset* of 5 nanoseconds (or 1.5 meters) during night-time. Its daily maximum is at 14:00 h local time at the ionospheric pierce point. The period and amplitude of the model are represented by third-degree polynomials in local time and geomagnetic latitude at the ionospheric pierce point. The vertical ionospheric delay is computed as follows:

$$I^s_{v,r}(t_i) \;=\; \begin{cases} c\left[5\cdot 10^{-9} + \left(\sum_{l=0}^{3} \alpha_l \left[\varphi^s_{m,r}(t_i)'/\pi\right]^l\right) \;\cos x^s_r(t_i)\right], & |x^s_r(t_i)| < \frac{\pi}{2} \\ c\cdot 5\cdot 10^{-9} \;\approx\; 1.5 \text{ m}, & |x^s_r(t_i)| > \frac{\pi}{2} \end{cases} \tag{4.40}$$

with $x_r^s(t_i) = \frac{2\pi(t_r^s(i)' - 14\cdot 60^2)}{\sum_{l=0}^{3}\beta_l\left[\varphi_{m,r}^s(i)'/\pi\right]^l}$, and where:

$I_{v,r}^s(t_i)$	:	vertical ionospheric delay [m]
c	:	velocity of light [m/s]
$\varphi_{m,r}^s(t_i)'$	:	geomagnetic latitude of the ionospheric pierce point
$t_r^s(t_i)'$	:	local time at the ionospheric pierce point [s]
α_l $(l = 0, \ldots, 3)$	:	Klobuchar coefficients for amplitude
β_l $(l = 0, \ldots, 3)$	:	Klobuchar coefficients for period

Note that the local time at the ionospheric pierce point is computed from the *Universal Time* (UT) as:

$$t_r^s(t_i)' = t_{UT} + \frac{24 \cdot 60^2}{2\pi}\lambda_r^s(t_i)' \tag{4.41}$$

with t_{UT} the Universal Time in seconds and $\lambda_r^s(t_i)'$ the longitude of the pierce ionospheric point. The eight Klobuchar coefficients are transmitted in (the header of) the GPS satellite's *navigation message* and updated daily. To convert the vertical delay to the slant delay, the following mapping function is used (ICD-GPS-200C, 2004):

$$F(e) = 1 + 16 \times (0.53 - e)^3 \tag{4.42}$$

where e is the satellite elevation angle in semicircles. The model presents an ideal smooth behavior of the ionosphere; therefore any significant fluctuations will not be modeled properly. The accuracy of the model is 50∼60% of the total effect [klobuchar, 1987] and under severe ionosphere activity at low elevations, the range error could be very large, up to 50 meters.

4.5.2 Global Ionosphere Maps

The International GNSS Service (IGS) of the International Association for Geodesy (IAG) maintains a global GNSS tracking network of more than 200 receivers. The availability of a huge number of the IGS permanent dual frequency GPS receivers, which are distributed over the Earths surface, has made the global monitoring of the ionosphere possible and Global Ionospheric VTEC Maps (GIM) are produced routinely. Four IGS Ionospheric Analysis Centers (CODE, ESA, JPL, and UPC) individually produce daily Global Ionosphere Maps using different techniques (Perez, 2005; Wienia, 2008). The official IGS product is a combined ionosphere map from all four IGS analysis centers. The GIM files are distributed on a daily basis; they give a value for the Vertical Total Electron Content (VTEC) every two hours (0, 2, 4, ..., 24 UTC) at the grid points. The resolution of the maps is 5° in longitude and 2.5° in latitude. The GIM are provided in the IONEX (IONosphere map EXchange) format (Schaer 1999). The maps are provided in two versions: a rapid product with a latency of 12 hours, and a final product with a latency of 3 days.

The *Center for Orbit Determination in Europe* (CODE), for example, employs spherical harmonics functions to model the global VTEC using about 200 worldwide GPS/GLONASS stations. The spherical harmonics are then developed into Global Ionosphere Maps. The other analysis centers use different approaches. Only the CODE distributes GIM in three

versions: *Predicted*, *Rapid* and *Final* GIM. The predicted version contains predictions of the ionospheric delay for 1-2 days ahead. The VTEC at a certain time and location is interpolated between two consecutive maps using four surrounding grid points. The VTEC provided by GIM has a standard deviation of 2∼8 TECU (CODE, 2007) depending on the epoch in the solar cycle, season, and location and to the best of our knowledge is unbiased (1 TECU corresponds to an ionospheric delay of 16 cm in L1).

GIM Slant Ionospheric Delay The interpolated VTEC from the GIM is converted to the vertical ionospheric range delay V_{GIM} as

$$V_{GIM} = \frac{40.3}{f^2} \cdot VTEC_{GIM} \tag{4.43}$$

where f is the frequency of signal [Hz]. The standard deviation of the VTEC provided by GIM is 2∼8 TECU (CODE, 2007). In order to get slant ionospheric delay along the line-of-sight between receiver and satellite a mapping function must be used to map the VTEC to the slant TEC value. Mapping is done based on the single-layer approximation for the ionosphere and the mapping function is $F(\xi) = \frac{1}{\cos(\xi)}$ where ξ is the zenith angle of the satellite at the ionospheric pierce point at the ionospheric height of 350 km.

4.6 Slant ionospheric delay measurements from GNSS

In section 4.2, the mathematical expressions for the code and phase observables, equations 4.6 and 4.12, constitute the basis of a functional model for the processing of the observations. The slant ionospheric delay can be determined either from least-squares processing of a GNSS network data or from the geometry-free linear combination (single receiver data processing). It is noted that the biases induced by the instrumental phase and code delays are major problems for accurate determination of the ionospheric delay from GNSS data.

4.6.1 Network processing

In a reference GPS network where the coordinates of receivers are known, pseudo observables for receiver-satellite ranges ρ_r^s can be computed. Therefore the *geometry-free* (Non-positioning) model of observations can be processed for estimating the ionospheric delays, tropospheric delays, clock parameters and ambiguities. In case the coordinates of the receivers are unknown, the coordinates of receivers beside other parameters (ionospheric delays, clock parameters and ambiguities) should be estimated by processing the *geometry-based* (positioning) model of observations. Due to rank deficiencies in the design matrix (for both the geometry-free and geometry-based models), the estimated ionospheric delays are biased by the receiver and satellite clocks of the L1 phase observable, see (Odijk, 2002), but the relative accuracy of the estimated ionospheric delay is high. The biases are eliminated by double-differencing the estimates of ionospheric delays. In order to estimate unbiased absolute ionospheric delays, one can treat the ionospheric delays as stochastic variables that result in the *ionospheric-weighted* model. For this purpose, for each receiver-satellite combination, an ionospheric pseudo-observable is added to the vector of GNSS

observables in the models. Besides the extension of the vector of observables, the covariance matrix of the observables is extended to account for the assumed precision of the ionospheric pseudo-observables. Although the rank deficiency of the design matrix is removed by adding ionospheric pseudo-observables, the absolute accuracy of the ionospheric estimates depend on the quality of the ionospheric pseudo-observables.

An important advantage of the network processing is that the double-difference (DD) ambiguities are estimable. It is well-known that DD ambiguities are *integer-valued* due to elimination of the initial phases of the signal by the double-differencing. This integer nature of DD ambiguities can be exploited for the ambiguity fixing. Many approaches exist for estimating the integer ambiguities, however not all of them are admissible. In (Teunissen, 2002, 2003, 2004), general definition for the admissible integer estimator is given and the class of admissible integer estimators are discussed. The *integer least-squares* (ILS) estimator is an admissible integer estimator that fully takes the ambiguity vc-matrix into account. The ILS procedure is mechanized in the LAMBDA (Least-Squares AMBiguity Decorrelation Adjustment) method for ambiguity resolution; see (Jonge and Tiberius, 1996). For more detail about the theory of the ambiguity resolution and its validation, we refer to (Verhagen, 2005). Fixing the DD ambiguities makes the functional model stronger that result in a more precise estimation of the other non-integer parameters in both the geometry-free and geometry-based models, see (Teunissen, 1995).

4.6.2 Geometry-free linear combination

The slant ionospheric delay I_r^s can be also measured with a so-called *geometry-free linear combination* of the dual frequency GNSS code and carrier phase observations of a single receiver data. This approach is computationally more simple than the network processing. It is noted that the absolute accuracy of the computed ionospheric delays from both approaches are in the same level, but the relative accuracy (in space) of the computed ionospheric delay of the network processing is better than that of the Geometry-free linear combination. In the following the functional model for the Geometry-free linear combinations for phase and code observations at a multiple of frequencies, from receiver r to satellite s, is derived.

The mathematical model (known as the *Gauss-Markov model*), for phase and code observables in equations 4.18, reads for j frequencies at epoch t_i

$$E\{y(t_i)\} = \underbrace{\begin{bmatrix} e_j & -\mu & \mathbb{I}_j & \mathbb{I}_j & \mathbb{I}_j & 0 & 0 \\ e_j & \mu & 0 & 0 & 0 & \mathbb{I}_j & \mathbb{I}_j \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} g(t_i) \\ I_r^s(t_i) \\ M \\ \delta^s(t_i) \\ \delta_r(t_i) \\ d^s(t_i) \\ d_r(t_i) \end{bmatrix}}_{x} \quad ; \quad D\{y(t_i)\} = Q_{y(t_i)} \tag{4.44}$$

where $E\{\}$ and $D\{\}$ are respectively the mathematical expectation and dispersion oper-

ators and $y(t_i)$ vector of observations is set up as

$$y(t_i) = \begin{bmatrix} \Phi(t_i) \\ P(t_i) \end{bmatrix} \quad \text{with} \quad \begin{cases} \Phi(t_i) = \left[\varphi^s_{r,1}(t_i)\ ,\ \dots\ ,\ \varphi^s_{r,j}(t_i)\right]^T \\ P(t_i) = \left[p^s_{r,1}(t_i)\ ,\ \dots\ ,\ p^s_{r,j}(t_i)\right]^T \end{cases} \tag{4.45}$$

In the design matrix A, $e_j = [1, \dots, 1]^T$ is j-vector and $\mathbb{I}_j$ stands for $j \times j$ identity matrix and $\mu = [\mu_1, \dots, \mu_j]^T$. The vector of unknown parameters x includes: $g(t_i)$ as defined in equation 4.17, the slant ionospheric delay of code observable on the L1 frequency, j-vector of ambiguities $M = [M^s_{r,1}, \dots, M^s_{r,j}]^T$, j-vector of the instrumental phase delays in satellite $\delta^s(t_i) = [\delta^s_{,1}(t_i), \dots, \delta^s_{,j}(t_i)]^T$, j-vector of the instrumental phase delays in receiver $\delta_r(t_i) = [\delta_{r,1}(t_i), \dots, \delta_{r,j}(t_i)]^T$, j-vector of the instrumental code delays in satellite $d^s(t_i) = [d^s_{,1}(t_i), \dots, d^s_{,j}(t_i)]^T$ and j-vector of the instrumental code delays in receiver $d_r(t_i) = [d_{r,1}(t_i), \dots, d_{r,j}(t_i)]^T$.

In the stochastic model of the GNSS observables, $Q_{y(t_i)}$ is variance covariance matrix of the phase and code observables, assuming the phase and code observables are *normally distributed* and uncorrelated,

$$Q_{y(t_i)} = \mathrm{blkdiag}[Q_\varphi\ ,\ Q_p] \otimes Q(t_i) \tag{4.46}$$

where 'blkdiag' stands for a *block-diagonal* matrix, Q_φ and Q_p denote the parts specific for respectively the phase and code observables (*cofactor matrices*) and $Q(t_i)$ the part which is similar for each observable type. For a compact notation, the matrix *Kronecker product* (denoted by the symbol $\otimes$) has been used. Through $Q(t_i)$ possible *satellite-elevation dependency* can be taken into account. When no elevation-dependent function is specified, it is assumed to be an identity matrix. In this case, the cofactor matrices Q_φ and Q_p may be referred to as phase and code variance covariance matrices.

Geometry-free observable By subtracting the phase (or code) observable of j^{th} frequency ($j > 1$) from that of the first frequency, both expressed in units of length, the geometry-dependent term $g(t_i)$ is eliminated and resulting

$$\begin{aligned} E\left\{\phi^s_{r,1j}(t_i)\right\} &= -\left(1-\mu_j\right) I^s_r(t_i) + DPB^s_{,1j}(t_i) + DPB_{r,1j}(t_i) + M^s_{r,1j} \\ E\left\{p^s_{r,1j}(t_i)\right\} &= \quad \left(1-\mu_j\right) I^s_r(t_i) + DCB^s_{,1j}(t_i) + DCB_{r,1j}(t_i) \end{aligned} \tag{4.47}$$

with $\phi^s_{r,1j}(t_i) = \phi^s_{r,1}(t_i) - \phi^s_{r,j}(t_i)$ and $p^s_{r,1j}(t_i) = p^s_{r,1}(t_i) - p^s_{r,j}(t_i)$ respectively the phase and code *geometry-free* linear combinations, $M^s_{r,1j} = M^s_{r,1} - M^s_{r,j}$ ambiguity of the phase geometry-free linear combination, $DPB_{r,1j}(t_i) = \delta_{r,1}(t_i) - \delta_{r,j}(t_i)$ and $DPB^s_{,1j}(t_i) = \delta^s_{,1}(t_i) - \delta^s_{,j}(t_i)$ respectively receiver and satellite *differential phase biases*, $DCB_{r,1j}(t_i) = d_{r,1}(t_i) - d_{r,j}(t_i)$ and $DCB^s_{,1j}(t_i) = d^s_{,1}(t_i) - d^s_{,j}(t_i)$ respectively receiver and satellite *differential code biases*.

The mathematical model of the phase and code observables, equation 4.44, can be transformed to mathematical model of the geometry-free linear combination with the following transformation matrix

$$D = \mathbb{I}_2 \otimes D_{j-1} \qquad \text{with} \qquad D_{j-1} = [e_{j-1}\ ,\ -\mathbb{I}_{j-1}] \tag{4.48}$$

After transformation, the observable vector of the geometry-free linear combination (specified by the superscript gf) reads as follows

$$y^{gf'}(t_i) = Dy(t_i) = \begin{bmatrix} \Phi^{gf}(t_i) \\ P^{gf}(t_i) \end{bmatrix} \quad \text{with} \quad \begin{cases} \Phi^{gf}(t_i) = \left[\varphi^s_{r,12}(t_i)\ ,\ \dots\ ,\ \varphi^s_{r,1j}(t_i)\right]^T \\ P^{gf}(t_i) = \left[p^s_{r,12}(t_i)\ ,\ \dots\ ,\ p^s_{r,1j}(t_i)\right]^T \end{cases} \tag{4.49}$$

The variance covariance (vc-) matrix of $y^{gf'}(t_i)$ is obtained by applying the propagation law of variances, i.e. $Q_{y^{gf'}(t_i)} = DQ_{y(t_i)}D^T$. When $Q(t_i) = \mathbb{I}_j$ is an identity j-matrix and the stochastic model of observables is assumed as $Q_\varphi = \sigma^2_\varphi\, \mathbb{I}_j$ and $Q_p = \sigma^2_p\, \mathbb{I}_j$ with σ^2_φ and σ^2_p the a priori variance factors of the phase and code observables respectively, the variance covariance matrix of $y^{gf'}(t_i)$ reads

$$Q_{y^{gf'}(t_i)} = \mathrm{blkdiag}[\sigma^2_\varphi, \sigma^2_p] \otimes C \qquad \text{with} \qquad C = \underbrace{\begin{bmatrix} 2 & 1 & \cdots & 1 \\ 1 & 2 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 2 \end{bmatrix}}_{(j-1)\times(j-1)} \tag{4.50}$$

From this vc-matrix it can be seen that the variances of the geometry-free observables increase by factor 2 and are correlated.

Rank-defect functional model Due to the transformation the design matrix becomes $A^{gf'} = DA$ and the functional model of geometry-free observables reads as follows (apostrophe $(')$ emphasize the rank deficiency of the functional model)

$$E\{y^{gf'}(t_i)\} = \underbrace{\begin{bmatrix} -D_{j-1}\mu & \mathbb{I}_{j-1} & \mathbb{I}_{j-1} & \mathbb{I}_{j-1} & 0 & 0 \\ D_{j-1}\mu & 0 & 0 & 0 & \mathbb{I}_{j-1} & \mathbb{I}_{j-1} \end{bmatrix}}_{A^{gf'}} \underbrace{\begin{bmatrix} I^s_r(t_i) \\ M^{gf} \\ DPB^s(t_i) \\ DPB_r(t_i) \\ DCB^s(t_i) \\ DCB_r(t_i) \end{bmatrix}}_{x'} \tag{4.51}$$

with $M^{gf} = [M^s_{r,12}, \dots, M^s_{r,1j}]^T$ vector of geometry-free ambiguities, $DPB^s(t_i) = [DPB^s_{,12}(t_i), \dots, DPB^s_{,1j}(t_i)]^T$ vector of the satellite differential phase biases and $DPB_r(t_i) = [DCB_{r,12}(t_i), \dots, DCB_{r,1j}(t_i)]^T$ vector of the receiver differential phase biases, $DCB^s(t_i) = [DCB^s_{,12}(t_i), \dots, DCB^s_{,1j}(t_i)]^T$ vector of the satellite differential code biases and $DCB_r(t_i) = [DCB_{r,12}(t_i), \dots, DCB_{r,1j}(t_i)]^T$ vector of the receiver differential code biases.

The functional model in equation 4.51 is an *underdetermined system* of equations (the number of observation is less than number of unknown parameters) and the design matrix $A^{gf'}$ is *rank-defect*. It implies that it is not possible to estimate separately all of the unknowns. To solve the mathematical model uniquely, the rank deficiencies in the columns of the design matrix must be eliminated. The design matrix $A^{gf'}$ has a rank deficiency of

$3j-2$ where $j-1$ rank-defects are for the columns associated to the receiver and satellite differential code biases, $2(j-1)$ rank-defects for the columns associated to the receiver and satellite differential phase biases and ambiguities.

The satellite differential code biases DCB^s and the receiver differential code biases DCB_r for some IGS stations are available in IONEX files that can be downloaded from the Center for Orbit Determination in Europe (CODE). Using these IONEX files, corrections for the differential code biases of satellites can be applied. However, unless the user's receiver is an IGS station, the receiver differential code bias (DCB_r) is not in the IONEX files and needs to be computed by the user. Therefore, using $DCB^s(t_i)$ from the IONEX files, the $DCB^s(t_i)$ can be fixed in the mathematical model that results in the $j-1$ columns in the design matrix, associated to DCB^s, are eliminated. It means the code observables are corrected by replacing the code geometry-free observables $P^{gf}(t_i)$ with the corrected data $(P^{gf}(t_i) - DCB^s(t_i))$ in the vector of observables $y^{gf'}(t_i)$.

The differential phase biases for satellite and receiver, i.e. DPB^s and DPB_r, are time dependent. Their absolute values along with their variations in time does not exceed a few centimeters (Jong, 1998). Since these parameters cannot be separately estimated from the data, in the subsequent analysis they will be dropped from the model. This will however introduce a cm-level bias in the estimates of the other parameters.

Then the design matrix will only have one rank deficiency left. The way to circumvent this rank deficiency is to treat the ionospheric delay $I_r^s(t_i)$ as a *stochastic variable*. It means that an *ionospheric pseudo-observable* is added to the vector of GPS observables. Beside the extension of the vector of observables, the vc-matrix of the GNSS observables is extended to account for the assumed precision of the ionospheric pseudo-observable. The ionospheric pseudo-observable may originate from different sources. In this thesis, the GIM is used to compute the ionospheric pseudo-observable (denoted by $I^s_{r,GIM}(t_i)$).

Note that the rank deficiency also can be eliminated by a time differencing of the geometry-free observables at two consecutive epochs that results in a time-differenced ionospheric delay. Since the time-differenced ionospheric delay can be estimated only from the phase geometry-free observables, it is very precise relatively in time. In this thesis, we need to measure the absolute ionospheric delay from the GPS data because of this it is necessary to remove the rank deficiency of the model 4.51 without time differencing.

Full-Rank functional model After fixing DCB^s, dropping DPB_r and DPB^s and adding the ionospheric pseudo-observable, the full-rank functional model of the geometry-free observables is obtained

$$E\{y^{gf}(t_i)\} = \underbrace{\begin{bmatrix} -D_{j-1}\mu & \mathbb{I}_{j-1} & 0 \\ D_{j-1}\mu & 0 & \mathbb{I}_{j-1} \\ 1 & 0 & 0 \end{bmatrix}}_{A^{gf}} \underbrace{\begin{bmatrix} I_r^s(t_i) \\ M^{gf} \\ DCB_r(t_i) \end{bmatrix}}_{x} \tag{4.52}$$

with

$$y^{gf}(t_i) = \begin{bmatrix} \Phi^{gf}(t_i) \\ P^{gf}(t_i) \\ I^s_{r,GIM}(t_i) \end{bmatrix} \quad \text{with} \quad Q_{y^{gf}(t_i)} = \begin{bmatrix} \sigma_\varphi^2 \cdot C & & 0 \\ & \sigma_p^2 \cdot C & \\ 0 & & \sigma^2_{I_{GIM}} \end{bmatrix} \tag{4.53}$$

where $\sigma_{I_{GIM}}$ stands for the standard deviation of ionospheric pseudo-observable. The standard deviation of the ionospheric pseudo-observable can be computed by $\sigma_{I_{GIM}} = 3\sigma_{V_{GIM}}$ with $\sigma_{V_{GIM}}$ as the standard deviation of the VTEC computed from GIM. Since the variation of the receiver differential code bias (DCB) during several hours does not change significantly, it is assumed that the vector of DCB_r is constant in time. Then the parameter vector x can be partitioned into two parts ($x = \left[x_1^T \ , \ x_2^T\right]^T$): a part for the time-varying or temporal parameters (denoted as x_1), a part for parameters which are assumed to be constant in time (denoted as x_2). The temporal part consists only of the ionospheric delay $I_r^s(t_i)$. The vector of ambiguities M^{gf} (if there is no *cycle slip*) and the vector of receiver differential code biases DCB_r form the non-temporal part x_2. As a consequence, the design matrix A^{gf} is partitioned as follows

$$A^{gf} = [A_1 \ , \ A_2] \quad \text{with} \quad A_1 = \begin{bmatrix} -D_{_{j-1}}\mu \\ D_{_{j-1}}\mu \\ 1 \end{bmatrix} \ ; \ A_2 = \begin{bmatrix} \mathbb{I}_{j-1} & 0 \\ 0 & \mathbb{I}_{j-1} \\ 0 & 0 \end{bmatrix} \tag{4.54}$$

Least squares solution In case of single-epoch processing all parameters are estimated, but the *redundancy* of the functional model is zero. Since the vector of parameters includes the non-temporal parameters x_2 adding new epoch data into the vector of observables increases the redundancy of the functional model for which the least squares solution of the model is estimated precisely.

When k epochs are involved in the mathematical model (in a time span without cycle slip) then the observable vector becomes $Y^{gf} = [y^{gf}(t_1), \ldots, y^{gf}(t_k)]$ that assumes the epochs are not correlated. The vc-matrix, for k epochs, reads $Q_{Y^{gf}} = \mathrm{blkdiag}[Q_{y^{gf}(t_1)}, \ldots, Q_{y^{gf}(t_k)}]$. The temporal parameter vector is also parameterized for k epochs, $x_1 = [I_r^s(t_1), \ldots, I_r^s(t_k)]$. The design matrix for k epochs is given as

$$A^{gf} = [\mathbb{I}_k \otimes A_1 \ , \ e_k \otimes A_2] \tag{4.55}$$

which is a full-rank matrix. The least squares solution of the mathematical model is

$$\hat{x} = \left(A^{gf^T} Q_{Y^{gf}}^{-1} A^{gf}\right)^{-1} A^{gf^T} Q_{Y^{gf}}^{-1} Y^{gf} \quad \text{with} \quad Q_{\hat{x}} = \left(A^{gf^T} Q_{Y^{gf}}^{-1} A^{gf}\right)^{-1} \tag{4.56}$$

The redundancy of the model is as a function of the number of observation epochs k and the number of frequency j (i.e. $df = 2(k-1)(j-1)$) which in the case of a single-epoch processing $df = 0$.

Analyzing the (formal) precision In order to evaluate the vc-matrix of the parameters, the model's *normal equations* can be partitioned for the temporal and non-temporal parameters as follows

$$Q_{\hat{x}} = \begin{bmatrix} Q_{\hat{x}_1} & Q_{\hat{x}_1\hat{x}_2} \\ Q_{\hat{x}_2\hat{x}_1} & Q_{\hat{x}_2} \end{bmatrix} = \begin{bmatrix} \mathbb{I}_k \otimes (A_1^T Q_{y^{gf}}^{-1} A_1) & e_k \otimes (A_1^T Q_{y^{gf}}^{-1} A_2) \\ e_k \otimes (A_2^T Q_{y^{gf}}^{-1} A_1) & k\ (A_2^T Q_{y^{gf}}^{-1} A_2) \end{bmatrix}^{-1} \tag{4.57}$$

which it is assumed $Q_{y^{gf}} = Q_{y^{gf}}(t_1) = \cdots = Q_{y^{gf}}(t_k)$. Using *Full block-partitioning* of an invertible matrix (see (Teunissen et al., 2005)), the vc-matrices for the temporal and

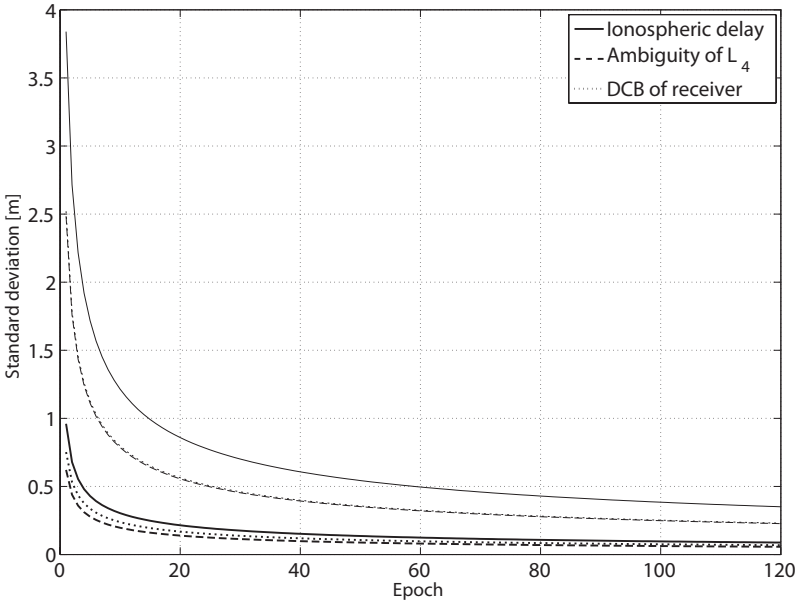

Figure 4.5: Standard deviation of the estimator of slant ionospheric delay and ambiguity and DCB of receiver in the model of geometry-free linear combination observables versus number of epochs for $\sigma_\phi = 3\ mm$ and $\sigma_p = 300\ mm$, the bold curves are for the case of $\sigma_{V_{GIM}} = 2$ [TECU] and the thin curves are for the case of $\sigma_{V_{GIM}} = 8$ [TECU].

non-temporal parameters are given

$$Q_{\hat{x}_1} = \left[\mathbb{I}_k \otimes (A_1^T Q_{y^{gf}}^{-1} A_1) - \tfrac{1}{k}(e_k\ e_k^T) \otimes (A_1^T Q_{y^{gf}}^{-1} A_2)(A_2^T Q_{y^{gf}}^{-1} A_2)^{-1}(A_2^T Q_{y^{gf}}^{-1} A_1)\right]^{-1}$$

$$Q_{\hat{x}_2} = \tfrac{1}{k}\left[(A_2^T Q_{y^{gf}}^{-1} A_2) - (A_2^T Q_{y^{gf}}^{-1} A_1)(A_1^T Q_{y^{gf}}^{-1} A_1)^{-1}(A_1^T Q_{y^{gf}}^{-1} A_2)\right]^{-1} \tag{4.58}$$

where diagonal arrays of k-matrix $Q_{\hat{x}_1}$ are the same and give the variance of the estimated ionospheric delay $\sigma_{\tilde{I}}^2$. In this thesis the estimated ionospheric delay is denoted as $\tilde{I}$ that a 'tilde' symbol emphasizes the measuring of slant ionospheric delay from the geometry-free linear combination. The first set of $(j-1)$ arrays in the diagonal of $2(j-1)$-matrix $Q_{\hat{x}_2}$ give the variances of the estimated ambiguity vector M^{gf} and the second set of $(j-1)$ arrays in the diagonal give the variance of the estimated receiver DCB.

It is possible to evaluate the (formal) vc-matrices of the parameters estimators without the use of actual GPS observations, since vc-matrices $Q_{\hat{x}_1}$ and $Q_{\hat{x}_2}$ only depend on the design matrix and vc-matrix of the observations.

In order to quantify the standard deviation of the parameters estimators, in equation 4.53, the square roots of a priori variance factors for phase and code observables are considered as $\sigma_{\phi,1} = 3$ mm, $\sigma_{p,1} = 300$ mm and for the standard deviation of the computed VTEC from GIM, $\sigma_{V_{GIM}}$, two different values 2 and 8 TECU (corresponding to 0.32 and 1.28 m) are used. To demonstrate the effect of number of epochs on quality of the vector of unknown parameters in equation 4.58, for the dual-frequency data $(j = 2)$, the standard

deviations of the estimated slant ionospheric delay, geometry-free ambiguity and receiver DCB are plotted as functions of the number of epochs in figure 4.5.

As expected an increase in the number of epochs results in a decrease in the standard deviations of the parameters estimators. Note that the standard deviations of the estimators do not seem to go to zero when increasing the number of epochs. In fact it seems that the standard deviations converge to a non-zero constant value which depends on the standard deviation of the ionospheric pseudo-observable $\sigma_{I_{GIM}}$. It implies that the computed slant ionospheric delay I_{GIM} from GIM and the estimated $\tilde{I}$ value from the geometry-free linear combination are correlated. It can be seen from the graphs that the standard deviation of the ionospheric delay becomes better than 1 TECU (a few decimeter) when the number of observation epochs are more than 10, even in the case of $\sigma_{V_{GIM}} = 8$ [TECU].

The standard deviations for the parameters estimator are computed for the case of $j = 3$ and compared with the standard deviations computed for the case of $j = 2$. The result was that the standard deviations are comparable for both cases. It is concluded that an increase in the number of frequencies does not improve significantly the standard deviations of the estimators in the model of geometry-free observables.

4.7 Summary

This chapter has focused on the ionospheric refraction error in the GNSS observations as consequence of the signal's propagation through the Earth's atmosphere. The ionospheric error consists of the first-, second- and third-order effects and an effect due to the bending of the signal. These effects are all function of the ionospheric total electron content (TEC) along the signal path between receiver and satellite. The mathematical model for extracting the ionospheric error (or TEC) using the geometry-free linear combinations of multi-frequency GNSS code and phase data has been set up. Besides the estimation of the ionospheric error, the receiver differential code bias (DCB) is also estimated from the least squares solution of the mathematical model. It has been found out that an increase in the number of GNSS frequencies does not improve significantly the standard deviation of the estimator of the ionospheric parameter in the model. However, the model will be used to estimate the ionospheric TEC data from the GNSS observations in the next chapters of this thesis.

NeQuick 3D Ionospheric Electron Density Profiler

5

NeQuick is a semi-empirical model that describes spatial and temporal variations of the ionospheric electron density. This chapter focus on the formulation of the NeQuick model and its performances under different ionospheric conditions. For comparison, the performance of the *Global Ionospheric Maps* (GIM), and the well known Klobuchar model will be included. In real-time application of NeQuick for the future Galileo navigation system, the solar daily activity needs to be introduced into the model by the effective ionization level (called Az parameter). As the Galileo system is under development, daily values of the Az parameter are not yet available, and therefore we present an alternative approach for estimation of the Az parameter using global ionospheric maps from IGS.

5.1 Ionospheric electron density model NeQuick

A family of *semi-empirical* 3D ionospheric electron concentration models recently developed by Abdus Salam ICTP (International Center of Theoretical Physics) in Trieste (Italy) and the Institute for Meteorology and Geophysics in Graz (Austria) consists of the three models *NeQuick*, *COSTprof*, and *NeUoG-plas*. These models are not *purely empirical* because they have been developed by taking into account the physical properties of the different ionospheric regions. In all three models, the ionosphere is vertically divided into two parts: a *lower part* (below the peak of the $F2$ region) and an *upper part* (above the peak of the $F2$ region). For the lower part of the ionosphere, the models have the same formulation based on the original DGR model (Giovanni and Radicella, 1990) that uses a sum of *Epstein* layers; the formulation is such that the model and its first derivative are always continuous (Radicella and Leitinger, 2001). Due to differing usages of the models, their formulations are different for the upper part of the ionosphere; only COSTprof and NeUoG-plas models take the *plasmasphere* into account, and they have different formulations for this uppermost region of the atmosphere (above 2000 km) (Hochegger et al., 2000).

All of the models are able to create electron density profiles, but among them NeQuick is a flexible *quick-run* model that has been proposed for single-frequency operation in the ESA *Galileo* project to compute the slant total electron content along arbitrary ground-to-satellite ray paths (Radicella et al., 2003).

NeQuick model NeQuick is an ionospheric *profiler* which calculates the electron density at a given time and location in the Earth's ionosphere. The NeQuick model uses the peaks of the three main ionospheric regions (E, $F1$, and $F2$ regions) as *anchor points* (Leitinger et al., 2005; Radicella and Leitinger, 2001) and the electron density at any location is computed based on the characteristic parameters (*peak electron density*, *peak height*) of the

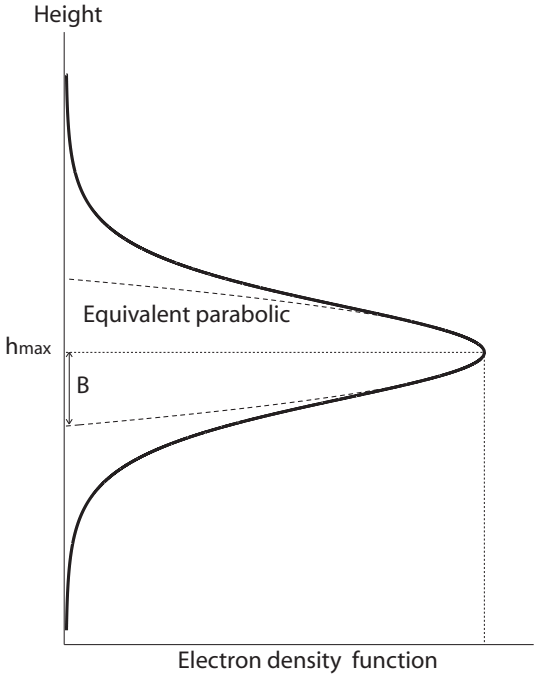

Figure 5.1: Typical shape of the electron density function of an Epstein layer

anchor points. NeQuick (Fortran 77) source code is only available in its 1994 version at *http://www.itu.int/ITU-R/software/study-groups/rsg3/databanks/ionosph/*. In the following, the formulation of the NeQuick model that the available source code uses is explained. The differences between the NeQuick version of 1994 and an recently improved version will be addressed. In general, different formulations are used by NeQuick to compute the electron density at a point located on the bottom side ($h \leq h_{max,F2}$) or the top side ($h \geq h_{max,F2}$) of the $F2$ peak electron density ($N_{max,F2}$), see figure 5.3. The formulation of NeQuick is based on Epstein layers; therefore to aid understanding, first the Epstein layer is explained.

Epstein layer In an Epstein layer, as introduced by (Rawer, 1983), an electron density function is built on Epstein functions and represented by the following analytical expression (Radicella and Zhang, 1995)

$$n(h) = \frac{4\, n_{\max}}{\left(1 + \exp\left(\frac{h - h_{\max}}{B}\right)\right)^2} \exp\left(\frac{h - h_{\max}}{B}\right) \tag{5.1}$$

where $n(h)$ is the electron density at height h, n_{max} is the peak electron density, h_{max} is the height of peak electron density, and B is called the *thickness* of the layer (see figure 5.1). The electron density function of an Epstein layer is a symmetric function and has a parabolic form around the height of its maximum electron density. Its derivative with respect to height is a continuous function.

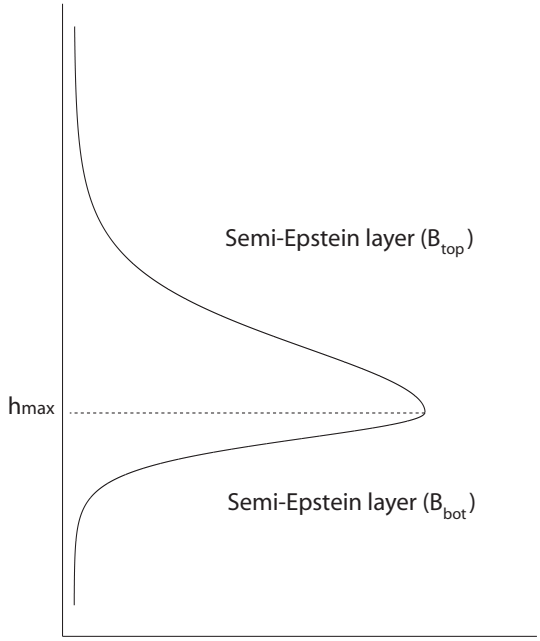

Figure 5.2: An asymmetric Epstein layer with two different thickness values (B_{top} and B_{bot}) for the top and bottom parts

5.1.1 NeQuick model formulation for the bottom side ($h \leq h_{max,F2}$)

In order to obtain better results from NeQuick, different values were considered for the thicknesses of the bottom and top parts of the Epstein layer associated to each ionospheric region (E, $F1$, and $F2$). This means that for each region two *semi-Epstein* layers that differ in their thickness values are used, and it results in an asymmetric Epstein layer (see figure 5.2).

Accordingly, the formulation of NeQuick for the bottom side of the ionosphere ($h \leq h_{max,F2}$) is based on five semi-Epstein layers. Two semi-Epstein layers are used for the E region (top and bottom), two for the $F1$ region (also top and bottom) and one for the bottom of the $F2$ region. It must be noted that NeQuick does not take any characteristic parameters of the D region into account.

Since the concept of well separated ionospheric regions cannot be maintained and their overlap must be admitted, the electron density of a given point with height less than the height of peak electron density of the $F2$ region (bottom side) is computed by summation of three semi-Epstein layers (Hochegger et al., 2000):

$$n(h) = \sum_{i=F2,F1,E} n_i(h) \tag{5.2}$$

where

$$n_i(h) = \frac{4\,A_i}{(1+\exp(z_i))^2}\exp(z_i) \tag{5.3}$$

The quantities z_{F2} and A_{F2}, for the semi-Epstein layer associated to the bottom of $F2$ region, are defined by formulas in Table 5.1.

Table 5.1: Definitions of the quantities z_{F2} and A_{F2}

Height interval [km]	z_{F2}	A_{F2}
$h_{max,F1} \leq h \leq h_{max,F2}$ $h_{max,E} \leq h \leq h_{max,F1}$ $100 \leq h \leq h_{max,E}$ $h \leq 100$	$\frac{h-h_{\max,F2}}{B_{bot,F2}}$	$n_{max,F2}$

Table 5.2: Definitions of the quantities z_{F1} and A_{F1}

Height interval [km]	B	z_{F1}	A_{F1}
$h_{max,F1} \leq h \leq h_{max,F2}$	$B_{top,F1}$	$\frac{h-h_{\max,F1}}{B} \exp\left\{\frac{10}{1+2\lvert h-h_{\max,F2}\rvert}\right\}$	$n_{\max,F1} - n_{F2}(h_{\max,F1})$
$h_{max,E} \leq h \leq h_{max,F1}$ $100 \leq h \leq h_{max,E}$ $h \leq 100$	$B_{bot,F1}$		

Definitions of the quantities z_{F1} and A_{F1} are given in Table 5.2. Note that in the definition of z_{F1}, the thickness parameter takes different values for the bottom side and top side of the semi-Epstein layers of the $F1$ region. Table 5.3 gives the definitions of the quantities z_E and A_E. In the definition of z_E, the thickness parameter takes different values for the bottom side and top side of the semi-Epstein layers of the E region.

Note that the semi-Epstein layers for the top sides of the E and $F1$ layers are lowered to zero at the peak of the $F2$ region due to the exponential factor in the formulation of z_E and z_{F1} (see figure 5.3). This is done to avoid secondary maxima in the vertical profile of electron density.

5.1.2 NeQuick model formulation for the top side ($h_{max,F2} \leq h$)

In the NeQuick model, the top side $F2$ region is a semi-Epstein layer but with a height-dependent thickness parameter. The electron density in the top side of the $F2$ region is computed by (Coisson et al., 2006):

$$n_{topside}(h) = \frac{4\, n_{\max,F2}}{(1+\exp(z))^2} \exp(z) \tag{5.4}$$

where z is defined by following formulae:

$$z = \frac{h - h_{\max,F2}}{H\left[1 + \frac{rg(h-h_{\max,F2})}{rH+g(h_{\max,F2})}\right]} \tag{5.5}$$

The denominator is a height-dependent thickness parameter of the top side of the $F2$ region and H acts as the scale height that is given by:

$$H = \frac{B_{top,F2}}{v} \tag{5.6}$$

$$v = (0.041163x - 0.183981)x + 1.424472 \tag{5.7}$$

Table 5.3: Definitions of the quantities z_E and A_E

Height interval [km]	B	z_E	A_E
$h_{max,F1} \leq h \leq h_{max,F2}$ $h_{max,E} \leq h \leq h_{max,F1}$	$B_{top,E}$	$\frac{h-h_{\max,E}}{B} \exp\left\{\frac{10}{1+2\lvert h-h_{\max,F2}\rvert}\right\}$	$n_{\max,E} - n_{F1}(h_{\max,E})$ $-n_{F2}(h_{\max,E})$
$100 \leq h \leq h_{max,E}$ $h \leq 100$	$B_{bot}E$		

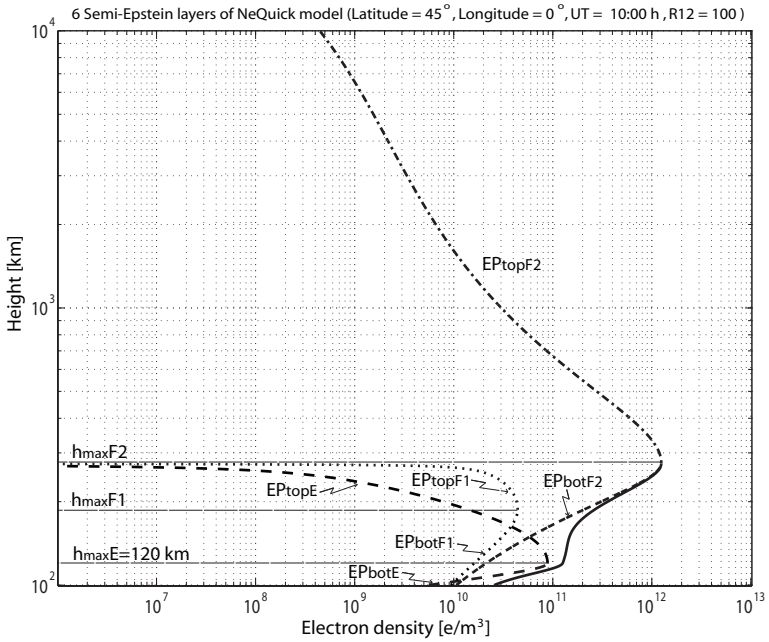

Figure 5.3: Six semi-Epstein layers used in NeQuick for 3 anchor points (peak of $F2$, $F1$, and E regions). The solid curve below the $F2$ region peak height is the NeQuick electron density profile for $h \leq h_{max,F2}$

$$x = \frac{B_{top,F2} - 150}{100} \tag{5.8}$$

In equations 5.5, the factor r is 100 and $g = 0.125$ is a height gradient for the scale height H and $rH + g(h_{max,F2})$ restricts the scale height increase in the top side of the $F2$ region (Leitinger et al., 2005).

5.2 Characteristic parameters of the anchor points

As seen in previous sections, for computation of the electron density at a given point in the top side or bottom side of the ionosphere, the characteristic parameters—peak electron densities and peak heights—of the three anchor points are needed. In NeQuick, the peak electron densities of the E, $F1$, and $F2$ regions are computed in electrons per cubic meter using the equation $n_{\max}X = 1.24 \cdot 10^{10} \cdot (foX)^2$ in which X stands for E, $F1$, or $F2$ (the inverse function of $f_p \simeq 8.98\sqrt{n}$ in section 3.4.6) when foX is in MHz (Rush, 1986). The peak height of the E region is fixed at $h_{max,E} = 120$ [km] but the peak heights of the $F1$ and $F2$ regions depend on the peak electron densities of the regions, will be explained next.

5.2.1 Peak height of the $F2$ region

The peak height of the $F2$ region is determined with the very widely used empirical formulation by Shimazaki 1995 in the form of Dudeney (Dudeney, 1983) based on the ionosonde parameters foE, $foF2$ and $M(3000)F2$ (symbol M)

$$h_{\max,F2} = \frac{1490\,\mu}{M+\Delta} - 176_{km} \tag{5.9}$$

where the quantities μ and Δ are defined as follows:

$$\mu = M\sqrt{\frac{0.0196\,M^2+1}{1.2967\,M^2-1}} \tag{5.10}$$

$$\Delta = \left\{ \begin{array}{ll} \frac{0.253}{\frac{foF2}{foE}-1.215} - 0.012 & ; \quad \text{if} \quad foE \neq 0 \\ -0.012 & ; \quad \text{if} \quad foE = 0 \end{array} \right\} \tag{5.11}$$

5.2.2 Thickness parameters of the semi-Epstein layers

The thickness parameters take different values for the bottom and top sides of each anchor point. The thickness parameter for the bottom-side semi-Epstein layer of the E region is fixed at $B_{bot,E} = 5$ [km]. The semi-Epstein layer of the top side of the E region takes the fixed value $B_{top,E} = 7$ [km] during the night (if $F1$ not present) but in the daytime (when $F1$ present) it is assigned $B_{top,E} = 0.5B_{top,F1}$ in [km].

For the bottom-side semi-Epstein layer of the $F1$ region, the thickness parameter is $B_{bot,F1} = 0.7B_{top,F1}$. The thickness value for the top side of $F1$ region is determined by

$$B_{top,F1} = \frac{h_{\max,F2} - h_{\max,F1}}{\ln\left(\frac{4\left(n_{\max,F1} - n_{F2}(h_{\max,F1})\right)}{0.1N_{\max,F1}}\right)} \tag{5.12}$$

In the semi-Epstein layer of the bottom side $F2$ region, the thickness parameter depends on the peak electron density $n_{max,F2}$ and the maximum vertical gradient of electron density $\max(dn/dh)$ in the ionosphere. The following equation is used for the thickness parameter of the bottom side of the $F2$ region:

$$B_{bot,F2} = \frac{0.385\,n_{\max,F2}}{\max\left(\frac{dn}{dh}\right)} \tag{5.13}$$

where the maximum electron density gradient with height is determined from two ionosonde parameters $foF2$ and $M(3000)F2$:

$$\max\left(\frac{dn}{dh}\right) = 0.01\exp\left\{-3.467 + 0.857\ln\left(foF2^2\right) + 2.02\ln\left(M\right)\right\} \tag{5.14}$$

The top-side thickness parameter of the $F2$ region is calculated by $B_{top,F2} = kB_{bot,F2}$ where the parameter k takes different formulations for April to September and for October

to March, and is always restricted to $2 \leq k \leq 8$.

$$k = \begin{cases} 6.705 - 0.014\,R_{12} - 0.008 h_{\max,F2} & \text{(April to September)} \\ -7.77 + 0.097\left(\frac{h_{\max,F2}}{B_{bot,F2}}\right)^2 + 0.153 n_{\max,F2} & \text{(October to March)} \end{cases} \tag{5.15}$$

where R_{12} is the sunspot number (see section 2.2).

5.3 Providing the ionosonde parameters for NeQuick

Previous sections show that the ionosonde parameters play significant roles in the formulation of NeQuick not only for providing the characteristic parameters of the anchor points but also for computing the thickness parameters of the semi-Epstein layers. Hence, to compute the NeQuick electron density at a given location and time, the ionosonde parameters (foE, $foF1$, $foF2$ and $M(3000)F2$) are needed.

In NeQuick, foE is computed from an empirical model developed by John Titheridge that depends on the solar zenith angle

$$(foE)^2 = \sqrt{F10.7}\ \left((1.112 - 0.019\ \epsilon)\cos\chi_{_{eff}}\right)^2 + 0.49 \tag{5.16}$$

with $\epsilon = m\ \frac{\exp(0.3\ \phi)-1}{\exp(0.3\ \phi)+1}$ as a seasonal factor where m set to -1 for November, December, January, and February and set to 1 for May, June, July, and August and elsewhere set to zero, $\chi_{_{eff}} = \chi$ when $\chi \leq 86.23°$ and $\chi_{_{eff}} = 90° - 0.24°\exp(20° - 0.2\chi)$ when $\chi > 86.23°$.

The $F1$ region critical transition $foF1$ is taken to be proportional to foE during daytime $foF1 = 1.4foE$ and 0 at night. For $foF2$ and $M(3000)F2$ the *CCIR* (*Comite Consultatif International des Radiocommunications*) maps of ITU-R (*International Telecommunication Union*) in the mode used by the *International Reference Ionosphere* (IRI) are used. These maps are explained in the following subsections.

5.3.1 CCIR maps of $foF2$ and $M(3000)F2$

The most common set of global maps for both parameters $foF2$ and $M(3000)F2$ are the CCIR maps that have been prepared using the monthly median values of $foF2$ and $M(3000)F2$ from all available ionosondes (about 150 stations) during the years 1954 to 1958, altogether about 10,000 station-months of data (Hanbaba, 1995).

A numerical mapping technique developed by Jones and Gallet (1962, 1965) was used to represent the *diurnal* and *geographic* variations of $foF2$ and $M(3000)F2$ in the CCIR maps. They first used Fourier analysis in the universal time system for monthly median diurnal variation (one hour time intervals) obtained by observations of each available ionosonde (Bilitza, 1990). The analysis of the data from each ionosonde station was done month by month. The least squares method was used to estimate the Fourier coefficients of each month for each station. Then, in order to provide separately a worldwide description for each coefficient, the orthonormal and spherical forms of the Legendre functions were applied

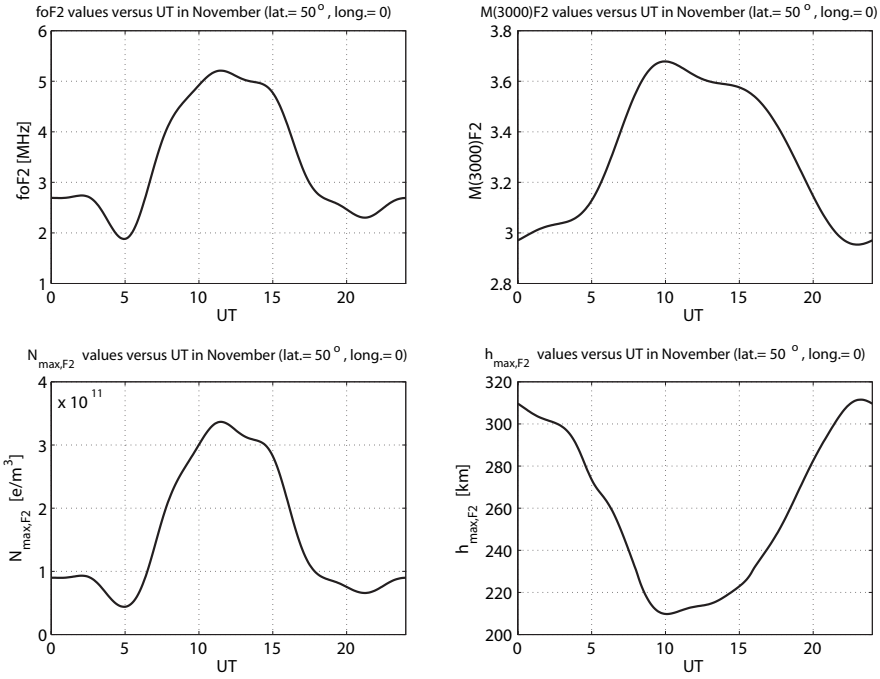

Figure 5.4: Typical daily variations of $foF2$ (top-left) and $M(3000)F2$ (top-right) computed from the CCIR maps at a point ($\varphi = 50°$ on the Greenwich meridian) in November when $R_{12} = 60$, and the corresponding values of $n_{max,F2}$ (bottom-left) and $h_{max,F2}$ (bottom-right) computed by NeQuick

for the expansion of the geographical variation of the each coefficient (Jones and Gallet, 1962a). It should be noted that the classical spherical harmonic functions were not used, however, since they are not orthogonal relative to the positions of stations. Orthogonality is essential because an optimum truncation of the series has to be made (Jones and Gallet, 1962b).

The CCIR maps for the $foF2$ parameter consist of 988 coefficients and for $M(3000)F2$ they consist of 441 coefficients for each month. The CCIR provides two sets of coefficients for each of $foF2$ and $M(3000)F2$, one for low and one for high solar activity ($R_{12} = 0$ to 100); coefficients for intermediate levels of solar activity are obtained by linear interpolation (Bradley, 1990). The complete CCIR maps of $foF2$ and $M(3000)F2$ for 12 months in a year consist of $(988 + 441) \cdot 2 \cdot 12 = 34296$ coefficients. These coefficients are available in 12 files (each for one month) named CCIRmm.ASC, where mm is $month + 10$, accessible from *ftp://nssdcftp.gsfc.nasa.gov/models/ionospheric/iri/iri2007/*.

Due to the strong magnetic control of the $F2$ region, the CCIR maps were published in a special magnetic field coordinate system in terms of geodetic latitude, longitude, and modified dip latitude that was first introduced by Rawer in 1963 (Bilitza, 1990). The modified dip latitude is computed using equation 2.23. The time system of the maps is universal time (UT) and R_{12} (the 12-months-running mean) defined in equation 2.14 is used to specify the solar activity level in a given month.

To show typical daily variation of $foF2$ and $M(3000)F2$ in November, their values computed from the CCIR maps at a point with $\varphi = 50°$ located on the Greenwich meridian

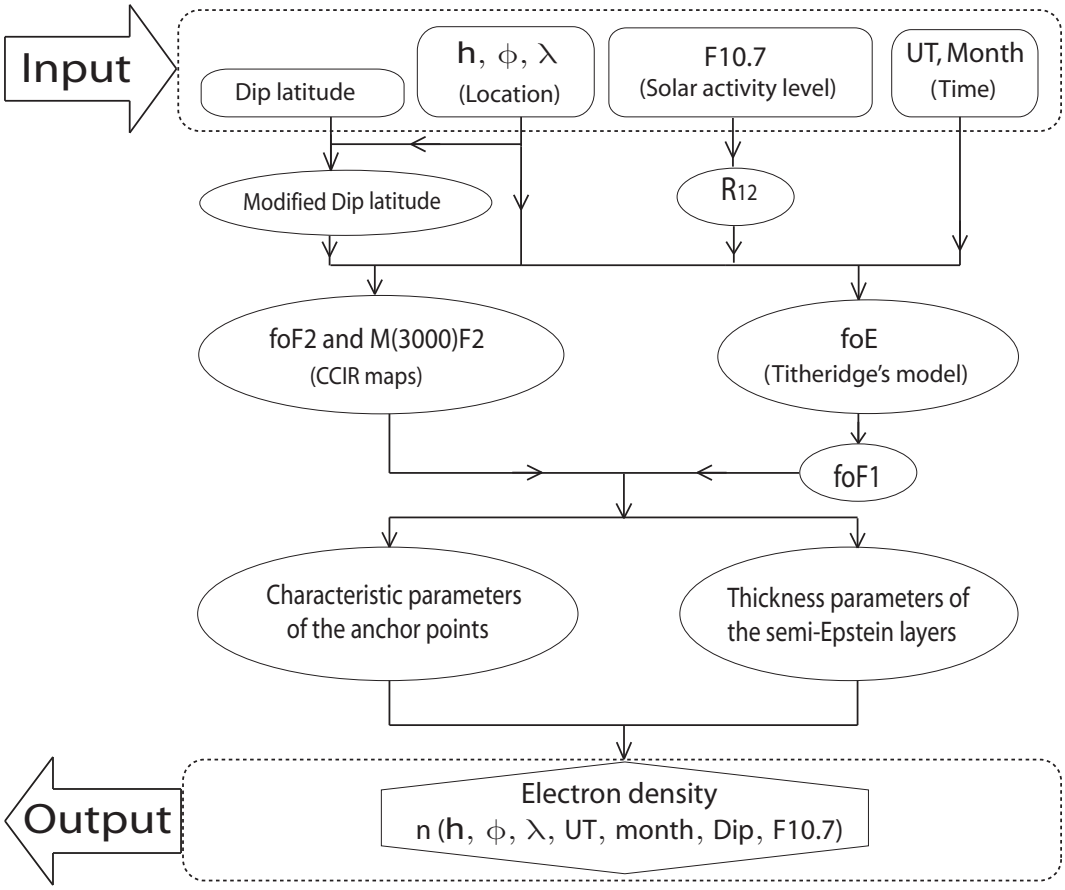

Figure 5.5: Presentation of NeQuick in a diagram

during a medium level of solar activity $R_{12} = 60$ are shown in figure 5.4. This figure also shows the corresponding values of the peak electron density and the peak height of the $F2$ region ($n_{max,F2}$, $h_{max,F2}$) computed using the NeQuick formulation.

5.3.2 Diagrammatic presentation of NeQuick

In order to aid understanding of the NeQuick processing procedure, the connection between input and output values through all the required processing is shown by a diagram in figure 5.5. The F10.7 index used for solar activity level is converted to R_{12} using the equation 2.16 in section 2.2.3. figure 5.5 shows all the required input values of NeQuick: location (geographic coordinates), time (UT and month), level of solar activity (F10.7), and a geomagnetic field related parameter (dip latitude). The NeQuick calculates dip latitude at a given point by third-order interpolation in geographic latitude and longitude using tabulated values in a grid point map that is available in the input file of 'diplats.asc'. This file was generated from the coefficients for spherical harmonics expansion of the IGRF for epoch 1977 (see section 2.3.2). The real geomagnetic inclination (Dip) is calculated from the dip latitude (Dipl) using the equation 2.22.

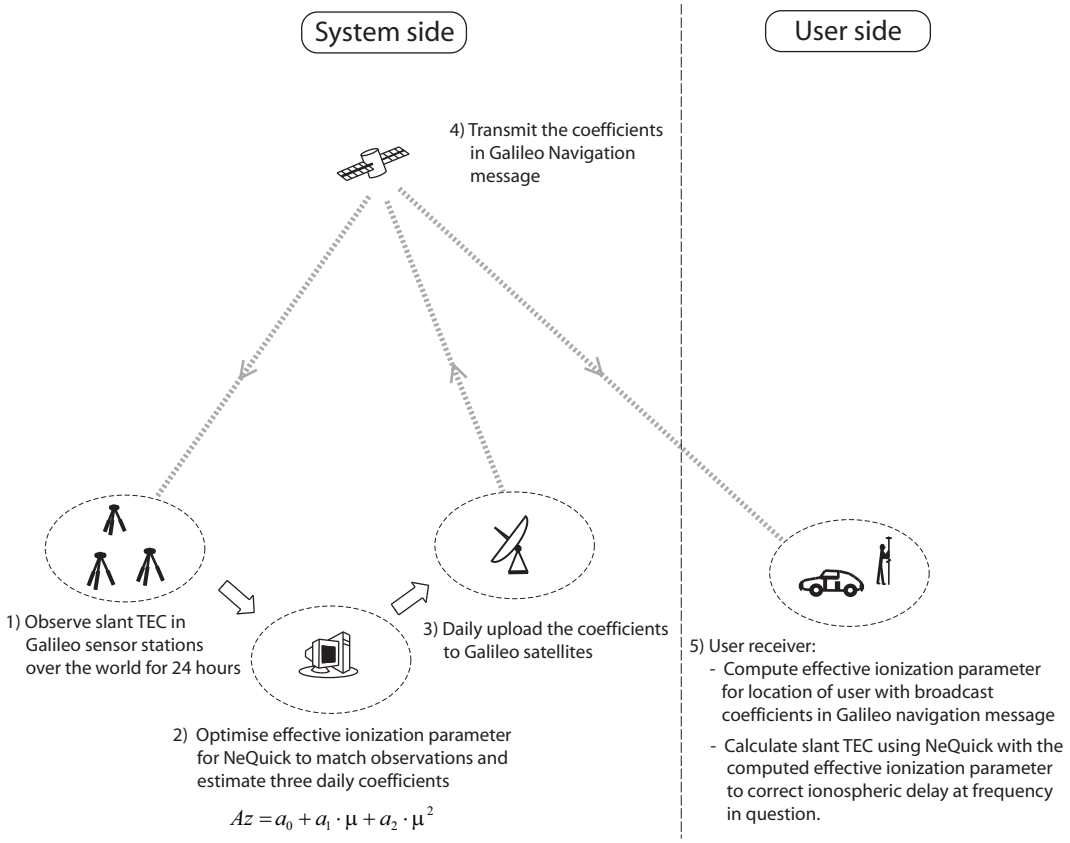

Figure 5.6: Galileo single-frequency ionospheric correction algorithm

5.4 NeQuick for the Galileo navigation system

Among all the empirical models that produce the spatial electron density distribution for a given time and location, the NeQuick model has been proposed as a real-time ionospheric correction model for single-frequency positioning operations in the future *Galileo navigation system* of the European Space Agency (ESA) to compute the slant total electron content (STEC) along arbitrary ground to satellite or satellite to satellite ray paths (Nava et al., 2005; Schluter et al., 2004; Radicella and Leitinger, 2001). Although the NeQuick model only provides electron density, the STEC can be computed by integrating along the signal path.

As seen so far, to compute the ionosonde parameters of the $F2$ region (i.e. $foF2$ and $M(3000)F2$) from the CCIR maps and foE from Titheridge's model, NeQuick was originally developed to use the 12-months-running mean R_{12} or monthly averaged F10.7 index of solar activity. To use the NeQuick model for real-time applications, such as a Galileo ionospheric correction model, the monthly averaged F10.7 index is replaced by a daily input parameter to take the daily variation of the solar activity and the user's local geomagnetic conditions into account. This daily NeQuick input parameter is the so-called *effective ionization level*, which denoted by Az in unit of solar flux $10^{-22}Wm^{-2}Hz^{-1}$, it characterizes the physical conditions of the ionosphere (Azpilicueta et al., 2003).

The Galileo system as a *stand-alone* navigation system relies on its own global network of permanent stations to provide the effective ionization level in real-time for single-frequency users. The algorithm that will be implemented, for correcting the ionospheric effects on the Galileo single-frequency user's observations, is shown in figure 5.6. It involves different operations for the ground, space and user segments. The algorithm consists of 5 steps; step 1 to step 4 are carried out by the Galileo ground and space segments. In the next section the *nominal approach* is described that is used by the ground segment for optimizing the effective ionization level for NeQuick to match observations is described.

5.4.1 Effective Ionization Level (Az parameter)

In the optimization of NeQuick to use for real-time single-frequency point positioning operations in the Galileo system, the *effective ionization level* (or Az parameter) plays a crucial role. Therefore, it can be said that it is an index to introduce daily solar activity into NeQuick, in which it is not only affected by local geomagnetic conditions but also by the optimization of the model to minimize model error with respect to the reference measurements.

The effective ionization level is a *key* input parameter for the NeQuick model as a Galileo ionospheric correction model. At a user's site it will be computed from a second order polynomial (Schluter et al., 2004) as a function of the user's modified dip latitude (μ):

$$Az = a_0 + a_1 \cdot \mu + a_2 \cdot \mu^2 \tag{5.17}$$

The three coefficients of the polynomial (a_0, a_1 and a_2) act as *NeQuick model parameters* and will be broadcast as part of the Galileo navigation message, see figure 5.6, (SIS-ICD, 2006).

5.4.2 Estimation of the effective ionization level (nominal approach)

In standard operation for Galileo, the worldwide coefficients a_0, a_1 and a_2 will be deduced by least-squares fit of a second-order polynomial to the estimated Az values from at least 20 permanent ground monitoring (or sensor) stations of the Galileo system with a near-uniform global distribution (Radicella et al., 2003). In this thesis, this is called as *nominal approach* for estimation of the coefficients. The Az value at each station will be estimated from the observations of the previous day. The coefficients will be updated once every 24 hours and broadcast to single-frequency users as a part of Galileo navigation messages. The Az value at each monitoring station will be estimated by minimizing the squared model error for a 24-hour period over a range of Az values ($Az \geq 64$) using the Brent optimization method (Brent, 1973):

$$\hat{Az} = \underset{\substack{Az \geq Az_{\min} \\ Az < Az_{\max}}}{\arg\min} \sum_{i=1}^{n} |TEC_{Measured} - TEC_{NeQuick}(Az)|_i^2 \tag{5.18}$$

where n is the number of observations from a station for all satellites in a day (Prieto-Cerdeira et al., 2006). The range of possible Az values is 64 to 193 (same as the range

of F10.7 index) in the original formulation of this algorithm. However, as it will be shown later, it is necessary to estimate Az over a much wider range (see subsection 5.5.2).

This method of estimation of the Az parameter has also been used in (Memarzadeh and van der Marel, 2006a) for estimating Az for individual stations. In Memarzadeh and van der Marel (2006a) the measured TEC was computed from dual frequency GPS observations, taking into account satellite *differential code biases* (DCB) computed by IGS. If the station is an IGS station the DCB of receiver can be obtained through IGS; if not, the receiver DCB has to be estimated (see subsection 4.6).

It must be mentioned that the use of NeQuick is not restricted to Galileo; it can also be used at least in post-processing applications of GPS and GLONASS. However, as Galileo is under development, values for Az are not yet available and therefore at the moment NeQuick cannot be used as a real-time ionospheric correction model for GNSS observations. In post-processing applications, users could estimate the local effective ionization level from daily permanent GPS station data, but since not all users are prepared or equipped to do this, this is not a very desirable solution.

In the section 5.5, we will present an alternative approach for estimation of the Az parameter using the *Global Ionospheric Maps* (GIM) from the IGS instead of raw GPS data.

5.4.3 Improved version of NeQuick

According to (Leitinger et al., 2005; Radicella and Zhang, 1995), several modifications have been applied to improve the 1994 version of NeQuick; they will be addressed in this section. The modifications in the model implementation and formulation are as follows:

- The upper limit of the Az parameter ($Az \leq 193$) is eliminated,
- $M(3000)F2$ values less than 1 are replaced by 1,
- The input file of dip latitudes (diplats.asc) should be updated every five years from the IGRF (using the field predicted for the following five years),
- In order to compute the A_{F2} quantity, the equation $A_{F2} = n_{max,F2} - 0.1 n_{max,F1}$ is used instead of $A_{F2} = n_{max,F2}$ (in Table 5.1),
- In Tables 5.2 and 5.3, the following equation is used for computation of the z_X quantity (index X is $F1$ or E):

$$z_X = \frac{h - h_{\max,X}}{B} \exp\left\{ \frac{10}{1 + 2\left|\frac{h - h_{\max,F2}}{B}\right|} \right\} \tag{5.19}$$

- The complicated formulation for the peak height of the $F1$ region (equation 4.10) is replaced by the following simple equation:

$$h_{max,F1} = \frac{h_{max,F2} - h_{max,E}}{2} \tag{5.20}$$

- During daytime under all conditions the peak frequency of the $F1$ region was originally $foF1 = 1.4foE$. In the improved NeQuick model for $foE < 2$ MHz, the peak frequency of the $F1$ region is set to $foF1 = 0$ and if $0.85foF2 < 1.4foE$ then $foF1 = 0.85 \times 1.4foE$,
- The following simplified formulations should be used for the thickness parameters of the $F1$ region:

$$B_{top,F1} = 0.3\,(h_{\max,F2} - h_{\max,F1}) \tag{5.21}$$

$$B_{bot,F1} = 0.5\,(h_{\max,F1} - h_{\max,E}) \tag{5.22}$$

It is important to know that the version of NeQuick for the Galileo navigation system is not finalized yet and it may still be modified. A *new version* of NeQuick has been introduced recently with major changes in the topside formulation and with important modifications in the bottom side description (Nava et al., 2008). The computations in this thesis were carried out before the new version of NeQuick was introduced, using the NeQuick source code version 1994 after applying the above mentioned modifications, was used in the computations.

5.5 Estimation of the effective ionization level using GIM

Since Global Ionosphere Maps routinely produced by the IGS are provided from more than 200 globally distributed IGS sites around the world we believe that GIM are robust and reliable enough (under different ionospheric climates) to employ them for estimation of the NeQuick model parameters. In this section a new approach is proposed to utilize the Global Ionospheric Maps for estimation of the NeQuick model parameters. The new approach is called the *alternative approach* in this thesis. Note that the proposed approach can be applied for estimation of the model parameters using the different versions of NeQuick model.

5.5.1 Estimation of the effective ionization level (alternative approach)

Using the VTEC times series computed from the GIM, the effective ionization level or Az parameter can be computed in three steps. In the first step, a daily time series of VTEC is produced for each grid point of the GIM. The time interval for VTEC is 2 hours, so the time-series for each grid point consists of 13 points (see figure 5.7). In the second step, the provided VTEC time series of a grid point is substituted into equation 5.18 (as the measured VTEC) to estimate a daily value of the Az parameter for the grid point by optimization of the NeQuick model as a function of the Az parameter. Since the NeQuick model varies smoothly with Az, to economize on computation time the modeled VTEC was computed at Az intervals of 16. The range over which the optimum Az is estimated has been increased from 64–193 to 0–209 (see figure 5.8). In figure 5.8, which shows a typical optimization of NeQuick, the vertical axis is a summation of the norm in equation 5.18 (model error) for which $n = 13$. As expected, the curve of model error versus the Az

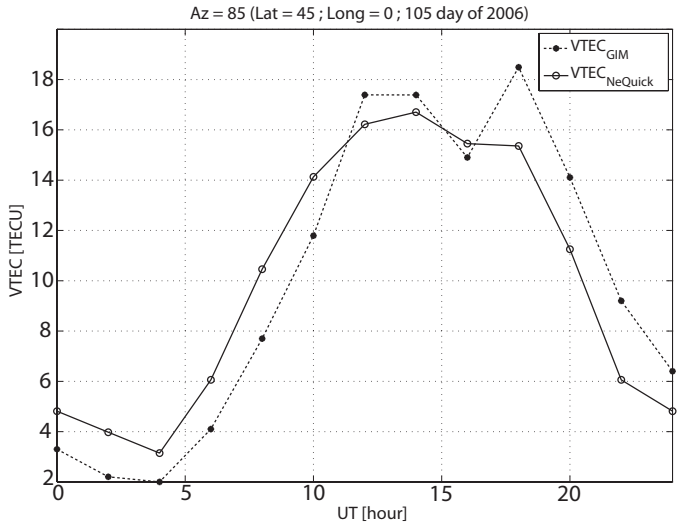

Figure 5.7: Daily VTEC time series computed by GIM and by NeQuick using the optimum effective ionization level ($Az = 85$) at a grid point ($\varphi = 45°$ on the Greenwich meridian) for day 105 of 2006

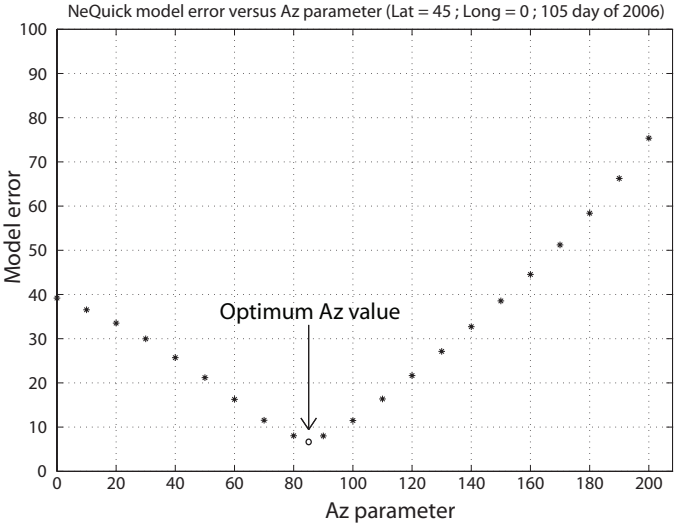

Figure 5.8: Typical example of the optimization of NeQuick as a function of the Az parameter at a grid point ($\varphi = 45°$ on the Greenwich meridian) for day 105 of 2006

parameter has a parabola shape with a unique minimum related to the optimum Az value. This second step results in a daily grid-based map of the Az parameter around the world.

In the third step, a second order polynomial is fit to the Az values with respect to modified dip latitude, and the coefficients of the polynomial are estimated. The main advantages of the alternative approach are:

- The VTEC time series produced from GIM is always healthy and available for every location over the world,
- There is no need to worry about the receiver Differential Code Bias (DCB)

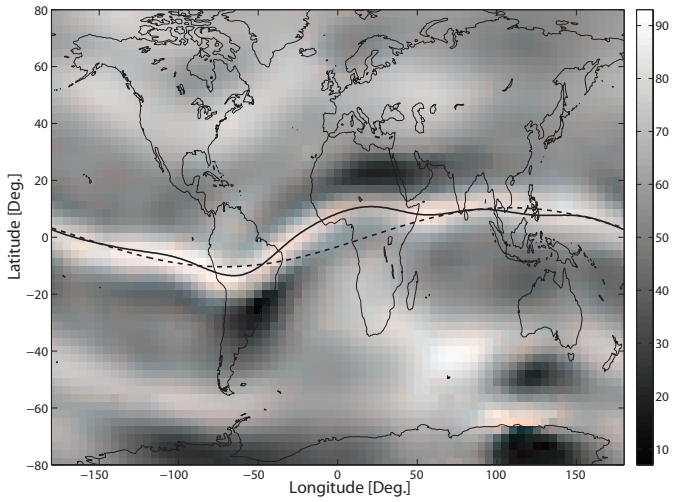

Figure 5.9: Daily Az grid-map (in unit of solar flux $10^{-22}Wm^{-2}Hz^{-1}$) using GIM for day 74 of 2006 (Geomagnetic equator and Dip equator denoted by dashed and solid curves, respectively)

A further advantage of this approach is that it provides a tool to investigate the spatial dependency of the Az parameter. The spatial variation of Az is studied in subsection 5.6.2. The alternative approach is validated and compared with the nominal approach in section 5.6.

In the alternative approach daily value of Az in a grid point is estimated using the 13 values of VTEC at the grid point ($n = 13$ as the size of VTEC time-series). It is also possible to use equation 5.18, with $n = 1$, to compute an Az value for each sample value in the VTEC time-series. This would therefore result in two-hourly Az maps. This could be used, for instance, to study the daily variability in Az, or to predict the Az for the next day, instead of using the values of the previous day as specified by the Galileo algorithm.

Another advantage of using two-dimensional Az maps is that the fitting approach of step 3, where a second-order polynomial in modified dip latitude is fitted to the Az data, can be evaluated and possibly improved.

5.5.2 Daily grid-based map of the effective ionization level

An example of a daily map of the effective ionization level Az, as provided by the second step of the alternative approach, is given in figure 5.9. Tith map has been produced using the final CODE GIM map for day 74 of 2006, showing the color coded Az at every grid point of the GIM.

It should be mentioned that we used 0 instead of 64 for the lower boundary of the Az parameter in equation 5.18. Since Az is not highly correlated with the F10.7 index (see subsection 5.6.3), due to local geomagnetic condition effects, the lower boundary of the Az parameter differs from the F10.7 index and it is reasonable to expect Az values less than 64 at some of the grid points.

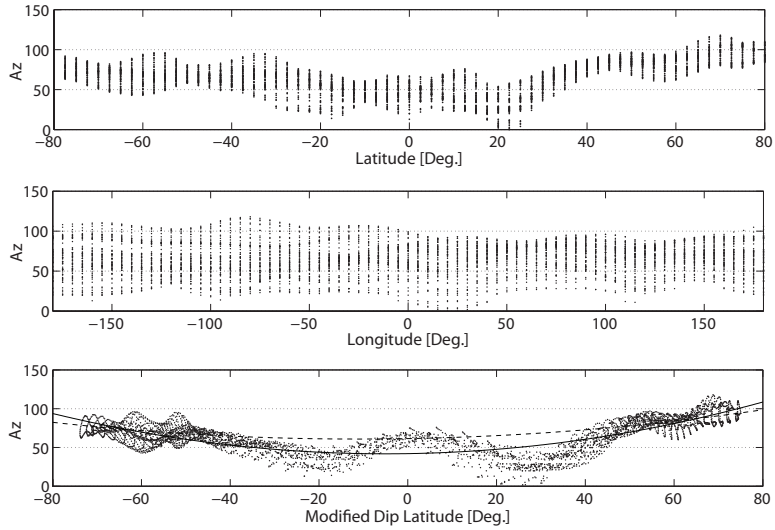

Figure 5.10: Estimated Az values at the GIM grid points (using GIM) versus latitude, longitude and modified dip latitude for day 74 of 2006. The solid curve in the bottom sub-figure shows the second-order polynomial fit to the estimated Az values with $Az_{min} = 0$ and the dashed curve shows the fit polynomial with $Az_{min} = 64$

Figure 5.9 gives a strong impression that the Az parameter is highly correlated with the geomagnetic latitude. To verify this, the estimated values of the Az parameter are plotted with respect to latitude, longitude and modified dip latitude in figure 5.10. Clearly, we can observe from figure 5.10 that there is a strong latitudinal dependency, especially with respect to the modified dip latitude, but that there is no apparent dependence on longitude. The scatter of Az when plotted against the modified dip latitude is smaller than when plotted against latitude, so clearly the modified dip latitude is a better argument for function fitting than the latitude. In order to estimate the NeQuick model parameters for Galileo, a second-order polynomial in modified dip latitude was fitted to the Az values. To demonstrate the NeQuick model parameters for different choices of the lower boundary of Az, a second-order polynomial was fitted to the Az values with $Az > 0$ (solid line in figure 5.10) and another second-order polynomial was fitted to Az values with $Az > 64$ (dashed line). The result is that the coefficients of the polynomial (the NeQuick model parameters) were changed significantly by different choices of the lower boundary of Az. More examples are given in the next sections.

5.5.3 Az parameter for single point positioning

The NeQuick model parameters, for the real-time single point positioning, will be provided for the users of the Galileo system in real-time by the Galileo navigation message. For this purpose, in the nominal approach, the model parameters are estimated every 24 hours using observations of the previous day in the permanent Galileo monitoring stations. Therefore, the corrections computed by NeQuick do not actually describe the ionosphere at the time of single point positioning, if the users apply the corrections in real time.

There are two possibility to estimate the NeQuick model parameters using GIM for the

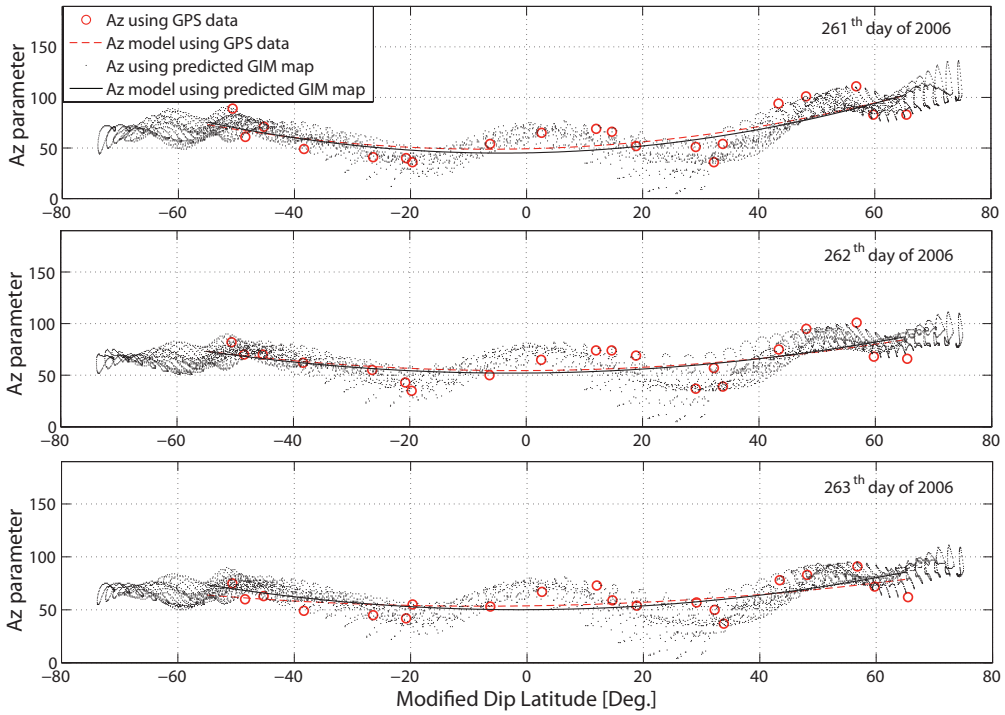

Figure 5.11: Az values at the GIM grid points (dots) estimated using the predicted version of GIM and predicted Az values at 20 IGS stations using observations of the previous day (circles) for days 261, 262 and 263 of 2006. Dashed and solid lines respectively show the fitted 2nd order polynomials for circles and dots.

single point positioning : 1) estimate using the *predicted* version of GIM (produced by the CODE), 2) predict using a time series of the model parameters over the previous days. Using the predicted model parameters for real-time application of the NeQuick model (or single point positioning) gives rise to questions about its efficiency and performance. One might ask: does the predicted model parameters have sufficient accuracy? To provide an answer to this question, the predicted version of GIM for days 261, 262 and 263 of year 2006 were processed to obtain the predicted version of the daily Az map. The Az values at the 20 IGS stations were also estimated using the daily GPS observations of days 260, 261, and 262 of 2006. The spatial distribution of Az values from the predicted version of GIM and the estimated Az values for 20 IGS stations (using data of the previous day) are depicted for the three days in figure 5.11. Note that in all processing in current and next sections, $Az_{min} = 0$ is used as the lower boundary of the Az parameter in the equation 5.18. Figure 5.11 clearly shows that there is high consistency between the values of Az estimated using the predicted version of GIM and Az values estimated for each IGS station using GPS data of the previous day. The second-order polynomials fit to Az values from the two approaches don't differ significantly from each other and we can conclude that use of the predicted version of GIM for computation of the NeQuick model parameters is suitable for real time application of NeQuick.

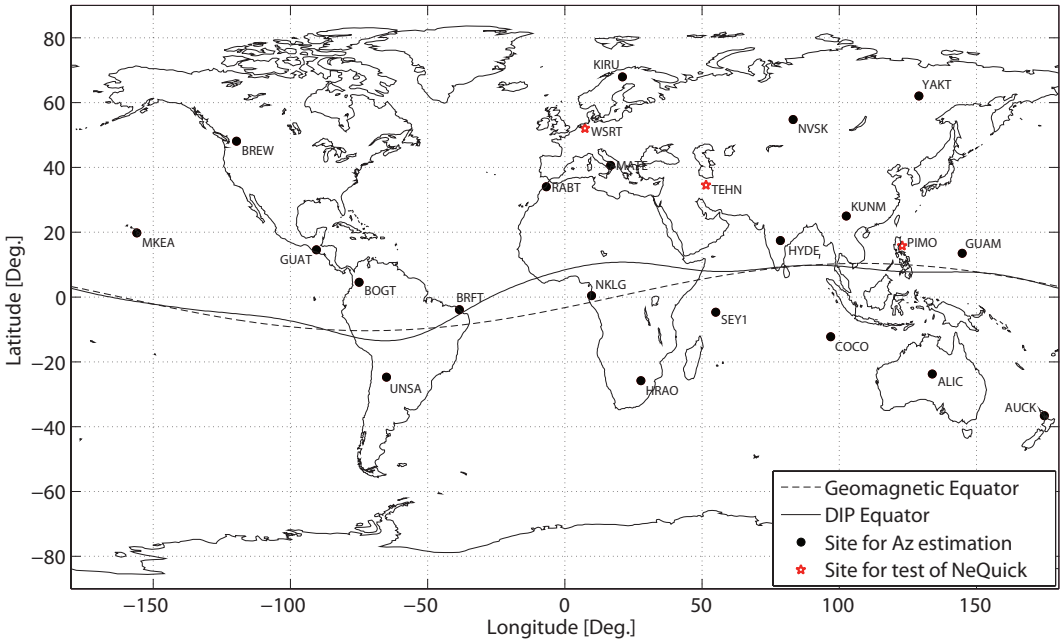

Figure 5.12: Distribution of the 20 IGS stations (black bold dots) used for estimation of the NeQuick model parameters and locations of the three IGS stations (red stars) used to test performance of the NeQuick model

5.6 Validation of the alternative approach

The goal here is to investigate the validity of Az parameter estimation using the GIM. For this purpose, the consistency of the alternative approach with the nominal approach is evaluated by comparing the estimated effective ionization level from the two approaches. The comparison will be done under geomagnetically different conditions of the ionosphere.

Description of data and processing Since the Galileo system is under development we used actual daily GPS observations from 20 IGS stations with a near-uniform global distribution (see figure 5.12) to estimate the NeQuick model parameters using the nominal approach. The daily effective ionization level for each IGS station was estimated using slant TEC values measured every 10 minutes for each satellite in view above an elevation mask angle of 40 degree during a 24-hour period. For estimation of the effective ionization level using the global ionospheric maps, the final version of GIM from the CODE was used.

5.6.1 Consistency of the approaches

To test the consistency of the proposed and the nominal approaches for estimating the Nequick model parameters under different geomagnetic conditions of the ionosphere, we processed data from 20 IGS stations on the 74th, 16th, 258th, and 349th days of year 2006 for which the Kp index varied between 0 and 9. For these four days, the effective ionization level estimated using GIM and the values estimated using GPS data at each IGS stations are shown versus modified dip latitude in figure 5.13. The estimated Az values using the GPS data at IGS stations are mostly in the range of the estimated values using

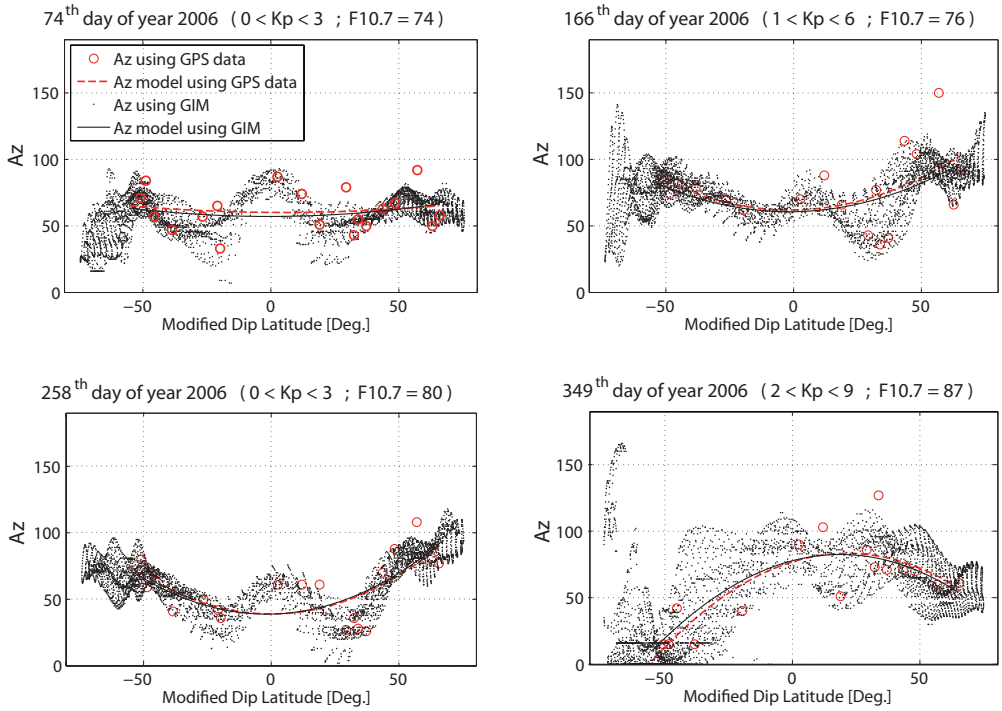

Figure 5.13: Estimated Az using GIM at the grid points (dots) and the estimated Az at 20 IGS stations using daily GPS data (circles) under different geomagnetic conditions of the ionosphere in days 74, 166, 258, 349 of 2006 (the solid curve is the fitted 2nd-order polynomial to the dots and the dashed red curve is the fitted 2nd-order polynomial to the circles)

GIM for the grid points whose modified dip latitudes are nearly the same as the station. The estimated Az for a few stations are out of the band estimated from the GIM. This has happened at stations where the quality of the measured slant TEC values were degraded due to increasing station-dependent errors (such as multipath and cycle-slip) in the GPS data of the station. However, it is obvious from figure 5.13 that there is a high consistency between the values of the Az estimated using GIM and the values estimated using daily GPS data at the IGS stations under different geomagnetic ionospheric conditions. To see the consistency between the estimated NeQuick model parameters from both approaches, in figure 5.13 a second-order polynomial was fit to the dots (Az values from the GIM grid points) and another second-order polynomial was fit to the circles (Az values from the 20 IGS stations). Note that to fit the polynomial to the dots, only Az values of the grid points located within the range of the maximum and minimum modified dip latitudes of the 20 IGS stations were used. If we compare the fitted polynomials (to the dots and circles), we can confidently conclude from the subfigures of figure 5.13 that estimation of the NeQuick model parameters using GIM is highly consistent with estimation by the nominal approach. This consistency was seen under different geomagnetic ionospheric conditions.

It can also be seen that a second-order polynomial is not necessarily the best-fitting function to model the behavior of the Az parameter against the modified dip latitude. Therefore it could be advantageous to use a different function for modeling the Az parameter as a function of the modified dip latitude. The next section shows more examples and presents more discussion about the best-fitting function for the spatial dependency of the Az parameter.

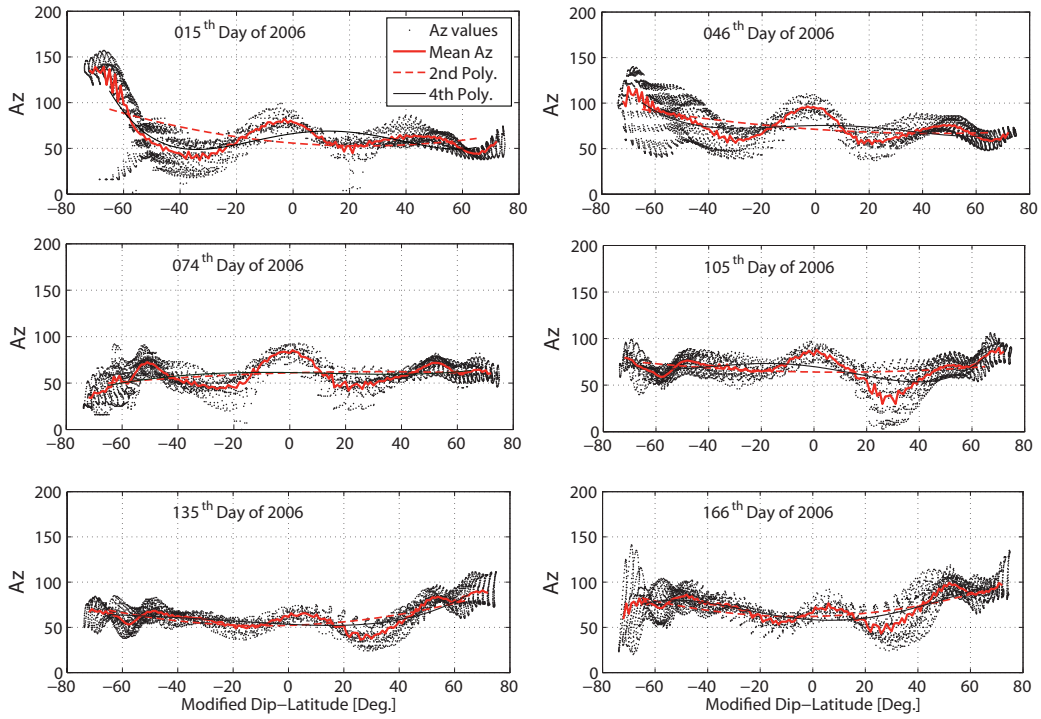

Figure 5.14: Estimated Az using GIM at the grid points (dots) versus modified dip latitude and the fitted 2nd- and 4th-order polynomials (dashed and solid curves, respectively) for days 15, 46, 74, 105, 135 and 166 of 2006 (the meridian mean Az value is indicated by the solid-bold red curve)

5.6.2 Modeling the spatial dependency of the Az parameter

In this subsection, the aim is to demonstrate that a second order polynomial is not the best-fitting function to model the behavior of the Az parameter versus modified dip latitude (μ). Employing GIM for estimation of the Az parameter provides a valuable opportunity to study the spatial variation of the Az parameter over the world. Daily Az maps using GIM were produced for 12 days in 2006 (the 15^{th} day of every month). Figures 5.14 and 5.15 show the estimated Az parameters (indicated by dots) with respect to modified dip latitude individually for each of the 12 days. In these figures, the solid-bold (red) curves show the meridian mean Az value versus modified dip latitude. From the plots in these figures, it can be seen that when the linear trend is ignored, the behavior of the Az parameter against modified dip latitude is almost symmetric and a local maximum around the geomagnetic equatorial region always exists. Consequently, a forth-order polynomial can be used as a simple function to express the spatial variation of the Az parameter against modified dip latitude. In polar regions, the fluctuation of the Az parameter is increased, especially during the summer period of the hemisphere, therefore a trigonometric function is preferable for higher dip latitudes.

For comparison, a fourth-order and a second-order polynomials are fitted to the estimated daily Az values in figures 5.14 and 5.15. Since fluctuations in the Az values increased in the polar regions, the higher latitudes ($\mu > 65^{\circ}$) were excluded in the fitting of the polynomials. In table 5.4, the RMS values (Root-Mean-Squared) of the fitted 2^{nd}- and

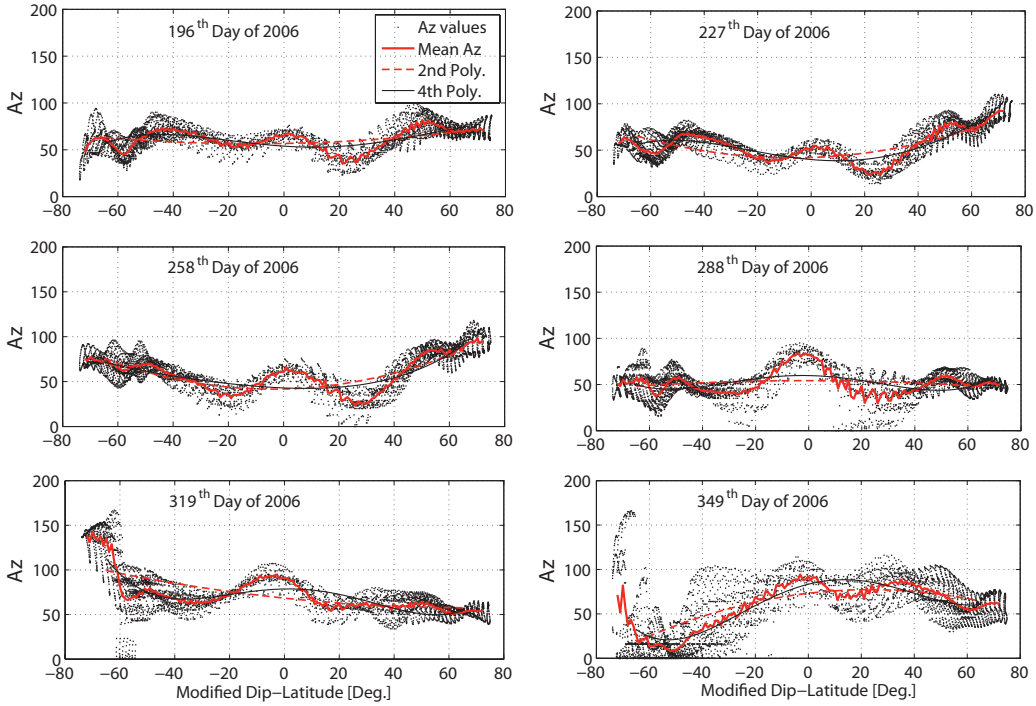

Figure 5.15: Estimated Az using GIM at the grid points (dots) versus modified dip latitude and the fitted 2nd- and 4th-order polynomials (dashed and solid curves, respectively) for days 196, 227, 258, 288, 319, and 349 of 2006 (the meridian mean Az value is indicated by the solid-bold red curve)

4^{th}-order polynomials are given. The RMS values for the 4^{th}-order polynomial are slightly smaller. This implies, at least, that the second-order polynomial proposed for the Galileo system may not be the best function to model the spatial distribution of the Az parameter with respect to modified dip latitude.

Table 5.4: RMS error of fitting 2^{nd}- and 4^{th}-order polynomials to the daily spatial distribution of the Az parameter (estimated using GIM) for 12 days of 2006

Day of year	015	046	074	105	135	166	196	227	258	288	319	349
2^{nd}**order polynomial**	25	18	14	14	12	17	13	13	13	14	21	29
4^{th}**order polynomial**	18	17	13	13	11	15	12	11	13	15	17	25

5.6.3 Correlation between Az and F10.7

As mentioned in section 5.4, for use of the NeQuick model as a Galileo ionospheric correction model, the monthly averaged F10.7 index is replaced by the daily Az parameter which takes into account the variable daily solar activity and the user's local geomagnetic conditions. It implies that the daily solar radiation intensity is introduced into the NeQuick model by Az parameter. Therefore, it is expected that Az is correlated to the solar radiation intensity (F10.7 index). This was investigated by the estimated Az values using GIM. Due to the spatial dependency of the Az parameter we investigated the correlation

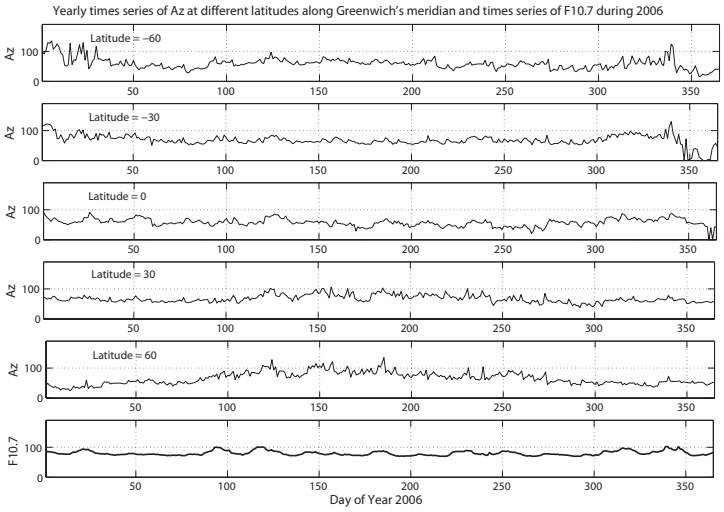

Figure 5.16: Yearly time series of the Az parameter at different latitudes ($\varphi = -60 : 30 : +60°$) along Greenwich's meridian and yearly time series of the F10.7 index during 2006

between Az and F10.7 at different latitudes ($\varphi = -60{:}10{:}+60°$) along Greenwich's meridian ($\lambda = 0°$). Time series of daily Az parameter estimated using GIM were produced for different years. For illustration of the yearly Az and F10.7 time series, only the time series of Az for latitudes $\varphi = -60{:}30{:}+60°$ and the time series of F10.7 during 2006 are plotted in figure 5.16.

In order to find out how the Az parameter is correlated with the solar radiation intensity, the correlation coefficient (ρ) between the yearly time series of Az and F10.7 at each point ($\varphi = -60 : 10 : +60°$) was computed as

$$\rho = \frac{\sigma_{x,y}}{\sigma_x \sigma_y} \tag{5.23}$$

where σ_x and σ_y are the standard deviations and $\sigma_{x,y}$ is the covariance of the yearly time series of Az and F10.7 that can be computed by (Teunissen, 2007)

$$\begin{bmatrix} \sigma_x^2 & \sigma_{x,y} \\ \sigma_{x,y} & \sigma_y^2 \end{bmatrix} = \frac{E^T E}{m-1} \tag{5.24}$$

$$E = \begin{bmatrix} x & y \end{bmatrix}_{m \times 2} \tag{5.25}$$

where m is the size of the time series. Note that x and y are the detrended time series (w.r.t. the mean) of the Az parameter and F10.7, respectively.

The correlation coefficient between Az and F10.7 versus latitude is shown in figure 5.17 for 2003, 2004, 2005 and 2006. It is obvious that the correlation between Az and F10.7 is always positive and increases in the equatorial region. The spatial dependency of the correlation between the Az parameter and F10.7 may be due to local geomagnetic conditions that Az takes into account. figure 5.17 shows that the correlation coefficient was always less than 0.65 and the Az parameter was not highly correlated with F10.7. Consequently, we concluded that limiting the Az parameter with the lower and upper limits for the F10.7 index (respectively 64 and 193) would be unreasonable.

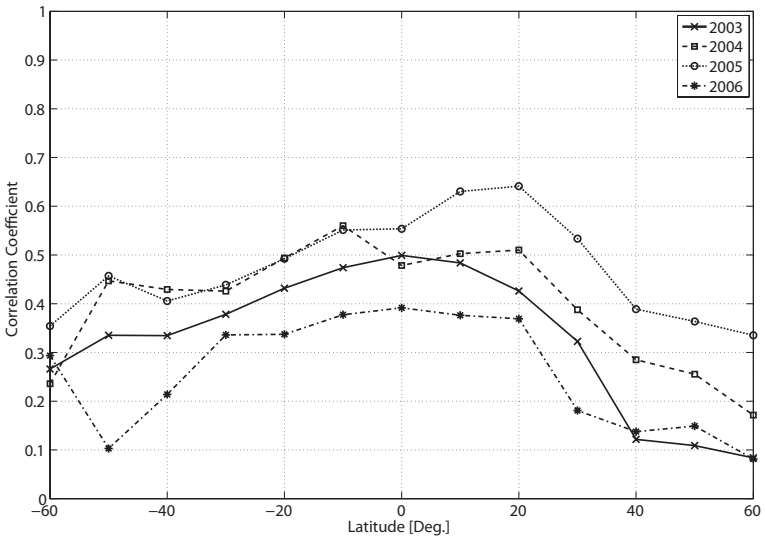

Figure 5.17: Correlation coefficient between Az and F10.7 versus latitude (estimated from yearly time series of Az and F10.7 provided for 2003, 2004, 2005, and 2006)

Lower and upper limits of the Az parameter To choose proper limits for the Az parameter, it is important to know the variations of $M(3000)F2$ and $foF2$ versus the F10.7 index in the CCIR maps. In figure 5.18, typical variations of $M(3000)F2$ and $foF2$ are given (top-right and top-left plots) for F10.7 values (from 0 to more than 1000) at a point located at $\varphi = 40^\circ$ and $\lambda = 10^\circ$ at $UT = 10:00h$ in July. In the lower plots of figure 5.18, the variations of $h_{max,F2}$ and $n_{max,F2}$ versus F10.7 are shown, based on the NeQuick model formulation (equations 5.2 and 5.9). It can be seen from the plot of $h_{max,F2}$ (bottom-right) that with increasing F10.7 the height of peak electron density of the $F2$ region goes to infinity. In reality, it does not occur due to decreasing atmospheric density with height. If we look at the plot of $n_{max,F2}$ (bottom-left), we see that for low values of F10.7, $n_{max,F2}$ goes to zero; this implies that the $F2$ region disappears for very low values of F10.7. However, from these plots one can say that it is not reasonable to choose a negative number for the lower limit of Az or very large values for the upper limit of Az. Therefore, a proper lower limit of Az can be 0 and a value around a few hundred would be a proper value for the upper limit. It should be noted that the CCIR maps were developed for two solar activities, one for low and one for high solar activity ($R_{12} = 0$ to 100), and the coefficients for intermediate levels of solar activity are obtained by linear interpolation (Bradley, 1990).

5.7 Performance of the NeQuick ionospheric model

In the previous section, the nominal approach and an alternative approach for estimation of the NeQuick model parameters were compared, and the validity of the alternative approach and its consistency with the nominal approach have been verified. In this section, we aim to study the performance of the NeQuick model in providing the correction of slant ionospheric distance delay for the GPS observations. To compare the NeQuick model with the different

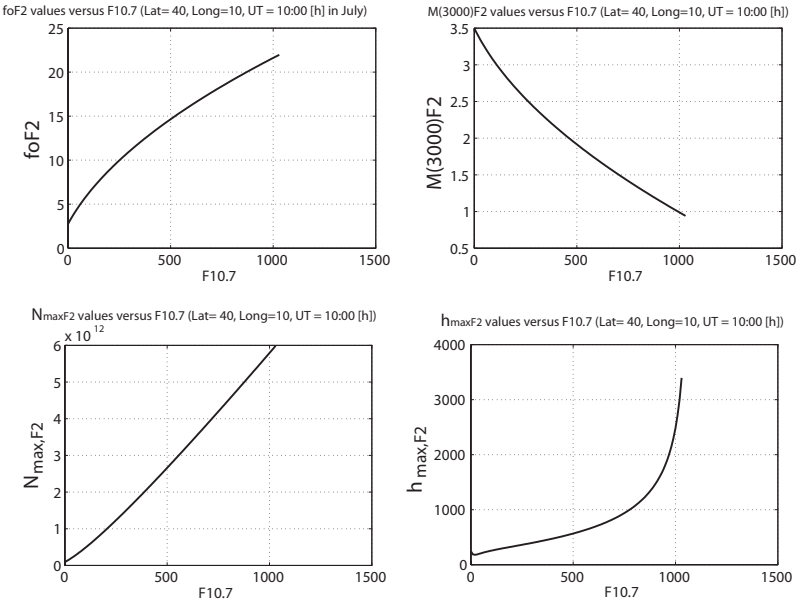

Figure 5.18: $M(3000)F2$ and $foF2$ versus F10.7 in the CCIR maps (top left and right subfigures) and variation of $n_{max,F2}$ and $h_{max,F2}$ versus F10.7 based on the NeQuick model formulation (bottom left and right subfigures)

types of ionospheric models, the performance of the CODE Global Ionospheric Maps (GIM) and the well-known Klobuchar model are also studied.

5.7.1 Data specifications and processing

Three IGS stations—PIMO ($\varphi = 15^\circ$ in Philippines), TEHN ($\varphi = 35^\circ$ in Iran), and WSRT ($\varphi = 53^\circ$ in the Netherlands)—that are located in the mid-latitude region (see figure 5.12) were used to study the performances of the ionospheric models. The slant ionospheric delay between these three IGS stations and all GPS satellites (in view) were measured, using the method of subsection 4.6.2, during 24 hours in the four days 74, 166, 258, and 349 of 2006 (one day in each season). The time interval of the measurements was 10 minutes and a cut-off angle of 15 degree was used for all GPS data. The ionospheric condition was geomagnetically different during the four days (days 74 and 258 were geomagnetically *quiet*, day 166 was *moderate*, and day 349 was *severe*). However one should consider that this analysis was based on a limited number of days and stations and for more reliable conclusion, it would be necessary to process a larger data set.

The measured slant ionospheric delays and the corresponding delays predicted by GIM, Klobuchar and NeQuick, and comparisons between them were processed by a package of Matlab scripts that written by the author.

To compare the performances of the models, the discrepancy between the slant ionospheric delay provided by each model I_{model} ($model$ stands for NeQuick, GIM, or Klobuchar) and the corresponding delay measured by the GPS phase data $\tilde{I}$ (see subsection 4.6) was computed:

$$\delta I_r^s(t) = \tilde{I}_r^s(t) - I_{r,model}^s(t) \tag{5.26}$$

where $\delta I_r^s(t)$ denotes the model error (or model residual) for epoch t between receiver r and satellite s. According to subsection 4.6, the accuracy level of the measured ionospheric delay $\tilde{I}$ is 1 TECU. Consequently, $\tilde{I}$ has been considered as a reference (or true ionospheric delay) and the daily bias in each model and the RMS (Root-Mean-Squared error) of model errors were computed by the first and second moments of $\delta I(t)$:

$$\text{BIAS} = E\{\delta I(t)\} \tag{5.27}$$

$$\text{RMS} = \sqrt{E\{\delta I(t)^2\}} \tag{5.28}$$

where $E\{\}$ is the mathematical expectation. The standard deviation (STD) of the model error is computed by $STD = \sqrt{(RMS)^2 + BIAS^2}$. To estimate model error in percentage (relative model error), the daily Average Percentage Model Error (APME) can be computed as follows (Memarzadeh and van der Marel, 2006a,b):

$$\text{APME} = \frac{1}{n}\sum_{i=1}^{n}\frac{\delta I(i)}{\tilde{I}(i)} \times 100 \tag{5.29}$$

where n is number of measured ionospheric delays from all stations to all satellites in a day. Note that the APME value increases for quiet ionospheric conditions, because the absolute ionospheric delay ($\tilde{I}$) is smaller during the geomagnetically quiet ionosphere. It implies that the APME is not a proper criterion to measure the correctness of a ionospheric model. That is why in the *Galileo specification* for single-receiver ionospheric correction, it is specified that the model error should be no more than 20 TEC Units or 30% of the actual slant TEC, whichever is larger (20 TECUs are equivalent to an execess group delay of 3.25 m at L1 frequency) (Prieto-Cerdeira et al., 2006). However, to compare the performances of the different models under the same conditions, the APME values can be a good criterion for our evaluation in this study.

Estimation of the receiver differential code bias (DCB_r) To measure the slant ionospheric delay using the geometry-free linear combinations, the satellite and receiver differential code biases (DCB^s and DCB_r) are needed. The DCB^s of GPS satellites were obtained from the CODE's IONEX files, but in our computations DCB_r of the receivers (for IGS stations: PIMO, TEHN, and WSRT) were estimated by the least squares adjustment of the functional model of geometry-free linear combinations.

Table 5.5: The monthly averaged differential code bias (DCB_r) of the PIMO and WSRT stations provided by the CODE-IGS center and the corresponding estimated values using daily GPS data; the discrepancies are given inside the brackets (values are in [ns])

Day of	Kp	WSRT		PIMO		TEHN	
2006	index	IONEX	Estimated	IONEX	Estimated	IONEX	Estimated
074	0-3	-1.90	-1.78 [-0.12]	-8.54	-7.87 [-0.67]	-	-3.80 [-]
166	1-6	-2.84	-2.69 [-0.15]	-7.84	-7.03 [-0.81]	-	-2.67 [-]
258	0-3	-3.05	-2.93 [-0.12]	-7.92	-7.10 [-0.82]	-	-2.65 [-]
349	2-9	-1.00	-1.12 [0.12]	-8.03	-7.57 [-0.45]	-	-3.58 [-]

It should be noticed that only monthly averaged values of DCB_r of IGS stations, which are used in GIM maps production, are available in the IONEX files. For comparison, the estimated DCB_r and corresponding values from the IONEX files are shown in table 5.5.

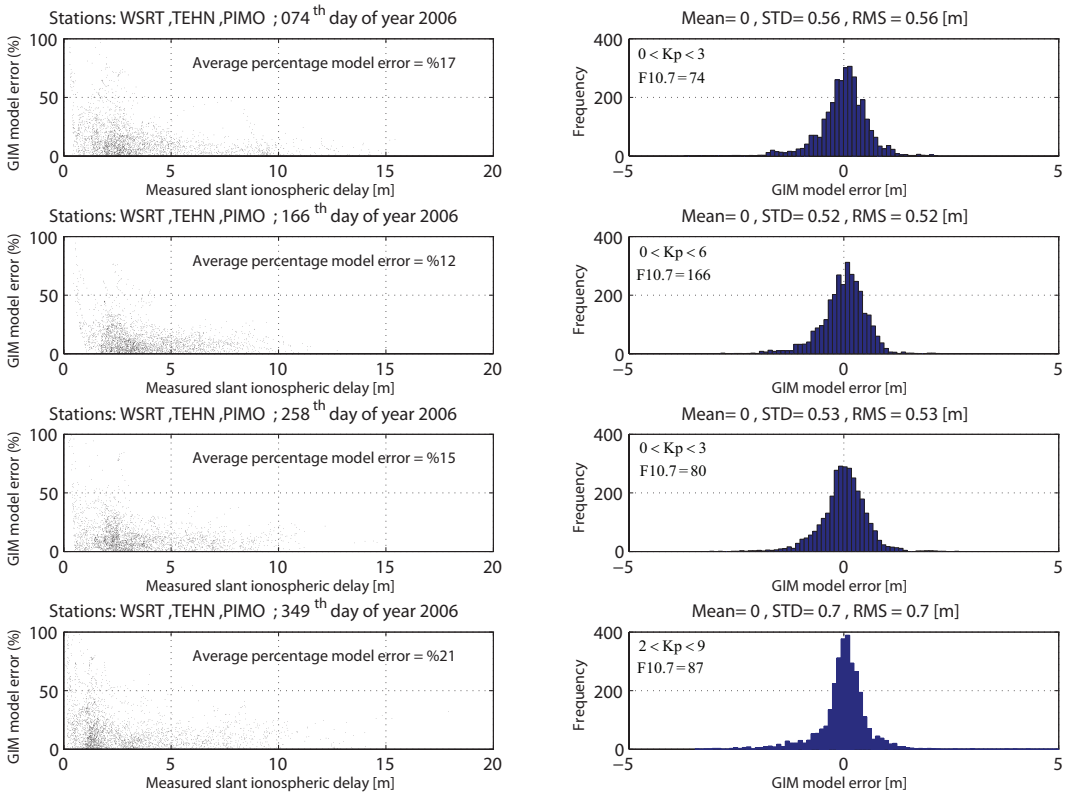

Figure 5.19: GIM model errors in percentage at three IGS stations (PIMO, TEHN, WSRT) versus measured slant ionospheric delay (dots in left subfigures) and histograms of the GIM model errors of all three stations (right subfigures) for days 74, 166, 258, 349 of 2006.

Note that TEHN station is one of the IGS stations but, since it is not used in producing GIM maps, the DCB_r of this station is not available from the IONEX files.

In Table 5.5, the discrepancies between IONEX and the estimated values for DCB_r are always less than one nanosecond and the standard deviation of the discrepancies is 0.34 nanosecond. According to figure 4.5, the standard deviation of the estimated DCB_r should be a few tenths of a nanosecond, so it is in good agreement.

5.7.2 Comparison between the model errors

In figure 5.19, histograms of the GIM model errors (right subfigures) and the relative model errors in percentage (left subfigures) at all three stations (PIMO, TEHN, WSRT) are shown for the four days (74, 166, 258, and 349 of 2006). As expected, the mean model error of GIM was 0 (for all days) and the RMS values were at levels of a few decimeters, which is consistent with the 2∼8 [TECU] accuracy of GIM maps.

In figure 5.20, the same plots for the NeQuick model error are shown. The Az parameter for the NeQuick model was provided using GIM maps (in the alternative approach). In order to evaluate the performance of the NeQuick model with Az parameter values from both the nominal approach (using 20 IGS stations) and the alternative approach, the BIAS, RMS, and APME values for the NeQuick model errors with Az from both approaches were

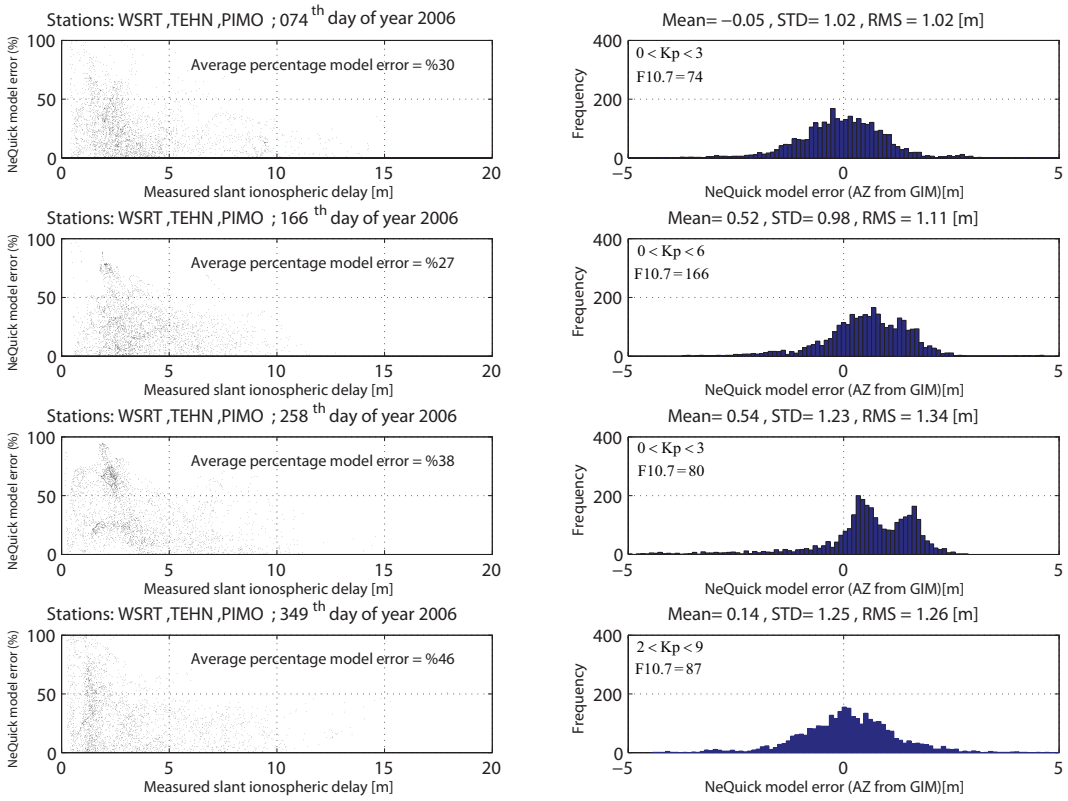

Figure 5.20: NeQuick model errors in percentage at three IGS stations (PIMO, TEHN, WSRT) versus measured slant ionospheric delay (dots in left subfigures) and histograms of the GIM model errors of all three stations (right subfigures) for days 74, 166, 258, 349 of 2006 (Az estimated from the final version of GIM maps).

computed. The BIAS, RMS and APME values of the Klobuchar model errors were also computed in same way.

For comparison between the models, the BIAS, STD, RMS, and APME values of the models are shown in figure 5.21. A remarkable result of figure 5.21 is obtained by comparing the RMS of NeQuick model error when Az was estimated from GIM with those when Az was estimated by the nominal approach. We see that there is only a very slight difference between the RMS values; it can be concluded that estimation of the NeQuick model parameters from GIM for application of the NeQuick model as an ionospheric delay correction model in GNSS is reliable enough. It can be observed that the APME values for NeQuick with Az provided by both approaches are also nearly the same under different ionospheric conditions. This shows that the accuracy of the NeQuick model is not degraded when the model parameters are estimated from GIM, and the NeQuick model parameters estimated using GIM are reliable enough under all ionospheric conditions.

In figure 5.21, as was expected, the BIAS, RMS, STD, and APME values for the GIM are less than those for the NeQuick and Klobuchar models. It follows that the GIM is the best of all. The BIAS, RMS, STD, and APME values for the Klobuchar model are larger than those for the NeQuick model. In other words, our results show that the NeQuick model error is always less than the Klobuchar model error. This results in the NeQuick model performance is better than the Klobuchar model (under geomagnetically different

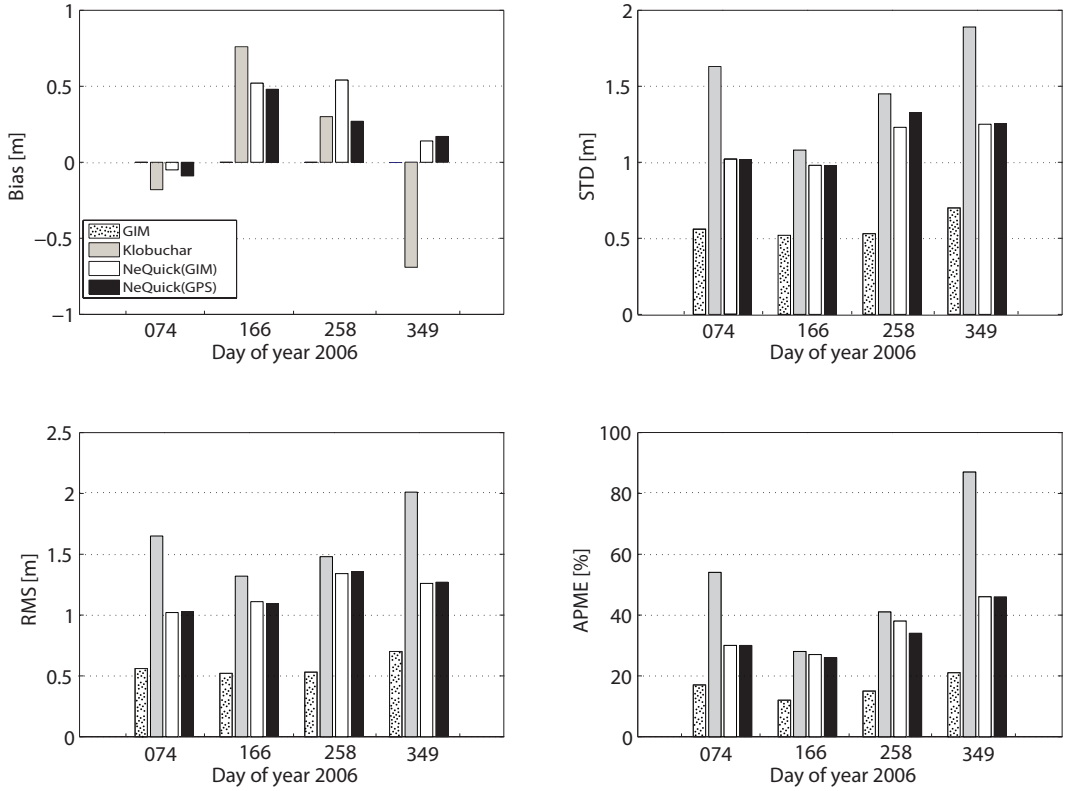

Figure 5.21: Comparison between Mean, STD, RMS, and APME of the GIM, Klobuchar and NeQuick (with the Az parameter provided from both GIM and 20 IGS stations) models errors during 74, 166, 258, 349 days of 2006

ionospheric conditions) in terms of an ionospheric correction model for the GNSS data.

5.8 Concluding remarks

The focus of this chapter was on the NeQuick ionospheric electron density model and estimation of the NeQuick model parameters. Formulation of the model based on the 1994 version was described and some differences between the 1994 NeQuick and the improved version were addressed. In this chapter, it is proposed to employ Global Ionospheric Maps (GIM) for estimation of the effective ionization level (Az parameter) which is a key input parameter for the NeQuick model. We believe that GIM routinely produced by the International GNSS Service is reliable and accurate enough to utilize for estimation of the daily NeQuick model parameters. We showed that the NeQuick model parameters provided from GIM are highly consistent with the nominal approach under all geomagnetic ionospheric conditions.

The alternative approach provides a valuable opportunity not only to study the spatial dependency of the Az parameter but also to investigate the correlation between the Az parameter and the solar radiation intensity. Our results in this chapter show that the second-order polynomial is only the simplest function to express the spatial distribution of the Az parameter against modified dip latitude but it is not the best-fitting function. A

study of the correlation between the Az parameter and F10.7 using their yearly time series from 2003 to 2006 shows that they are not highly correlated (the maximum correlation is less than 0.65). Resulting in that use of the same lower limits for Az as for F10.7 (64) is unreasonable in practice. We also found that the correlation coefficient between Az and F10.7 is always positive and depends on latitude. The correlation coefficient is higher in equatorial regions.

By taking advantage of the availability of the predicted version of GIM from the Center for Orbit Determination in Europe (CODE), one of the IGS analysis centers, predicted NeQuick model parameters for real-time GNSS applications can be computed. Comparing the predicted model parameters with the corresponding values from the nominal approach verified the consistency and validity of the predicted model parameters.

We compared slant ionospheric delays derived from NeQuick, using model parameters computed from both GIM and the nominal approach, with the corresponding delay measured by GPS phase data in three IGS sites. This demonstrates that the accuracy of the NeQuick model is the same using the model parameters from GIM as using model parameters from the nominal approach.

6 Physics-Based Modeling of TEC

In this chapter a physics-based model to describe the regular variation of TEC during a day as function of solar zenith angle is developed. The model is based on the theory of ionosphere formation which was presented in chapter 3.

The physics of the ionosphere is used for the definition of the normal ionosphere with an electron density profile based on the Chapman production function in subsection 3.2.2. In section 6.2 a recursive model for VTEC in the normal ionosphere is introduced and the least squares estimation of the model parameters using VTEC measurements is discussed. The performance of the recursive model is investigated in section 6.3 using GIM maps as input.

6.1 Normal ionosphere

In order to express ionospheric VTEC as a function of the solar incidence zenith angle, the earth's ionosphere is considered as an *ideal* ionized medium that it is defined based on (Chapman, 1931a) and (Chapman, 1931b). The idealized ionosphere is *simple* but *similar* representation of the actual ionosphere. It is:

- An atmosphere with only *two types* of gases (O_2 and O). The more heavy gas (O_2) dominates the lower altitudes and the lighter gas (O atomic oxygen) dominates the higher altitudes.
- An *isothermal* atmosphere where density varies *exponentially* with height.
- An atmosphere with a *spherical stratification* which rotates with the earth, at the same rotation rate.
- An atmosphere where gases are ionized by absorption of the solar radiation (*photoionization process*) and the rate of ion production is determined by the *Chapman production function*: the ion production rate is proportional to the air-density and the intensity of the radiation reaching that point.
- The rate of electron loss at the lower altitudes is proportional to the squared electron density $\alpha n_e^2(h)$ (α quadratic recombination coefficient) and proportional to the electron density $\beta n_e(h)$ (β linear recombination coefficient) at higher altitudes.
- The quadratic and linear recombination coefficients are *independent* of height.

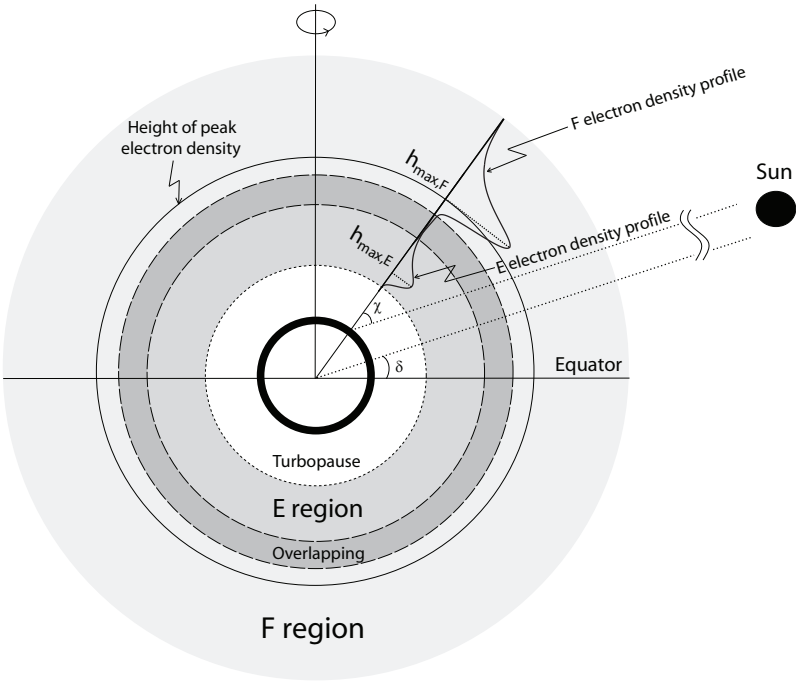

Figure 6.1: Illustration of the normal ionosphere with E- and F-regions.

- No transport of charged particles (electrons and ions). The charged particles do not move from the place which they were formed, this implies that the geomagnetic field plays no role in the formation of the normal ionosphere.

According to above assumptions, the idealized ionosphere consists of two ionized regions: the idealized E-region (lower ionosphere) and the idealized F-region (upper ionosphere). The two regions overlap as shown in figure 6.1. The overlapped region is under the altitude of the peak electron density of the F-region. These two idealized ionized regions define an ideal ionosphere that it refers to a *normal ionosphere* in this thesis.

6.1.1 Vertical electron density profile in the normal ionosphere

The electron density at a given point in the normal ionosphere is computed by summation of the electron densities of the two idealized E- and F-regions

$$n_e(t,h) = n_{e,E}(t,h) + n_{e,F}(t,h) \tag{6.1}$$

with $n_e(t,h)$ the normal electron density at height of h, and with $n_{e,E}(t,h)$ and $n_{e,F}(t,h)$ the electron densities of the normal E- and F-regions. The electron density varies over time because of the earth's rotation. The rotation of the earth results in a changing the earth-sun geometry at a given point in the normal ionosphere. The temporal variation of the normal electron density can be expressed as the sum of the variation of electron densities in the two normal regions

$$\frac{\partial n_e(t,h)}{\partial t} = \frac{\partial n_{e,E}(t,h)}{\partial t} + \frac{\partial n_{e,F}(t,h)}{\partial t} \tag{6.2}$$

The VTEC in the normal ionosphere is the integral of equation 6.1 over height

$$N(t) = \underbrace{\int_0^\infty n_{e,E}(t,h)dh}_{N_E(t)} + \underbrace{\int_0^\infty n_{e,F}(t,h)dh}_{N_F(t)} \tag{6.3}$$

where N stands for the normal vertical total electron content and N_E and N_F are respectively the vertical total electron contents for the normal E- and F-regions. The temporal variation of the VTEC in the normal ionosphere is the integral of equation 6.2 over height

$$\frac{\partial N(t)}{\partial t} = \int_0^\infty \frac{\partial n_e(t,h)}{\partial t}dh = \underbrace{\int_0^\infty \frac{\partial n_{e,E}(t,h)}{\partial t}dh}_{\frac{\partial N_E(t)}{\partial t}} + \underbrace{\int_0^\infty \frac{\partial n_{e,F}(t,h)}{\partial t}dh}_{\frac{\partial N_F(t)}{\partial t}} \tag{6.4}$$

The time derivative of the normal VTEC in the E- and F-regions can be obtained by integration of the electron density continuity equation over height for each layer.

6.1.2 VTEC in the normal E-region

The continuity equation of electron density at a given point in the normal E-region is written

$$\frac{\partial n_{e,E}(\chi,h)}{\partial t} = q(\chi,h) - \alpha\ n_{e,E}^2(\chi,h) \tag{6.5}$$

with α is the quadratic recombination coefficient (constant in time and height) and χ the solar zenith angle at the point. The charged particles can't move in the normal ionosphere consequently the term associated to the transport process is *absent* in the equation 6.5. The solar zenith angle at the point depends on the location of point, local time, and season. Therefore, it can be computed as a function of latitude of the point φ, and the local time t, and the solar declination angle δ, i.e. $\chi = \chi(t,\varphi,\delta)$ as follows

$$\cos\chi = \sin\delta\sin\varphi + \cos\delta\cos\varphi\cos T \tag{6.6}$$

where $T = (t-12)\frac{\pi}{12}$ is the longitude of the point (in radians) measured eastward from the *sun meridian* (or the *noon meridian*) at the local time t in hour ($T = 0$ at local noon time).

In equation 6.5, $q(\chi,h)$ is the ion production rate at solar zenith angle $\chi(t,\varphi,\delta)$ and at height h given by the *Chapman production function* (equation 3.29)

$$q(\chi,h) = q_{max,E}^0\ a(\chi,h) \tag{6.7}$$

with

$$a(\chi,h) = \ \exp\{1 - z - Ch(\chi,h)\ \exp\{-z\}\} \tag{6.8}$$

and with $z = \frac{h-h_{max,E}^0}{H_E}$, H_E is the scale height of the normal E-region, $h_{max,E}^0$ the height of the maximum rate of ion production $q_{max,E}^0$ when the sun is overhead ($\chi = 0$). $Ch(\chi,h)$ is the *Chapman grazing incidence function* (equation 3.28).

In (Chapman, 1931a), the equation 6.5 has been re-written as follows

$$\sigma_E \ \frac{\partial v_E}{\partial T} + v_E^2 \ = \ a(\chi, h) \tag{6.9}$$

with $v_E = \frac{n_{e,E}(\chi,h)}{n^0_{max,E}}$ the dimension free ratio of electron density with respect to the maximum electron density $n^0_{max,E}$ when the sun is overhead $\chi = 0$ and σ_E is a time constant (in radians) related to the decay time. The non-linear first order differential equation 6.9 becomes equivalent to equation 6.5 if σ_E is

$$\begin{aligned} \sigma_E &= \frac{1}{\epsilon \ \sqrt{\alpha \ q^0_{max,E}}} \\ n^0_{max,E} &= \sqrt{\frac{q^0_{max,E}}{\alpha}} \end{aligned} \tag{6.10}$$

with $\epsilon = \frac{43200}{\pi}$ conversion factor (radian to second). To have physical interpretation of σ_E, suppose the electron density at $t_0 = 0$ is $n_{e,E}(t_0)$ and the electron production is switched off at $t = 0$, i.e. $q(t_0) = 0$, the continuity equation 6.5 yields after some re-writing and integration over

$$\int_{n_{e,E}(t_0)}^{n_{e,E}(t)} \frac{dn_{e,E}}{n^2_{e,E}} = -\alpha \int_{t_0}^{t} dt = -\alpha \ (t - t_0) \tag{6.11}$$

The integration on the left hand side leads to

$$\frac{n_{e,E}(t) - n_{e,E}(t_0)}{n_{e,E}(t) n_{e,E}(t_0)} = -\alpha \ (t - t_0) \tag{6.12}$$

The decay time for reducing electron density by a factor $e = 2.718\ldots$ is then

$$\tau_{_E} = \frac{e - 1}{n_{e,E}(t_0) \ \alpha} \tag{6.13}$$

Hence there is no exponential decay in the E-region and the speed of decay depends on the initial electron density. $\tau_{_E}$ is known *lifetime* of electrons in the E-region. Substituting equation 6.10 into 6.13 gives

$$\tau^0_{_E} = \epsilon(e - 1)\sigma_E \ \approx \ 2362.8 \ \sigma_E \ \ \text{[in seconds]} \tag{6.14}$$

where $\tau^0_{_E}$ is the lifetime of electrons when $\chi = 0$. This equation shows that σ_E is proportional to the life time of electrons in the E-region when the sun is overhead. During the day time, the lifetime in the E-region is about 10 seconds, which corresponds to $\sigma_E = 4.2e-4$ radians. These small values for σ_E leads to the E-region being formed in photochemical equilibrium.

Differential equation 6.9 is in form of *Riccati differential equation*. The general solution of the Riccati differential equation can be expressed as follows

$$v_E(T) = v_{E,0} + g(T) \left[C \ + \ \frac{1}{\sigma_E} \int g(T) dT \right]^{-1} \tag{6.15}$$

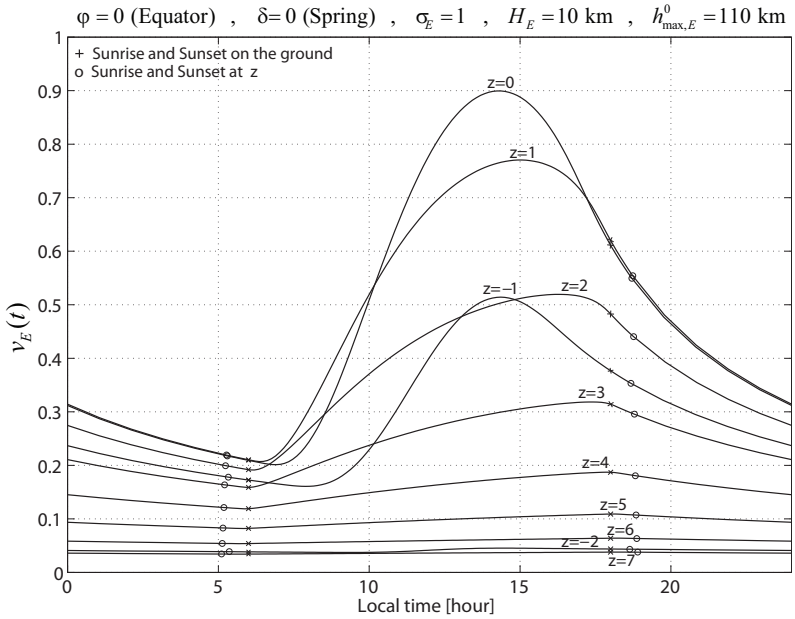

Figure 6.2: Temporal variation of $v_E(t) = \frac{n_{e,E}(\chi,h)}{n^0_{max,E}}$ for various heights in a day at *equator* ($\varphi = 0$) at *equinoxes* ($\delta = 0$) for $\sigma_E = 1$ in radians.

with

$$g(T) = \exp\left\{\frac{-2}{\sigma_E}\int v_{E,0}\, dT\right\} \tag{6.16}$$

where C is an arbitrary constant and $v_{E,0}$ is a particular solution of the Riccati differential equation.

Equation 6.15 represents the variation of v_E during the hours of daylight. During the hours of darkness the right-hand side of differential equation 6.9 must be replaced by zero, this leads to the following solution for night time

$$v_E(T) = \frac{\sigma_E}{T - C} \tag{6.17}$$

where C is an arbitrary constant. It results in the following condition at sunrise and sunset (denoted as the suffixes r and s respectively)

$$\frac{1}{v_{E,r}} - \frac{1}{v_{E,s}} = \frac{T_r - T_s}{\sigma_E} \tag{6.18}$$

The solution of the Riccati differential equation 6.15 has to fulfill this condition, which determines the arbitrary parameter C involved in the general solution 6.15. This boundary value problem (BVP), with $[T_r \;\; T_s]$ as boundaries and equation 6.18 as a boundary condition, is solved using numerical methods.

Solutions of equation 6.15 under boundary condition 6.18 versus local time for various heights (at distance z above and below $h^0_{max,E}$), for $\sigma_E = 1$ radians at equator ($\varphi = 0$) at equinoxes ($\delta = 0$) are shown in figure 6.2. In this figure, the graphs show that for $\sigma_E = 1$

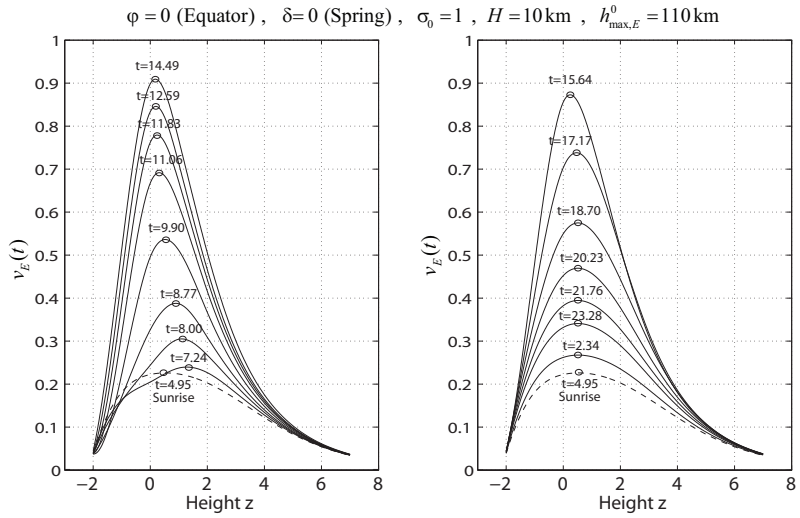

Figure 6.3: Vertical profile of $v_E(t) = \frac{n_{e,E}(\chi,h)}{n^0_{max,E}}$ versus z at *equator* ($\varphi = 0$) for various t (local time in hour) at *equinoxes* ($\delta = 0$) for $\sigma_E = 1$ in radians.

radians the maximum value of $v_E = \frac{n_{e,E}(\chi,h)}{n^0_{max,E}}$ occurs after noon for heights close to $h^0_{max,E}$, at greater heights the maximum occurs near sunset. Since sunrise and sunset times are height dependent, for each line in figure 6.2, sunrise and sunset times on the ground are indicated by a 'cross' and those at height of $h = zH_E + h^0_{max,E}$ indicated by a 'circle'. According to equation 6.17, v_E is decreasing during night time. This decrease is continued for a considerable time interval after sunrise for many heights in figure 6.2. In other words, v_E starts to increase only after a considerable time after sunrise.

The solutions of the BVP for various heights at equator $\varphi = 0$ at equinoxes ($\delta = 0$) have been used to compute the vertical profile of v_E at various local times. Figure 6.3 shows the vertical profile of v_E during two consecutive sunrises at height $h^0_{max,F}$. The plots on the left show the vertical profiles from sunrise to the local time which v_E reaches its maximum. The plots on the right show profiles from the time of maximum v_E to sunrise on the next day. For $\sigma_E = 1$, the maximum of v_E (at all heights) occurs after noon at about local time $14{:}30$. Note that maximum v_E is less than 1, which implies that $n_{e,E}(\chi, h)$ does not reach $n^0_{max,E}$.

Integrating over height

Integrating equation 6.9 over the height h yields

$$\sigma_E \frac{\partial V_E}{\partial T} + \int_0^\infty v_E^2 dh \; = \int_0^\infty a(\chi, h) dh \; = \; A(\chi, H_E, h^0_{max,E}) \tag{6.19}$$

where $V_E(t) = \frac{N_E(t)}{n^0_{max,E}}$ is the ratio between vertical total electron content of the normal E-region N_E and peak electron density of the normal E-region $n^0_{max,E}$ (when the sun is overhead $\chi = 0$). $A(\chi, H_E, h^0_{max,E})$, or is short $A(\chi)$, is a function of solar zenith angle χ, scale height H_E and the height of peak electron density $h^0_{max,E}$ (when $\chi = 0$). V_E gives the *depth* of an idealized E-region which has the same electron content as the normal

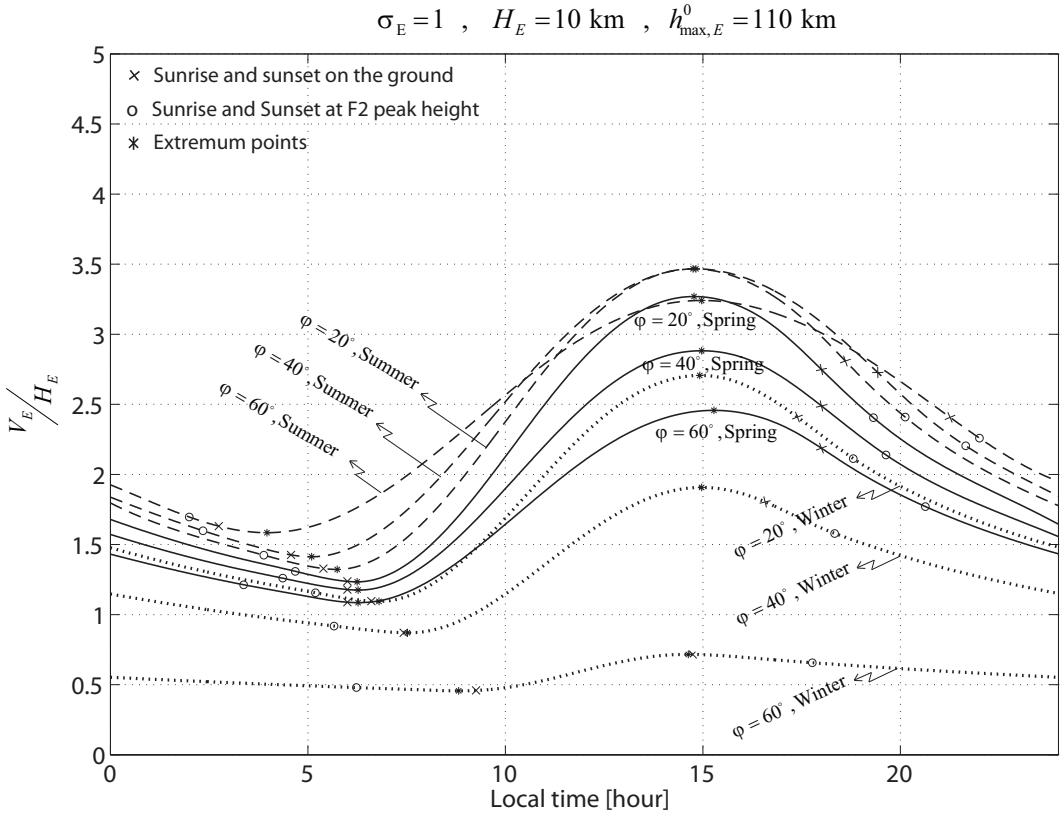

Figure 6.4: Temporal variation of $V_E(t) = \frac{N_E(t)}{n^0_{max,E}}$ (divided by H_E) for various latitudes at *equinoxes* $\delta = 0$ (solid curves) and at *summer solstice* $\delta = +23.5°$ (dashed curves) and at *winter solstice* $\delta = -23.5°$ (dotted curves) for $\sigma_E = 1$ in radians.

E-region but with an *uniform* electron density equal to $n^0_{max,E}$.

Using the definition of the shape factor in equation 4.36 and taking $n^0_{\max} = n_{\max}\sqrt{Ch(h_{max,E},\chi)}$ into account (based on equation 3.39), equation 6.19 can be written in terms of the shape factor as

$$\sigma_E \ \frac{\partial V_E}{\partial T} + \frac{\Omega(\chi)}{\sqrt{Ch(h_{max,E},\chi)}} V_E \ = \ A(\chi) \tag{6.20}$$

where due to time dependency of the shape factor $\Omega(\chi)$ the differential equation 6.20 is a non-linear first order differential equation. Since evaluation of $\int_0^\infty v_E^2 dh$ and $\Omega(\chi)$ are difficult, equation 6.19 or 6.20 are solved by numerical methods.

In fact, $V_E(t) = \int v_E \, dh$, which implies that V_E can be computed numerically from the graphs in figure 6.3. Note that $V_E(t) = H_E \int v_E \, dz$, therefore the area computed from the graphs should be multiplied to the scale height H_E. To study the behavior of V_E, the BVP has been solved for various values of σ_E, namely 1, 0.5, 0.1, 0.04 and 0.01, for different latitudes ($\varphi = 0°,\ 20°,\ 40°,\ 60°$) at equinoxes ($\delta = 0°$) and solstice ($\delta = \pm 23.5°$).

Figure 6.4 shows the temporal variation of V_E (divided by H_E) versus local time in a day for different latitudes at the equinoxes and solstices for $\sigma_E = 1$. The total electron content is equal to $V_E(t)H_E$ in normal E-region. Note that the normal E-region extends to heights

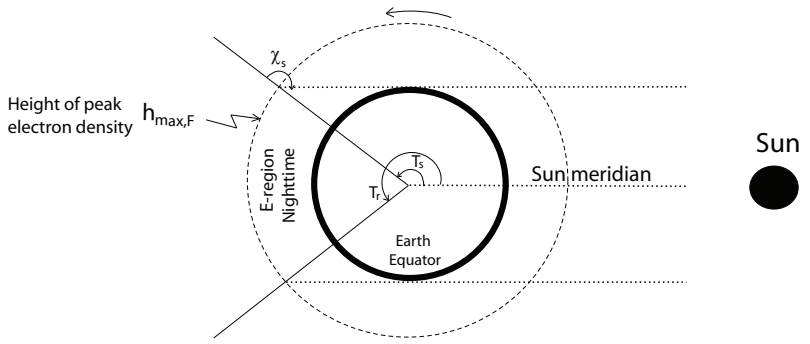

Figure 6.5: Nighttime duration in the normal E-region (in the equator plane).

below the height of peak electron density of the F-region $h_{max,F}$. Hence, it can be assumed that sunrise and sunset time of the normal E-region are the same as those for the height of $h_{max,F}$ (see figure 6.5). In figure 6.4, sunrise and sunset times at $h_{max,F}$ are indicated by a 'circle' and on the ground ($h = 0$) indicated by a 'cross'.

The graphs in figure 6.4 show that the daily pattern of V_E has minimum and maximum points (indicated by '$*$'); the minimum occurs before local noon and the maximum occurs after noon. An increase in latitude causes an increase of solar zenith angle, and as a consequence the level of V_E at day time and night time is decreased when the latitude increases for both equinoxes and winter solstice. For summer solstice, the level of V_E is increased during day time, but it is decreased during night time when the latitude increases. According to equation 6.19, $\frac{\partial V_E}{\partial T}$ becomes negative after the maximum of V_E and its absolute value increases with increasing latitude. For the summer solstice, the variation of the slope versus latitude is quite large, which leads to a change in the level of V_E at different latitudes during night time. As expected the level of V_E (at all latitudes) is higher during the summer solstice (dashed curves) and is lower during the winter solstice (dotted curves) with respect to the level of V_E during the equinoxes (solid curves).

It can be seen from the graphs of V_E for the equinoxes that the time of the minimum point is more or less the same for all latitudes for $\sigma_E = 1$ radians. During summer solstice, the time between the minimum point and local noon is geater for higher latitudes. The opposite is observed during winter solstic. Note that for $\sigma_E = 1$ for many of the graphs, the decrease in V_E is continued a considerable time interval after sunrise. The time of the maximum point depends on latitude in the same way as for equinoxes and solstices. The time between the maximum and local noon is increased by an increase in latitude.

The recombination process in the normal E-region is controlled by σ_E. A large value for σ_E corresponds to a small values for α. The daily pattern of V_E (in unit H_E) for various σ_E values are plotted in figure 6.6. It can be seen that larger values of σ_E (increasing the lifetime of electrons) results in degradation of the recombination term, which allows V_E to continue to increase after noon and the maximum of V_E occurs after noon. The time of maximum V_E departs more from the local noon time for large value of σ_E. Since the recombination term is proportional to squared electron density n_e^2, an increase in electron

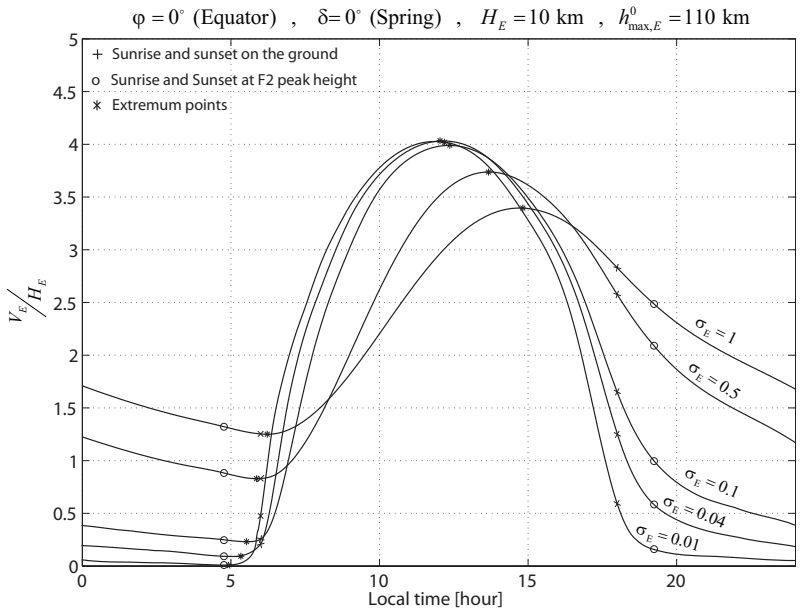

Figure 6.6: Temporal variation of $V_E(t) = \frac{N_E(t)}{n^0_{max,E}}$ (divided by H_E) for various σ_E at *equator* ($\varphi = 0$) at *equinoxes* ($\delta = 0$).

density results in an increase in the recombination term. Because of this, level of the maximum V_E is decreased for large values of σ_E. The minimum of V_E also is affected by σ_E in a way that, for large values of σ_E, the time distance between the minimum and local noon is decreased.

For smaller values of σ_E, which corresponde to larger values of α, the recombination term gets large values. As a consequence, the continuity equation tends to the photochemical equilibrium and the pattern of V_E tends to be symmetric around local noon time.

Photochemical equilibrium condition

If it is assumed that the normal E-region is formed in the *photochemical equilibrium* ($q(\chi, h) = \alpha\ n^2_{e,E}(\chi, h)$) then the normal E-region will be a *Chapman layer* and the vertical electron density profile is described by equation 3.37. Under photochemical equilibrium condition, if the earth does not revolve, i.e. $\frac{dv_E}{dT} = 0$, differential equation 6.20 is reduced to

$$\frac{\Omega(\chi)}{\sqrt{Ch(h_{max,E}, \chi)}}\ V_E\ =\ A(\chi) \tag{6.21}$$

Substitution of equation 3.37 into 4.37 gives the shape factor $\Omega(\chi)$ as follows

$$\Omega(\chi) = \frac{A(\chi)\ \sqrt{Ch(h_{max,E}, \chi)}}{B(\chi)} \tag{6.22}$$

with

$$B(\chi, H_E, h^0_{max,E}) = \int_0^\infty \sqrt{a(\chi, h)}\ dh \tag{6.23}$$

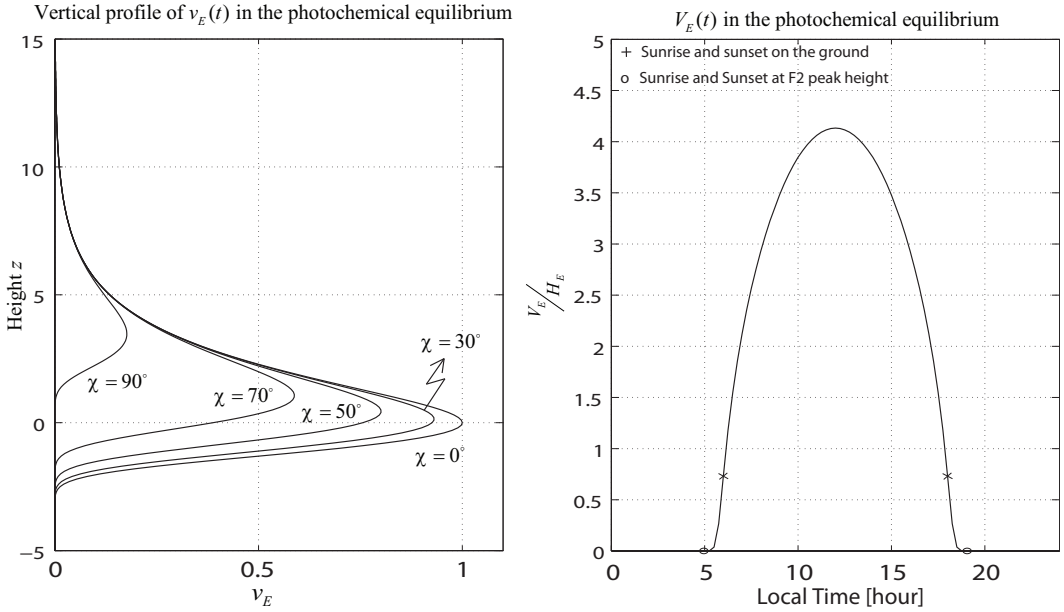

Figure 6.7: (right side) Temporal variation of $V_E(t) = \frac{N_E(t)}{n^0_{max,E}}$ (divided by H_E) in the normal E-region formed in the photochemical equilibrium (a simple Chapman-layer) at *equator* $\varphi = 0$ at equinoxes $\delta = 0$, (left side) vertical profile of $v_E(h) = \frac{n_{e,E}(h)}{n^0_{max,E}}$ in a simple Chapman-layer for different solar zenith angles.

where B is function of solar zenith angle χ, scale height H_E and the height of peak electron density $h^0_{max,E}$ (when $\chi = 0$). For convenience, it is denoted as $B(\chi)$ in this chapter. Substituting equation 6.22 into 6.21 yields

$$V_E(T) \;=\; B(\chi) \tag{6.24}$$

It implies that in the normal E-region under the photochemical equilibrium condition, the vertical total electron content N_E as a function of solar zenith angle χ is described by

$$N_E(T) = n^0_{max,E}\; B(\chi) \;=\; \frac{q^0_{max,E}}{\alpha}\; B(\chi) \tag{6.25}$$

Under the condition the earth does not revolve, N_E will be constant in time. Due to the earth rotating there is a daily variation of $N_E(T)$. To plot pattern of V_E under the photochemical equilibrium $B(\chi)$ was evaluated by the numerical integration method on the equator ($\varphi = 0$) at *equinoxes* ($\delta = 0$) and the result (divided by H_E) is shown in the right side of figure 6.7. It is clear that under photochemical equilibrium condition the maximum value of V_E, in the normal E-region, occurs at *local noon-time* (maximum solar zenith angle) and its pattern is *symmetric* with respect to the local noon-time. V_E is zero for nighttime which implies that the normal E-region (under photochemical equilibrium condition) disappeares between sunset and sunrise. Comparing the pattern of V_E in the photochemical equilibrium with the patterns in figure 6.6 shows that for decreasing values of σ_E the pattern of V_E tends to be similar to photochemical equilibrium. On the left side of figure 6.7, the vertical profile of $v_E(h) = n_e(h)/n^0_{max,E}$ (a Chapman-layer) is shown for different solar zenith angles. It can be seen that, the maximum value of $v_E(h)$ is increased and the height of maximum v_E is decreased when the solar zenith angle is increased.

6.1.3 VTEC in the normal F-region

In this subsection, an expression for the vertical total electron content of the normal F-region as a function of solar zenith angle is derived. The loss of electrons in the F-region is proportional to the electron density. Disregarding transport processes, the electron density continuity equation is written as

$$\frac{\partial n_{e,F}(\chi,h)}{\partial t} = q(\chi,h) - \beta\ n_{e,F}(\chi,h) \tag{6.26}$$

where β is the linear recombination coefficient in $[s^{-1}]$ (constant in time and in height) and $q(\chi,h)$ denotes the ion production rate at time t and at height h from equation 6.7. If local time in seconds, t is expressed in terms of the longitude measured eastward from the noon meridian T in radians, $t = T/\epsilon + 12 \cdot 3600$, the continuity equation can be rewritten as

$$\epsilon\frac{\partial n_{e,F}(\chi,h)}{\partial T} = q^0_{max,F}\ a(\chi,h) - \beta\ n_{e,F}(\chi,h) \tag{6.27}$$

with $\epsilon = \pi/43200$ the conversion factor from second to radian. The function $a(\chi,h)$ is

$$a(\chi,h)\ =\ \exp\{1 - z - Ch(\chi,h)\ \exp\{-z\}\} \tag{6.28}$$

with $z = \frac{h-h^0_{max,F}}{H_F}$, H_F is the scale height of the normal F-region and $h^0_{max,F}$ the height corresponding to the maximum rate of ion production $q^0_{max,F}$ (when $\chi = 0$). If we define $v_F = \frac{n_{e,F}(\chi,h)}{n^0_{max,F}}$ as the ratio of electron density with $n^0_{max,F} = \frac{q^0_{max,F}}{\beta}$ the maximum electron density in the normal F-region when the sun is overhead $\chi = 0$, then equation 6.27 can be written as follows

$$\sigma_F\frac{\partial v_F}{\partial T}\ +\ v_F = a(\chi,h) \tag{6.29}$$

where $\sigma_F = \frac{\epsilon}{\beta}$ (in $[radians]$) is constant in time and in height with $\tau = \frac{\sigma_F}{\epsilon} = \frac{1}{\beta}$ known as the *recombination time constant* in the F-region. In fact τ is a measure for the *lifetime* of electrons in the F-region: It is the time it takes for the electron density $n_{e,F}(h)$ to be reduced to $\frac{n_{e,F}(h)}{e}$ when electron production is swiched off. Hence σ_F is proportional to the lifetime of electrons in the F-region.

The general solution of the linear first order differential equation 6.29 in terms of time t measured from noon meridian (local time minus 12 hours) is obtained as follows

$$v_F(t) = e^{-\beta t}\left(\beta\int_0^t e^{\beta t'}\ a(\chi,h)\ dt' + C\right) \tag{6.30}$$

where C is an arbitrary constant. This solution represents the variation of v_F during the hours of *daylight*; during the hours of *darkness* at height h the right-hand side of equation 6.29 must be replaced by zero, which leads to following differential equation

$$\sigma_F\frac{\partial v_F}{\partial T}\ +\ v_F = 0 \tag{6.31}$$

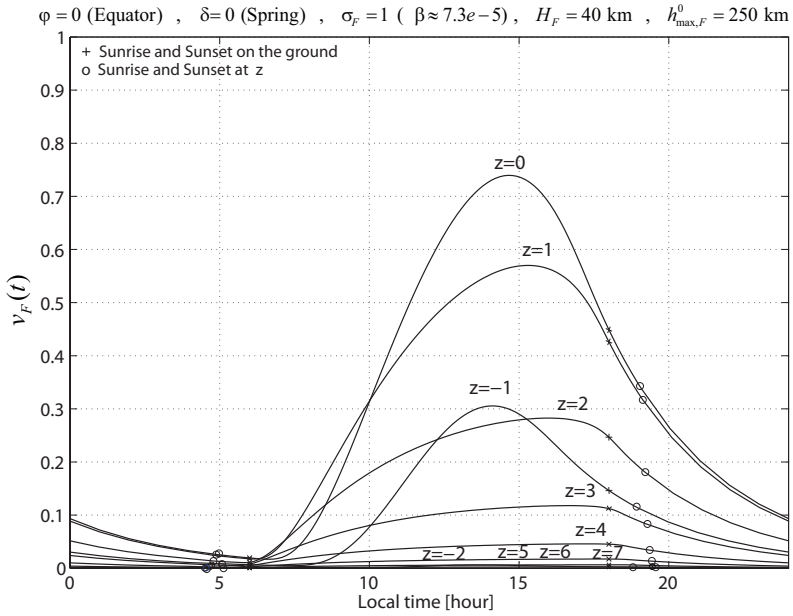

Figure 6.8: Temporal variation of $v_F(t) = \frac{n_{e,F}(\chi,h)}{n^0_{max,F}}$ for various heights in a day at *equator* ($\varphi = 0$) at *equinoxes* ($\delta = 0$) for $\sigma_F = 1$ in radians.

with solution

$$v_F(t) = Ce^{-\beta t} \tag{6.32}$$

If the time of sunrise and sunset at height h are denoted by t_r and t_s, the following condition is obtained

$$v_r = v_s \; e^{-\beta(t_r - t_s)} \tag{6.33}$$

with v_r and v_s the function values of v_F for sunrise and sunset at height h. This condition also implies that during night time, the normal F-region does not disappear, but the ionization process is stopped and only the recombination process is acting. The particular solution for day time should be found subject to the condition of equation 6.33 which determines the arbitrary constant C involved in the general solution. This boundary value problem has been solved for the equator $\varphi = 0^\circ$ at equinoxes $\delta = 0^\circ$ with $\sigma_F = 1$ radians (corresponding to the lifetime of electrons $\tau = 3.8$ hours). Figure 6.8 shows the solutions for different heights. In this figure, the sunrise and sunset points for height z are denoted by a 'circle' and those of $h = 0$ (on the ground) denoted by a 'cross'. At the equator, for $\sigma_F = 1$ the maximum value of $v_F = \frac{n_{e,F}(\chi,h)}{n^0_{max,F}}$ occurs distinctly after noon and is actually closer to sunset for larger heights.

Vertical profiles of v_F, using the solutions of the BVP at different heights, are shown in figure 6.9 for different local times. It can be seen that for $\sigma_F = 1$ at the equator, $n_{e,F}$ reaches its maximum at local time about $t = 14:53$ at the level of $z = 0.15$, which is about 6 km above height $h^0_{max,F} = 250\ km$.

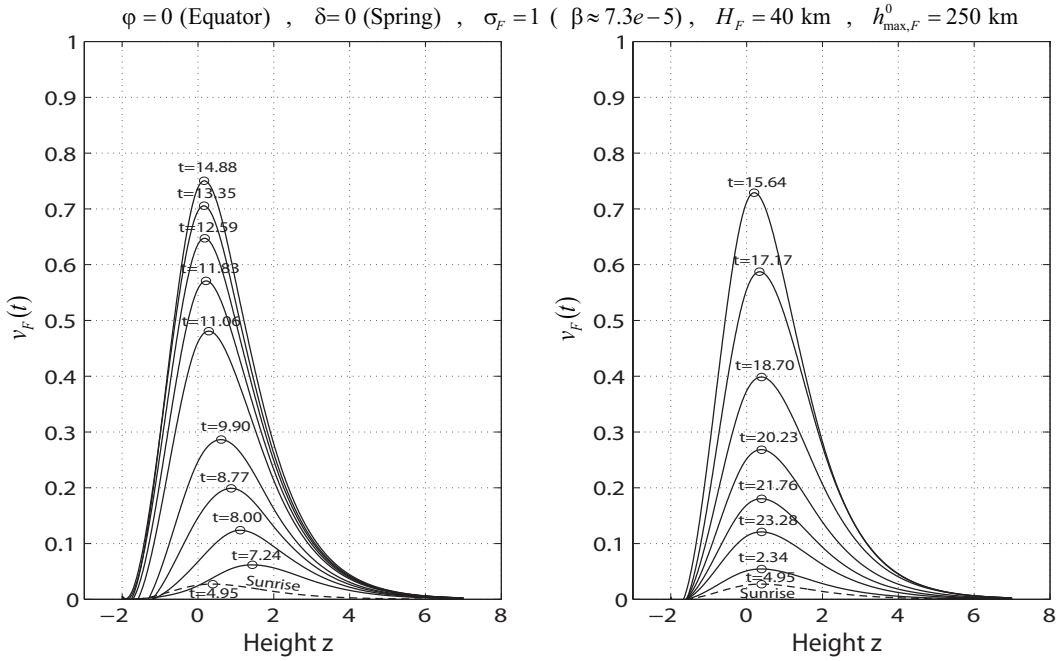

Figure 6.9: Vertical profile of $v_F(t) = \frac{n_{e,F}(\chi,h)}{n^0_{max,F}}$ versus z at *equator* ($\varphi = 0$) for various t (local time in hour) at *equinoxes* ($\delta = 0$) for $\sigma_F = 1$ in radians.

Integration over height

Integrating equation 6.29 over height h results in following differential equation

$$\sigma_F \frac{\partial V_F}{\partial T} \; + \; V_F = A(\chi) \tag{6.34}$$

with $V_F = \frac{N_F}{n^0_{max,F}}$ the rate of vertical total electron content in the normal F-region (denoted as N_F) with respect to the peak electron density $n^0_{max,F}$. V_F gives the *depth* of an hypothetical F-region with an *uniform* electron density equal to $n^0_{max,F}$.

The general solution for V_F in terms of t (local time measured from noon meridian) is given as

$$V_F(t) = e^{-\beta t} \left(\beta \int_0^t e^{\beta t'} \; A(\chi) \; dt' + C \right) \tag{6.35}$$

which results in a solution of $N_F(t)$

$$N_F(t) = n^0_{max,F} \; V_F(t) \; = \; n^0_{max,F} \; e^{-\beta t} \left(\beta \int_0^t e^{\beta t'} \; A(\chi) \; dt' + C \right) \tag{6.36}$$

Note that $A(\chi)$ is shorthand $A(\chi, H_F, h^0_{max,F})$, a function of scale height H_F, he height of peak electron density $h^0_{max,F}$ (when $\chi = 0$) and solar zenith angle χ. To evaluate $V_F(t)$ over a day, it is necessary evaluate $A(\chi)$, which is defined as follows

$$A(\chi) = \int_0^\infty a(\chi, h) \; dh \; = \; \int_0^\infty \exp\{1 - z - Ch(\chi, h) \; \exp\{-z\}\} \, dh \tag{6.37}$$

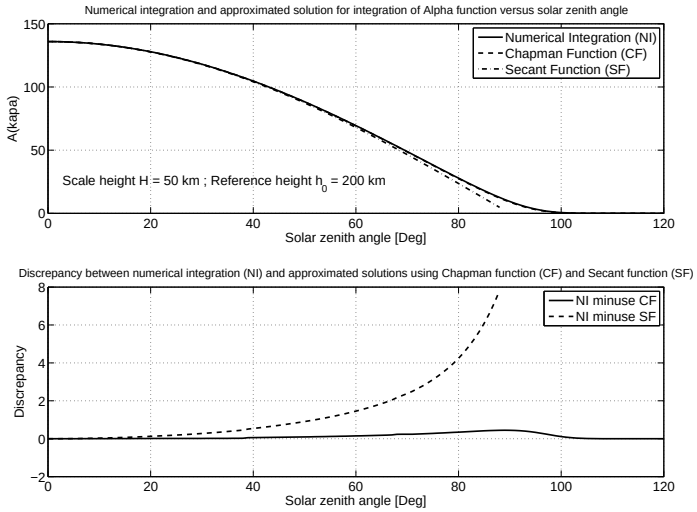

Figure 6.10: (Top) $A(\chi)$ function versus χ provided by numerical integration (denoted as NI) and using approximated solution with $\sec\chi$ (denoted as SF) and the approximated solution with $Ch(\chi, h)$ (denoted as CF), (Bottom) approximation errors in SF and CF with respect to NI.

with $\chi = \chi(t, \varphi, \delta)$. In (Feltens, 1998), using $\sec\chi$ instead of the Chapman grazing function, which is acceptable for low solar zenith angles $\chi \leq 70^\circ$, it is shown that

$$\int a(\chi, h)\ dh\ = H_F \cos\chi \exp\{1 - \sec(\chi)\ \exp\{-z\}\} \qquad (\chi \leq 70^\circ) \tag{6.38}$$

Using this equation, integration of $a(\chi, h)$ over h in terms of the Chapman grazing function can be approximated as follows

$$\int a(\chi, h)\ dh\ \approx \frac{H_F}{Ch(\chi, h)} \exp\{1 - Ch(\chi, h)\ \exp\{-z\}\} \tag{6.39}$$

where the height dependency of $Ch(\chi, h)$ is considered. Unlike $Ch(\chi, h)$, $\sec\chi$ is not height dependent, which results in errors in equation 6.38 and 6.39 for high values of χ. In figure 6.10, $A(\chi)$ is computed using numerical integration (NI) of equation 6.37 and computed with the approximations of equations 6.38 (SF) and 6.39 (CF). As expected, the difference between SI and NI is large for high solar zenith angles. The differences between CF and NI are quite small in general except for small excursions around 90°. Therefore, equation 6.39 is a good approximation of equation 6.37.

$A(\chi)$ is only zero when $\chi = 180^\circ$. This occurs only at midnight for latitudes in the equatorial region where the solar zenith angle reaches zero at local noontime in specific period of the year. For other latitudes the time interval between sunset and sunrise tends to zero when height goes to infinity. If the height of the upper limit of the F-region is considered as infinity, then the duration of the night in the F-region will be zero which results in photoionization during the whole day (daytime and nighttime). However, values of $A(\chi)$ for higher solar zenith angles ($\chi > 120^\circ$) are very small (less than $\sim 10^{-8}\ km$), which results in very low level photoionization. From this follows that differential equation 6.34 is valid for the entire range of T, i.e. from zero to 2π. Disregarding daily variation

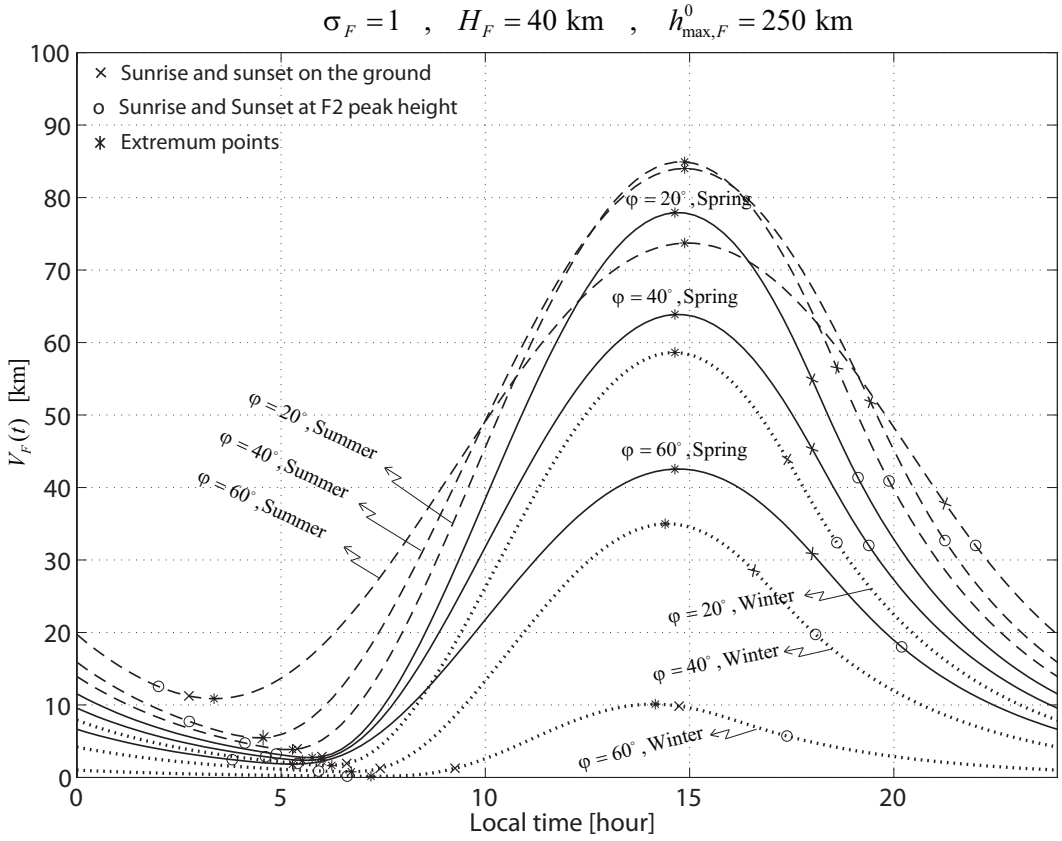

Figure 6.11: Temporal variation of $V_F(t) = \frac{N_F(t)}{n^0_{max,F}}$ in $[km]$ for various latitudes at *equinoxes* $\delta = 0$ (solid curves) and at *summer solstice* $\delta = +23.5^\circ$ (dashed curves) and at *winter solstice* $\delta = -23.5^\circ$ (dotted curves) for $\sigma_F = 1$ in radians.

of the VTEC due to changing in the solar declination angle δ (seasonal variation), one can assume that the temporal variation of VTEC is repeated everyday. This assumption leads to the condition $V_F(T = 0) = V_F(T = 2\pi)$, so that C in equation 6.35 can be obtained subject to this condition. Solving equation 6.34 under condition $V_F(T = 0) = V_F(T = 2\pi)$ results in a Boundary Value Problem (BVP). In order to study variability of V_F in time and space, the BVP for various values of σ_F for different latitudes at the equinoxes and the solstices have been solved using numerical methods.

Figure 6.11 shows the daily pattern of V_F for various latitudes at equinoxes ($\delta = 0$) and solstices ($\delta = \pm 23.5$) for $\sigma_F = 1$ radians, which corresponds to a lifetime of electrons $\tau_F = 3.8$ hours ($\beta = 7.3e - 5$). In this figure, sunrise and sunset times at level $h_{max,F}$ are denoted by a 'circle' and on the ground ($h = 0$) denoted by a 'cross'. The graphs show that daily pattern of V_F has minimum and maximum points (denoted by '$*$'), and the minimum occurs before local noon and the maximum occurs after noon. As expected, at all latitudes, the level of V_F is higher for the summer solstice and it is lower for the winter solstice with respect to those for the equinoxes. With increasing latitude, therefore increasing solar zenith angle, the level of V_F is decreased during *day time* for the equinoxes and solstices. The descent of V_F is faster for large values of V_F after the maximum. Hence, for high latitudes where the level of V_F is lower, the slope of the pattern is small for the equinoxes and solstices. For the summer solstice, the variation of the slope versus latitude

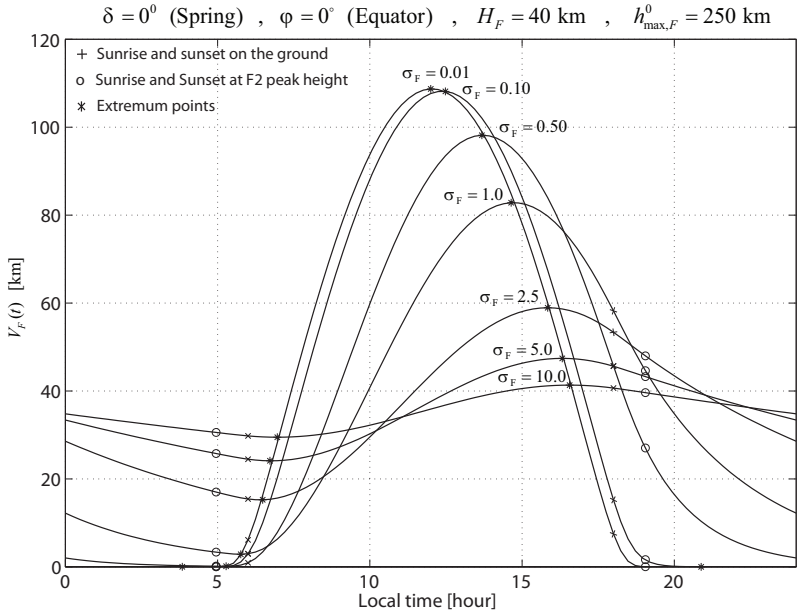

Figure 6.12: Temporal variation of $V_F(t) = \frac{N_F(t)}{n^0_{max,F}}$ in $[km]$ for various σ_F values at *equator* ($\varphi = 0$) at *equinoxes* ($\delta = 0$).

is quite large. Because of this, the level of V_F is increased for *night time* the latitude increases. For the equinoxes and winter solstice, the level of V_F is decreased for night time when latitude is increased.

For $\sigma_F = 1$, at equinoxes, for all latitudes the peak of V_F occurs after noon at the same time $14{:}43$. For the summer and winter solstices, time of the peak depends on the latitude. For summer solstice, the time distance between local noon and the peak is increased for large latitudes (due to increasing in sunlight duration). For winter solstice, the duration of sunlight is decreased with increasing latitude and therefore the oppesit effect is observed. As can be seen from the graphs, the time of the minimum point of V_F is changed considerably with respect to latitude and season. For $\sigma_F = 1$ at equinoxes, the minimum of V_F occurs before sunrise on the ground. An increase in latitude increases the sunlight duration in the F-region, consequently electron production is started earlier at higher latitudes.

The daily pattern of V_F at the equator at equinoxes is shown in figure 6.12 for various values of σ_F. For larger values of σ_F (corresponding to smaller recombination coefficient β), the life time of electrons is longer which leads to continuation of the accumulation of electrons after noon time. Because of this, for large values of σ_F, the maximum point of the pattern of V_F departs more from local noon time and for small values like $\sigma_F = 0.01$ (or $\beta \approx 7.3e-3$, corresponding to the lifetime $\tau_{_F} = 137$ second) the maximum occurs at local noon time. Although for larger values of σ_F more electrons are accumulated, also the recombination βn_e is increased and hence the level of V_F is decreased during day time. According to equation 6.34, the derivative of V_F is proportional to inverse of σ_F. Hence, the descent of V_F after the maximum, speeds up for small value of σ_F. As a consequence, in figure 6.12, the level of V_F during night time for large value of σ_F is higher. Also the decrease in the slope of V_F results in the minimum to be later after sunrise on the ground for large values of σ_F.

6.1.4 Combined VTEC of the normal ionosphere

So far, we achieved vertical total electron content of the normal E- and F-regions individually. The entire VTEC for the normal ionosphere is obtained by integration equation 6.1 over height

$$N(t) = \int_0^\infty n_e(t,h)dh \;=\; \underbrace{\int_0^\infty n_{e,E}(t,h)dh}_{N_E} \;+ \underbrace{\int_0^\infty n_{e,F}(t,h)dh}_{N_F} \tag{6.40}$$

where N denotes the entire normal VTEC. It is assumed that the normal E-region is formed under photochemical equilibrium condition. Substituting N_E and N_F from respectively equations 6.25 and 6.36 yields

$$N(t) \;=\; n^0_{max,E}\; B(\chi) \;+\; \beta\; n^0_{max,F}\; e^{-\beta t} \int_0^t e^{\beta t'}\; A(\chi)\; dt' + C \tag{6.41}$$

where C is an arbitrary constant. This is a continuous function that describes the behavior of the total electron content in the normal ionosphere when the E-region is formed in the photochemical equilibrium. This function will be used as a basis for derivation of a novel ionospheric model for temporal variability of the vertical TEC, see section 6.2.

6.1.5 Slant TEC in the normal ionosphere

To complete this section, the *slant* TEC of the normal F-region is addressed in this subsection. In order to derive slant TEC along a signal path ρ between a given receiver P on the ground and satellite, the continuity equation 6.29 should be integrated along the signal path. It leads to following differential equation

$$\sigma_F \frac{\partial W_F}{\partial T} \;+\; W_F = \int_0^\rho a(\chi(h,\xi_{_P},Az),h)\; d\rho = A(\chi_{_P},\xi_{_P},Az) \tag{6.42}$$

with $W_F = \frac{S_F}{n^0_{max,F}}$ the rate of slant total electron content in the normal F-region (denoted as S_F) and $n^0_{max,F}$, $\chi_{_P}$ and $\xi_{_P}$ solar and satellite zenith angles on the ground ($h = 0$), Az azimuth of the signal path.

The geometry of the slant signal path between receiver and satellite is depicted in figure 6.13. This figure indicates that the relation between differential increments $d\rho$ along slant direction and dh along vertical direction is $d\rho = \frac{dh}{\cos\xi(h)}$, with $\xi(h)$ computed from equation 4.34 for each point along the signal path. Hence, $A(\chi_{_P},\xi_{_P},Az)$ in equation 6.42 can be re-written in terms of height h as follows

$$A(\chi_{_P},\xi_{_P},Az) = \int_0^\infty a(\chi(h,\xi_{_P},Az),h) \frac{dh}{\sqrt{1-\left(\frac{R_e}{R_e+h}\right)^2\;\sin\xi_{_P}{}^2}} \tag{6.43}$$

where the solar zenith angle is function of satellite zenith angle and azimuth and height along the signal path. To show the solar zenith angle as a function of height, the geometry

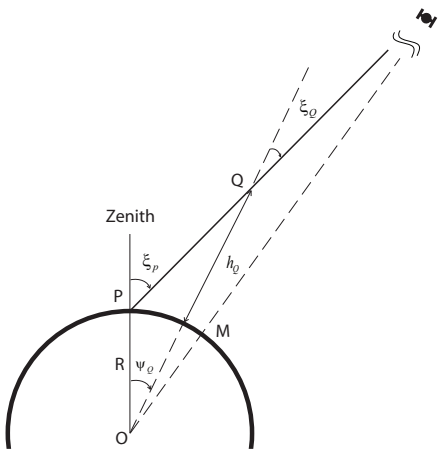

Figure 6.13: Satellite zenith angle variation along a slant signal path.

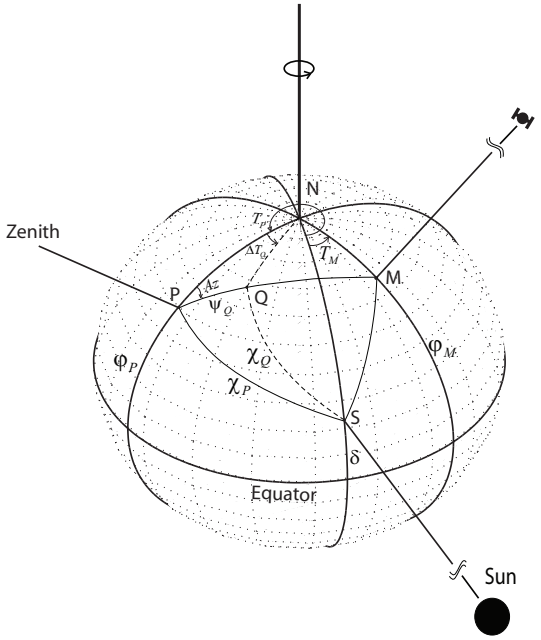

Figure 6.14: Geometrical illustration of the solar zenith angle variation along a slant signal path.

between receiver and satellite and the sun is illustrated in a spherical system in figure 6.14. In this figure, the arc between receiver P and M indicates vertical projection of the signal path on the sphere and Q indicates the projection of an arbitrary point with height h in the signal path. T is the longitude measured eastward from the sun meridian and ψ_Q denotes the geocentric angle between P and Q, which as indicated in figure 6.13, can be expressed as a function of height

$$\psi_Q(h) = \xi_P - \xi_Q(h) \tag{6.44}$$

where $\xi_Q(h)$ is computed by equation 4.34. Using the *cosine equation* in the spherical geometry for spherical triangle NPQ, latitude of Q is given as

$$\sin\phi_Q = \sin\phi_P \cos\psi_Q + \cos\phi_P \sin\psi_Q \cos Az \tag{6.45}$$

and the *spherical sine equation* gives ΔT as

$$\sin\Delta T_Q = \frac{\cos\phi_Q}{\sin Az}\sin\psi_Q \tag{6.46}$$

with $\Delta T_Q = T_Q - T_P$. It is clear from equations 6.44, 6.45 and 6.46 that ϕ_Q and T_Q along the signal path are function of height and azimuth and satellite zenith angle. Resulting in the solar zenith angle χ_Q, computed from equation 6.6, is also function of height and azimuth of satellite zenith angle.

If it is assumed that the satellite is not moving and the signal path is fixed in space the general solution of the differential equation 6.42 is obtained as

$$W_F(t) = e^{-\beta t}\left(\beta\int_0^t e^{\beta t'}\ A(\chi_P(t'), \xi_P, Az)\ dt' + C\right) \tag{6.47}$$

where t is local time measured from noon meridian and C is an arbitrary constant. In fact, $W_F(t)$ describes temporal evolution of the electron content in the normal F-region in terms of the azimuth and zenith angle of the signal path.

Equation 6.47 is theoretically of interest because it expresses the slant TEC in an arbitrary direction without taking into account a mapping function. However, the practical evaluation of this equation is complicated due to complexity of the integrand. This an expense that one should pay not to use a mapping function for modeling of the slant TEC. One has to deal with the same problem in the E-region. Therefore, in the subsequent procedure for TEC modeling, we derive a novel model for a vertical TEC using equation 6.41.

6.2 Recursive model of VTEC in the normal ionosphere

In this section, the goal is to derive a *recursive model* for prediction of the temporal variation of the total electron content in the normal ionosphere. It is assumed that the normal E-region is formed in the *photochemical equilibrium*. Therefore, *continuous* expression for temporal VTEC variation in the normal ionosphere is given by equation 6.41. For temporal prediction purposes it is necessary to obtain a *discrete* expression for the temporal variation of the VTEC.

Discrete expression of VTEC

The equivalent discrete solution of equation 6.36 between two consecutive epochs t_{i-1} and t_i can be written as follows

$$N_F(t_i) = e^{-\beta\Delta t} N_F(t_{i-1}) \; + \; \beta \; e^{-\beta\Delta t} \; n^0_{max,F} \int_{t_{i-1}}^{t_i} e^{\beta\Delta t} \; A(\chi) \; dt \tag{6.48}$$

where $\Delta t = t_i - t_{i-1}$ is time interval between the epochs in seconds. For simplicity, we used the average of zenith angles between two consecutive epochs, so that the function of $A(\chi)$ can be taken out of the integration. The analytical solution of the integration becomes

$$\int_{t_{i-1}}^{t_i} e^{\beta(t-t_{i-1})} \; dt \; = \; \frac{1}{\beta} \left(e^{\beta(t_i - t_{i-1})} - 1 \right) \tag{6.49}$$

Substituting equation 6.49 into 6.48 yields

$$N_F(t_i) = e^{-\beta\Delta t} N_F(t_{i-1}) \; + \; n^0_{max,F} \; A(\bar{\chi}) \left(1 - e^{-\beta\Delta t} \right) \tag{6.50}$$

where $\bar{\chi} = \frac{\chi(t_i) + \chi(t_{i-1})}{2}$. The discrete form of equation 6.41 is obtained by using equation 6.50

$$N(t_i) \; = \; n^0_{max,E} \; B(\chi) \; + \; e^{-\beta\Delta t} N_F(t_{i-1}) \; + \; n^0_{max,F} \; A(\bar{\chi}) \left(1 - e^{-\beta\Delta t} \right) \tag{6.51}$$

According to equation 6.40, $N_F(t) = N(t) - N_E(t)$, which allows to write equation 6.51 as follows

$$\begin{aligned} N(t_i) \; &= \; e^{-\beta\Delta t} N(t_{i-1}) \; + \; n^0_{max,F} \; A(\bar{\chi}) \left(1 - e^{-\beta\Delta t} \right) \\ &+ \; n^0_{max,E} \; \left[B\left(\chi(t_i)\right) - e^{-\beta\Delta t} B(\chi(t_{i-1})) \right] \end{aligned} \tag{6.52}$$

Equation 6.52 is a *discrete* and *recursive* expression of the temporal variability of the entire total electron content in the normal ionosphere. In fact, it is as a function of the solar zenith angle. The first term on the right side of equation 6.52 shows the *electron content* from the previous epoch t_{i-1} that survived the recombination process, the second and third terms express the *appearance* of the new electrons over the time interval $\Delta t = t_i - t_{i-1}$, respectively in the normal E-region and normal F-region. The equation is for both daytime and nighttime. At night, both $B(\chi)$ and $A(\chi)$ tend to zero and the recombination process is the dominating process. Note that $A(\chi)$ tends to zero later than $B(\chi)$, because in the F-region night is started later than in the E-region.

Equation 6.52 can be used as a *physics-based model* of the ionosphere, which describes the temporal variation of VTEC due to the earth's rotation and revolution around the Sun.

6.2.1 Parametrization of the VTEC model

Several physical parameters are involved in equation 6.52 which need to be specified. The maximum electron densities at the normal E-region and F-region, when the sun is overhead $\chi = 0$, are respectively $n^0_{max,E} = \sqrt{\frac{q^0_{max,E}}{\alpha}}$ and $n^0_{max,F} = \frac{q^0_{max,F}}{\beta}$. According to equation 3.18, when $\chi = 0$, the maximum rate of ion production for the E- and F- regions are respectively obtained as (in $[\frac{electron}{s\ km\ m^2}]$)

$$q^0_{max,E} = \frac{\eta_E\ \phi^{photon}_\infty}{H_E\ e} \tag{6.53}$$

$$q^0_{max,F} = \frac{\eta_F\ \phi^{photon}_\infty}{H_F\ e} \tag{6.54}$$

where e is the Napierian number $(2.718\cdots)$, η_E and η_F are respectively the ionization efficiencies for O_2 and O gases (in $[\frac{electron}{photon}]$), ϕ^{photon}_∞ denotes the solar photon radiation flux on top of the atmosphere (in $[\frac{photon}{m^2\ s}]$) and H_E and H_F are scale height of the E- and F-regions (in $[km]$). Since we assumed that the normal ionosphere is formed by an *isothermal atmosphere*, H_E and H_F do not depend on height if the vertical gradient of the earth's gravitational acceleration is ignored (see equation 2.7). As mentioned in subsection 3.2.2 for an atomic gas and for wavelengths less than the minimum wavelength of the ionizing radiation of the atom, the ionization efficiency $\eta_F \approx \eta_E = 1$. Using this assumption and taking equations 6.53 and 6.54 into account, equation 6.52 yields

$$\begin{aligned} N(t_i) \ =\ & e^{-\beta\Delta t} N(t_{i-1}) \ +\ \tfrac{1}{e\ \beta\ H_F}\ A(\bar{\chi})\left(1 - e^{-\beta\Delta t}\right)\ \phi^{photon}_\infty \\ & +\ \tfrac{1}{\sqrt{e\ \alpha\ H_E}}\ \left[B\left(\chi(t_i)\right) - e^{-\beta\Delta t} B(\chi(t_{i-1}))\right]\ \sqrt{\phi^{photon}_\infty} \end{aligned} \tag{6.55}$$

In equation 6.55, the first term gives the number of electrons in the vertical column which survive the recombination afte Δt. This term shows the *persistent electrons*. The second term, which is proportional to the solar photon radiation flux ϕ^{photon}_∞, gives the appearance of new electrons in the normal F-region. The third term, which is proportional to $\sqrt{\phi^{photon}_\infty}$, can be interpreted as the appearance of *new* electrons in the normal E-region. Summation of the second and third terms gives total number of the produced new electrons.

In equation 6.55, N or VTEC is computed in $[\frac{electron}{m^2}]$ with time t in *seconds*, the scale heights in *km*, $A(\chi)$ and $B(\chi)$ are in $[km]$. The unit of β is $[\frac{1}{s}]$ and α is $[\frac{m^3}{electron\ s}]$.

One can consider the solar photon radiation flux ϕ_{∞}^{photon}, the linear recombination coefficient β and the quadratic recombination coefficient α as *unknown parameters* in equation 6.55. The scale heights H_F and H_E can be considered as known parameters, and two functions values $A(\chi)$ and $B(\chi)$ can be computed by numerical integration for solar zenith angle $\chi(t, \varphi, \delta)$ for a given time t, season δ and latitude φ. Note that for evaluating $A(\chi)$ and $B(\chi)$, $h^0_{max,E}$ and $h^0_{max,F}$ are required as additional known parameters.

6.2.2 Providing the model parameters

$h^0_{max,E}$ and $h^0_{max,F}$

Due to spatial and time dependency of the solar zenith angle the height of the peak electron density is time and space dependent, i.e. $h_{max} = h_{max}(\chi)$ with $\chi(t, \varphi, \delta)$. In the real earth's ionosphere, due to *laterally variation* of the atmospheric density, the spatial dependency of the height of peak electron density is amplified. Therefore, we assume h^0_{max} (when $\chi = 0$) is space dependent, but constant over time of a day.

In order to take the spatial dependency of h^0_{max} into account, we assume that $h^0_{max,F} = h_{max,F}(\chi_{min})$. The height of peak electron density h_{max,F_2} can be determined as a function of the F2-region propagation factor $M(3000)F2$ (denoted as M) (Schmid et al., 1973)

$$h_{max,F_2} = 1346.92 - 526.40\ M + 59.825\ M^2 \tag{6.56}$$

where h_{max,F_2} is in km. The F2-region propagation factor M is function of UT and location and the monthly average sunspot number R_{12}, which can be computed from the ITU-R (International Telecommunication Union) maps in the mode used by the International Reference Ionosphere (IRI). More details of these maps are given in subsection 5.3.1. Since the solar zenith angle reaches its minimum at local noon time M should be computed from CCIR maps for UT corresponding to local noon time at given point. Then $h_{max,F}(\chi_{min})$ can be determined from 6.56.

For the E-region we assume a fixed height of peak electron density as $h_{max,E} = 110\ km$.

Scale heights

In order to obtain an appropriate value for the scale height of the normal F2-region, we assumed that the scale height of the normal F-region is equal to the scale height at the height of the peak electron density of the F-region when the solar zenith angle is minimum χ_{min}, i.e. $H_F = H_{F_2}(h_{max,F_2}(\chi_{min}))$. According to (Cappellari et al., 1976), the scale height at the F2-region can be expressed as a function of the height of maximum electron density as follows

$$H_{F_2} = \frac{5}{3}\left[30 + 0.2\left(h_{max,F_2} - 200\right)\right] = \frac{h_{max,F_2} - 50}{3} \tag{6.57}$$

where H_{F_2} is in km. So, first tye height of maximum electron density at minimum solar zenith angle $h_{max,F_2}(\chi_{min})$ is computed using equation 6.56, and then $H_{F_2}(h_{max,F_2}, \chi_{min})$ can be determined from equation 6.57.

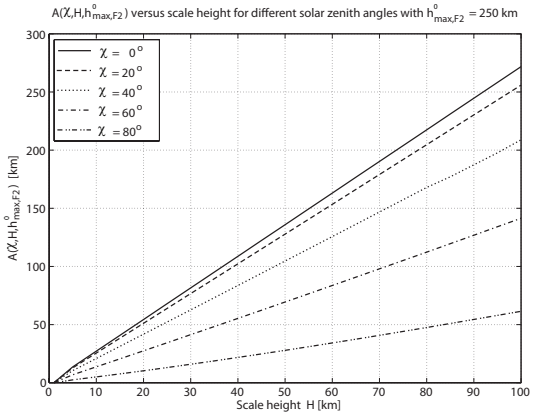

Figure 6.15: $A(\chi, H, h^0_{max})$ function versus scale height H for different χ at $h^0_{max,F} = 250\ km$.

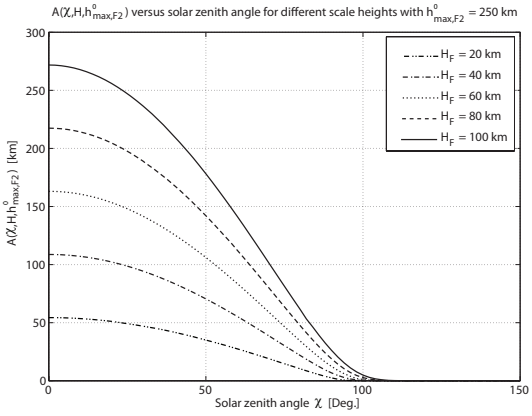

Figure 6.16: $A(\chi, H, h^0_{max})$ function versus solar zenith angle χ for different H at $h^0_{max,F} = 250\ km$.

According to (Titheridge, 2000), the average scale height of molecular oxygen O_2 at latitude of $40°N$ at a height of $h_{max,E} = 110\ km$ during a year under solar flux from F10.7=70 to 210 is equal to $H_E = 6.44\ km$. Therefore, we can use this average scale height as a constant scale height in equation 6.55.

Tabulating A and B functions

To evaluate both particular functions $A(\chi, H, h^0_{max})$ and $B(\chi, H, h^0_{max})$, defined respectively in equations 6.37 and 6.23, numerical integration should be used. Since the Chapman grazing incidence function $Ch(\chi, h)$ is involved in the integrands, i.e. $a(\chi, H, h^0_{max})$ and $\sqrt{a(\chi, H, h^0_{max})}$, the numerical integration will be time consuming. The graphs in figure 6.15 show the variation of the A function versus scale height H for different solar zenith angle χ. The variation of the A function versus solar zenith angle χ is depicted in figure 6.16 for various scale heights H. The graphs in these figures were drawn for $h^0_{max,F} = 250\ km$. The graphs of the B function versus scale height and solar zenith angle for $h^0_{max,H} = 110\ km$ are depicted respectively in figures 6.17 and 6.18. Note that for larger solar zenith angles ($\chi > 150°$) the A and B functions values tend to zero. From these figures it is clear that both A and B functions are smooth functions. Therefore, to speed up the computation process, the functions can be tabulated over different values of their input arguments.

The full range of solar zenith angle is $0 \leq \chi \leq 180°$ and the range of the scale height H is from 1 to 100 km. Tests showed that the A and B functions are rather insensitive against the height of peak electron density h^0_{max}, therefore the functions have only been tabulated for appropriate intervals along solar zenith angle and along scale height for fixed h^0_{max}. We used $h^0_{max,E} = 110\ km$ and $h^0_{max,F} = 250\ km$ as appropriate heights for the normal E- and F-regions, respectively. The interval for solar zenith angle is $2°$ and for scale height $5\ km$. Several checks were performed beforehand to ensure that the A and B functions vary smoothly inside of these intervals. We used the fixed height $250\ km$ as height of the peak electron density in function A (appropriate peak height for the F-region when $\chi = 0$) and the fixed height $110\ km$ as height of the peak electron density in function B. Figures 6.19 and 6.20 show the tabulated A and B functions.

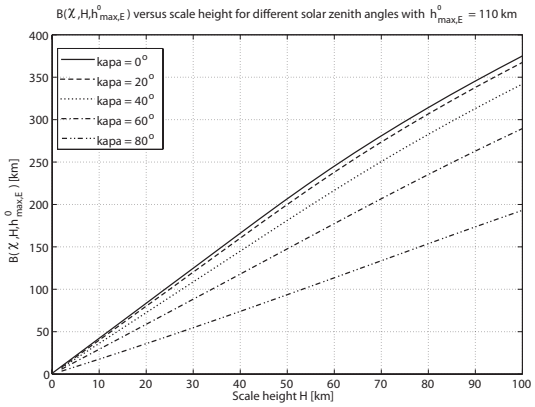

Figure 6.17: $B(\chi, H, h^0_{max})$ function versus scale height H for different χ at $h^0_{max,E} = 110\ km$.

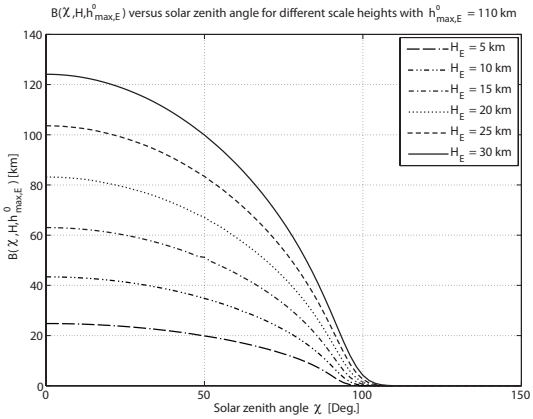

Figure 6.18: $B(\chi, H, h^0_{max})$ function versus solar zenith angle χ for different H at $h^0_{max,E} = 110\ km$.

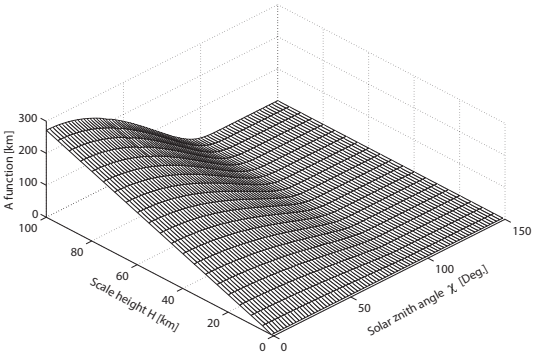

Figure 6.19: $A(\chi, H, h^0_{max})$ function versus solar zenith angle χ and scale height H for $h^0_{max,F} = 250\ km$.

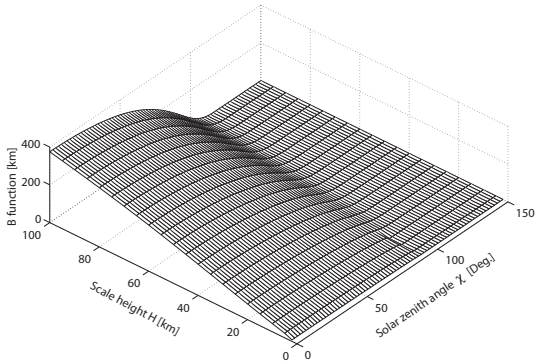

Figure 6.20: $B(\chi, H, h^0_{max})$ function versus solar zenith angle χ and scale height H for $h^0_{max,E} = 110\ km$.

For any given values of solar zenith angle χ and scale height H, the function value of A and B can be computed by the four nearest grid values using a *bilinear interpolation*. The function values of A and B for large solar zenith angles $\chi > 150°$ are set to zero.

6.2.3 Functional model for estimating the parameters

Three model parameters (ϕ^{photon}_{∞}, β, α) in equation 6.55 should be essentially estimated by fitting the model to daily VTEC observations at a given point. Due to non-linearity of the model, the model should be linearized and then the parameters can be estimated in the least squares iteration process. Since the model has a recursive form, for *initialization* at first epoch it is necessary to consider $N(t_0)$ (denoted as N_0) as an additional unknown parameter. Let x denotes vector of unknown parameters and $\tilde{y}$ denotes vector of observations (uncorrelated VTEC values) at $m > 4$ epochs at a given point

$$x = \begin{bmatrix} \alpha \\ \beta \\ \phi_{\infty} \\ N_0 \end{bmatrix} \quad ; \quad \tilde{y} = \begin{bmatrix} N_1 \\ N_2 \\ \vdots \\ N_m \end{bmatrix} \quad ; \quad Q_{\tilde{y}} = \begin{bmatrix} \sigma^2_{N_1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma^2_{N_m} \end{bmatrix} \tag{6.58}$$

where x is 4×1 vector, $\tilde{y}$ is m-vector, and $Q_{\tilde{y}}$ is $m\times m$ vc-matrix of the observation vector. For convenient the solar photon radiation flux is denoted by ϕ_∞ and VTEC observation at time t_i is denoted as N_i.

The functional model of the observation vector and vector of parameters is shown by m-vector of functions as $f(x,y)=0$ where y stands for the true observation vector. When redundant observations are available ($m>4$) the functional model will be inconsistent (due to observation random error), i.e. $f(x,\tilde{y})\approx 0$, that assuming $E\{\tilde{y}\}=y$ leads to $f(x,E\{\tilde{y}\})=0$. The functional model is given as follows

$$f(x,\tilde{y})=\begin{bmatrix} f_1(x,N_0,N_1)\\ f_2(x,N_1,N_2)\\ \vdots\\ f_m(x,N_{m-1},N_m)\end{bmatrix}\approx 0 \tag{6.59}$$

with $f_i(x,N_{i-1},N_i)$ for $i=1,\ldots,m$ as observation equation (base on equation 6.55) is

$$f_i(x,N_{i-1},N_i)\ =N_i\ -e^{-\beta\Delta t}N_{i-1}\ -\frac{A(\bar{\chi})\left(1-e^{-\beta\Delta t}\right)}{e\ \beta\ H_F}\ \phi_\infty\ -\frac{\left[B_i-e^{-\beta\Delta t}B_{i-1}\right]}{\sqrt{e\ \alpha\ H_E}}\ \sqrt{\phi_\infty} \tag{6.60}$$

with $\Delta t=t_i-t_{i-1}$, $B_i=B(\chi(t_i))$, and $\bar{\chi}=\frac{\chi(t_i)+\chi(t_{i-1})}{2}$. Note that the vector of parameters is considered constant in time.

The least squares estimator of the vector of parameters $\hat{x}$ is obtained by minimizing the squared norm of the vector of residuals $\|\hat{y}-\tilde{y}\|^2_{Q^{-1}_{\tilde{y}}}$ subject to the $f(\hat{x},\hat{y})=0$ (the 'hat' symbol indicates the least squares estimator) (Teunissen et al., 2005). This is a *constrained optimization problem* with *objective function* $\|\hat{y}-\tilde{y}\|^2_{Q^{-1}_{\tilde{y}}}$ and with *constraints* $f(\hat{x},\hat{y})=0$.

According to the *Lagrange's theorem*, if $\hat{x}$ is the *constrained minimizer* then the gradient of the objective function is a linear combination of the gradients of the constraints. This is a *necessary condition* (or *Lagrange condition*) for the constrained minimizer. The Lagrange theorem states that the constrained minimizer is a *stationary point* of a linear combination of the objective function and the constraints

$$L(x,y,\ell)=\|y-\tilde{y}\|^2_{Q^{-1}_{\tilde{y}}}+\ell^T f(x,y) \tag{6.61}$$

where $L(x,y,\ell)$ is the so-called *Lagrangian function* and ℓ is an m-vector of unknown *Lagrange multipliers*. In fact, the necessary condition in Lagrange's theorem is equivalent to the necessary condition for *unconstrained optimization* applied to the Lagrangian function (see Chong and Zak (2001)). At a stationary point, the gradient of the Lagrangian function is zero, i.e. $\frac{\partial L}{\partial X}(\hat{X})=0$ with $X=[x^T,y^T,\ell^T]$. Note that a stationary point of the Lagrangian function is only a candidate for the constrained minimizer of the objective function. For a stationary point to be indeed the constrained minimizer, the sufficient condition is that the *Hessian matrix* of the Lagrangian function is a positive definite matrix. This implies that the least squares estimator $\hat{X}$ is a stationary point of the Lagrangian function, i.e. $\frac{\partial L}{\partial X}(\hat{X})=0$, for which $\frac{\partial^2 L}{\partial X^2}(\hat{X})$ is a positive definite matrix.

Since the functional model $f(x,y)=0$ is *non-linear* and *non-explicit* with respect to the unknown vector x and observation vector y, a stationary point of the Lagrangian function

can be reached with two approaches: 1) using *Newton's method* (see (Teunissen, 1990)), 2) first substituting the linearized objective function (with respect to both x and y) in 6.61 and then *Gauss-Newton method* (in the least-squares iteration process) gives a stationary point of the Lagrangian function. Note that Newton's method is also an *iterative* approach and converges to the solution faster than the Gauss-Newton method. Both approaches converge to a unique solution, but estimating the vc-matrix of the estimator is complicated in Newton's method due to the non-explicit objective function. Because of this we use the second approach in this thesis. The linearized form of the objective function is derived in the following.

6.2.4 Linearization of the functional model

According to *Taylor's expansion theorem*, the *linear approximation* or *linearized version* of the functional model $f(x, y) = 0$ about given approximate values of the unknown vector (denoted as x^0) and the observation vector ($y^0 = \tilde{y}$)is given as follows

$$f(x,y) \;\approx\; f(x^0,y^0) \;+\; \partial_{x^T} f(x^0,y^0)\,(x-x^0) \;+\; \partial_{y^T} f(x^0,y^0)\,(y-y^0) \;\approx\; 0 \quad (6.62)$$

where first term on the right-hand side is referred to as the *zero-order* term and the second and third terms are referred to the *first-order* terms. The $m \times 4$ matrix $\partial_{x^T} f(x^0, y^0)$ and $m \times m$ matrix $\partial_{y^T} f(x^0, y^0)$ are partial derivatives of the functional model with respect to x and y, respectively. These matrices are

$$\partial_{x^T} f(x^0,y^0) = \begin{bmatrix} \frac{\partial}{\partial\alpha} f_1(x^0,y^0) & \frac{\partial}{\partial\beta} f_1(x^0,y^0) & \frac{\partial}{\partial\phi_\infty} f_1(x^0,y^0) & \frac{\partial}{\partial N_0} f_1(x^0,y^0) \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial}{\partial\alpha} f_m(x^0,y^0) & \frac{\partial}{\partial\beta} f_m(x^0,y^0) & \frac{\partial}{\partial\phi_\infty} f_m(x^0,y^0) & \frac{\partial}{\partial N_0} f_m(x^0,y^0) \end{bmatrix} \quad (6.63)$$

and

$$\partial_{y^T} f(x^0,y^0) = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ \frac{\partial}{\partial N_1} f_2(x^0,y^0) & 1 & 0 & \cdots & 0 \\ 0 & \frac{\partial}{\partial N_2} f_3(x^0,y^0) & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & & \vdots \\ 0 & 0 & \cdots & \frac{\partial}{\partial N_{m-1}} f_m(x^0,y^0) & 1 \end{bmatrix} \quad (6.64)$$

where

$$\frac{\partial}{\partial\alpha} f_i \;=\; \frac{\left[B_i - e^{-\beta\Delta t} B_{i-1}\right]}{2\sqrt{e\;\alpha^3\;H_E}}\;\sqrt{\phi_\infty} \quad (6.65)$$

$$\frac{\partial}{\partial\beta} f_i \;=\; \Delta t\; e^{-\beta\Delta t} N_{i-1} \;-\; \frac{A(\bar{\chi})}{e\;H_F\;\beta^2}\left[e^{-\beta\Delta t}(\beta\Delta t + 1) - 1\right]\phi_\infty \;+\; \Delta t\; B_{i-1}\; e^{-\beta\Delta t}\sqrt{\phi_\infty} \quad (6.66)$$

$$\frac{\partial}{\partial \phi_\infty} f_i \;=\; (e^{-\beta \Delta t} - 1)\, \frac{A(\bar{\chi})}{e\; H_F\; \beta} \;-\; \frac{1}{2}\left[B_i - B_{i-1}\; e^{-\beta \Delta t}\right]\; \frac{1}{\sqrt{\phi_\infty}} \tag{6.67}$$

$$\frac{\partial}{\partial N_0} f_i \;=\; \begin{cases} -e^{-\beta \Delta t} & \text{if } \; i = 1 \\ \\ 0 & \text{if } \; i \neq 1 \end{cases} \tag{6.68}$$

$$\frac{\partial}{\partial N_{i-1}} f_i \;= -e^{-\beta \Delta t} \tag{6.69}$$

It is convenient, the linearized functional model in equation 6.62 is rewritten as follows

$$A^0\; \Delta x \;+\; B^0\; \Delta y \;+\; f(x^0, y^0) \approx\; 0 \tag{6.70}$$

where $\Delta x = x - x^0$, $\Delta y = y - y^0$ and m-matrices of $A = \partial_{x^T} f(x, y)$ and $B = \partial_{y^T} f(x, y)$ and m-vector of $f(x^0, y^0)$ are known.

6.2.5 Least-squares solution of the model parameters

Least squares solution of the functional model After replacing the non-linear functional model by the linearized functional model (from equation 6.70) the Lagrangian function yields

$$L(x, y, \ell) \approx \|\Delta y\|^2_{Q^{-1}_{y^0}} + \ell^T \left(A^0\; \Delta x \;+\; B^0\; \Delta y \;+\; f(x^0, y^0)\right) \tag{6.71}$$

where is still a non-linear function. The least squares solution of the functional model is obtained by taking the derivatives of the Lagrangian function with respect to Δx, Δy, and ℓ and equating them to null vectors. The result is

$$\begin{aligned} &\tfrac{\partial L}{\partial \Delta y} = 2\; \Delta y^T Q^{-1}_{y^0} \;+\; \ell^T B^0 = 0 \\ &\tfrac{\partial L}{\partial \ell} = A^0 \Delta x \;+\; B^0 \Delta y \;+\; f(x^0, y^0) = 0 \\ &\tfrac{\partial L}{\partial \Delta x} = \ell^T A^0 = 0 \end{aligned} \tag{6.72}$$

The transpose of each equation in 6.72 leads to the following *hypermatrix*

$$\underbrace{\begin{bmatrix} Q^{-1}_{y^0} & B^{0^T} & 0 \\ B^0 & 0 & A^0 \\ 0 & A^{0^T} & 0 \end{bmatrix}}_{(2m+4)\times(2m+4)} \underbrace{\begin{bmatrix} \Delta y \\ \ell \\ \Delta x \end{bmatrix}}_{(2m+4)\times 1} + \underbrace{\begin{bmatrix} 0 \\ f(x^0, y^0) \\ 0 \end{bmatrix}}_{(2m+4)\times 1} = \underbrace{\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}}_{(2m+4)\times 1} \tag{6.73}$$

This is the *normal equation system* in hypermatrix form that can be solved directly for hypervector $\left[\Delta y^T,\ \ell^T,\ \Delta x^T\right]$ (as a solution of the linearized functional model) by inverting the coefficient hypermatrix.

The solution of the non-linear functional model can be estimated from the solution of the linearized model in an *iterative* process using *Gauss-Newton method*. It means that in iteration cycle $j+1$ the solution of the linearized model is estimated using the solution of the previous iteration cycle j as approximate values of unknown vector. Therefore, the normal equation system is written for iteration cycle j as follows

$$\begin{bmatrix} Q_{y^0}^{-1} & B^{j^T} & 0 \\ B^j & 0 & A^j \\ 0 & A^{j^T} & 0 \end{bmatrix} \begin{bmatrix} \Delta y^{j+1} \\ \ell^{j+1} \\ \Delta x^{j+1} \end{bmatrix} + \begin{bmatrix} 0 \\ w^j \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \tag{6.74}$$

where $A^j = \partial_{x^T} f(x^j, y^j)$ and $B^j = \partial_{y^T} f(x^j, y^j)$ are design matrices and m-vector w^j is called the *misclosure vector* as a known vector

$$w^j = f(x^j, y^j) + B^j\ (y^0 - y^j) \tag{6.75}$$

In equation 6.74, Δx^{j+1} and Δy^{j+1} are unknown vectors. Δx^{j+1} is called *vector of corrections* and Δy^{j+1} is called *residual vector*. Approximate values of vectors x and y at iteration cycle $(j+1)$ yield

$$x^{j+1} = x^j\ +\ \Delta x^{j+1} \tag{6.76}$$

$$y^{j+1} = y^0\ +\ \Delta y^{j+1} \tag{6.77}$$

Note that approximate observation vector for iteration cycle $(j+1)$ computed by adding the jth residual vector to the observation vector $y^0 = \tilde{y}$ while $(j+1)$th approximate unknown vector is obtained by adding jth vector of corrections to the approximate unknown vector of jth iteration cycle. To stop the iteration one needs a stop criterion. The iteration is terminated when the difference between two successive increments, i.e. Δx^j and Δx^{j+1}, is small enough.

The least squares estimators $\Delta\hat{y}^{j+1}$ and $\Delta\hat{y}^{j+1}$ can be obtained using the partitioning technique of the normal equation system as follows

$$\begin{aligned} \Delta\hat{x}^{j+1} &= -\ \left(A^{j^T} M^j A^j\right)^{-1} \left(A^{j^T} M^j\ w^j\right) \\ \Delta\hat{y}^{j+1} &= -\ Q_{y^0} B^{j^T} M^j\ (A^j \Delta\hat{x}^{j+1}\ +\ w^j) \end{aligned} \tag{6.78}$$

where $M^j = \left(B^j Q_{y^0} B^{j^T}\right)^{-1}$ is a full rank m-matrix. Then subistituting these into 6.76 and 6.77 gives respectively least squares estimator of the unknown vector $\hat{x}^{j+1}$ and the estimator of observation vector $\hat{y}^{j+1}$. The corresponding variance covariance matrices are obtained as

$$\begin{aligned} Q_{\hat{x}^{j+1}} &= Q_{\Delta\hat{x}^{j+1}} = \left(A^{j^T} M^j A^j\right)^{-1} \\ Q_{\hat{y}^{j+1}} &= Q_{y^0} - Q_{\Delta\hat{y}^{j+1}} \end{aligned} \tag{6.79}$$

with

$$Q_{\Delta\hat{y}^{j+1}} = Q_{y^0} B^{j^T} M^j \left[I - A^j (A^{j^T} M^j A^j)^{-1} A^{j^T} M^j \right] B^j Q_{y^0} \tag{6.80}$$

It is noticed that application of the variance propagation law to equation 6.75 leads to $Q_{w^j} = M^{j^{-1}}$.

Besides the estimation part, the least squares processing should contain a procedure to validate the obtained solutions. For this purpose, the *overall functional model test* as a general test on the validity of the model can be carried out.

6.3 Performance of the VTEC model

In this section, the VTEC model of equation 6.55 is validated and performance of the model is examined using VTEC data at *local* and *global* scales. It should be noted that the model only describes regular variation of VTEC due to changing in solar zenith angle. Therefore, the evaluations are carried out using VTEC data under geomagnetically quiet ionospheric condition (Kp < 4).

During initial testing it was found that the least squares solution of the linearized model is not convergence to the certain solution of the non-linear model by iterative solutions. This problem can be overcome by either having good approximate value for the unknown vector x^0 or by reducing the number of estimated parameters. This is due to the fact that the three model parameters α, β and ϕ_∞ are highly correlated. Therefore, α is fixed to an appropriate constant value which results in reduction in the size of unknown vector (from 4 to 3).

During day time in the E-region, the mean value of α for the dissociative recombination of electrons and molecular ions (O_2^+ or NO^+) is 10^{-1} $[\frac{m^3}{electron\ s}]$ (Titheridge, 2000). Therefore, in the all processing in this section, we fixed the quadratic recombination coefficient to $\alpha = 10^{-1}$ $[\frac{m^3}{electron\ s}]$.

The model parameters in equation 6.55 must be essentially estimated by fitting the model to daily VTEC observations at a given ionospheric point. We used the Global Ionospheric Maps (GIM) that are routinely produced by the International GNSS Service (IGS). More information about GIM is given in subsection 4.5.2. Since the maps are produced for every 2 hours in UT time system, a daily VTEC time series with 2 hours time interval in UT time are used. The generated time series can be used as an observation vector y to estimate the VTEC model parameters x in the least squares process. A diagonal covariance matrix is considered as the stochastic model of observation vector. As standard deviation of the provided VTEC value (from GIM) we used $\sigma_{_N} = 2$ TECU.

6.3.1 Local test of the VTEC model

In this subsection, the goodness of fit is is valuated for daily VTEC time series provided from the GIM maps at latitude 35° and longitude 50° (mid-latitude region) for two days (130 and 210) of 2004 (for both days Kp < 4). For each day, daily VTEC time series at the

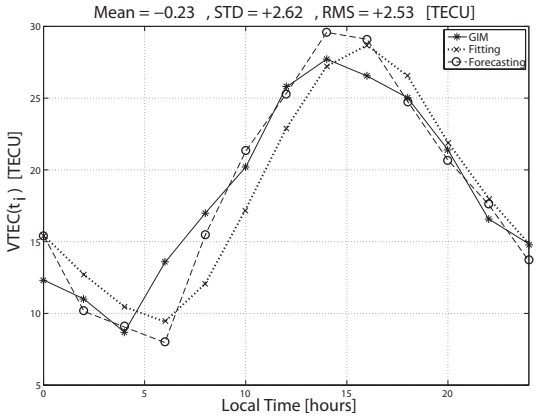

Figure 6.21: Daily VTEC from GIM maps (solid curve) and the modeled VTEC (dashed curve) in a ionospheric point ($\varphi = 35^\circ$ and $\lambda = 50^\circ$) for day 130 of 2004 (the dotted curve is for forecasting VTEC at each epoch using the model).

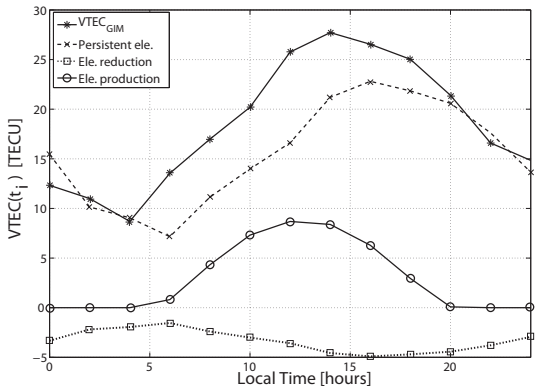

Figure 6.22: Electron production and reduction processes and persistent electrons of the VTEC in a ionospheric point ($\varphi = 35^\circ$ and $\lambda = 50^\circ$) for day 130 of 2004.

point with 2 hours time interval was generated using the GIM maps to estimate the model parameters in the least squares processing. The size of observation vector was $m = 13$ (VTEC values of every 2 hours from 0 to 24 in local time) and the size of unknown vector was 3 ($x = [\beta \;\; \phi_\infty \;\; N_0]^T$).

For day 130 of 2004, the recombination coefficient of the F-region and the solar photon radiation intensity were estimated as $\beta = 2.71 \times 10^{-5}$ in $[s^{-1}]$ and $\phi_\infty = 1.44 \times 10^{13}$ in $[\frac{photon}{m^2 \; s}]$, respectively. According to the estimated β, the lifetime of electrons for the F-region is $\tau_{_F} = 10.25$ hours. Based on $H_E = 6.44$ kilometer and $\alpha = 10^{-1}$ in $[\frac{m^3}{electron \; s}]$ and using the estimated ϕ_∞ in equations 6.53 and 6.10 gives $\sigma_E = 8 \times 10^{-9}$ in seconds. Using equation 6.14 gives the lifetime of electrons at the ionospheric point in the E-region (when the sun is overhead) as $\tau_{_E}^0 \approx 2 \times 10^{-4}$ in seconds. Because of the photochemical equilibrium condition in the model for the E-region, it is reasonable to have such a short lifetime for electrons in the E-region.

The modeled VTEC, using the estimated daily parameters, are shown in figure 6.21 for day 130 of 2004. In this figure, the solid curve is for the daily VTEC time series provided by the GIM maps and the dashed curve is for VTEC values computed by the fitted model. It is clear that the model follows the daily pattern of the VTEC. The mean residual (modeled VTEC minus VTEC from GIM) is less than one TECU, which is an indicator that the model is not significantly biased. The Root-Mean-Square value of the residuals was 2.53 TECU. The model error is increased about sunrise because the photoionization process in the real ionosphere is stats slightly earlier with respect to that of the model.

The earlier start of ion production in pre-dawn region could partly be due to the *conjugate region heating* as a special feature of ionosphere. During the early morning hours, the energetic *photoelectrons* created in the conjugate ionosphere in the F-region (by photoionization) are able to travel along the geomagnetic field lines from the conjugate region to the local region. In passing the energetic photoelectrons downwards into the dark hemisphere, collisions with neutral atmospheric particles occur and produce appreciable heating and ionization in pre-dawn region ionosphere (Banks, 1969). This phenomena is illustrated

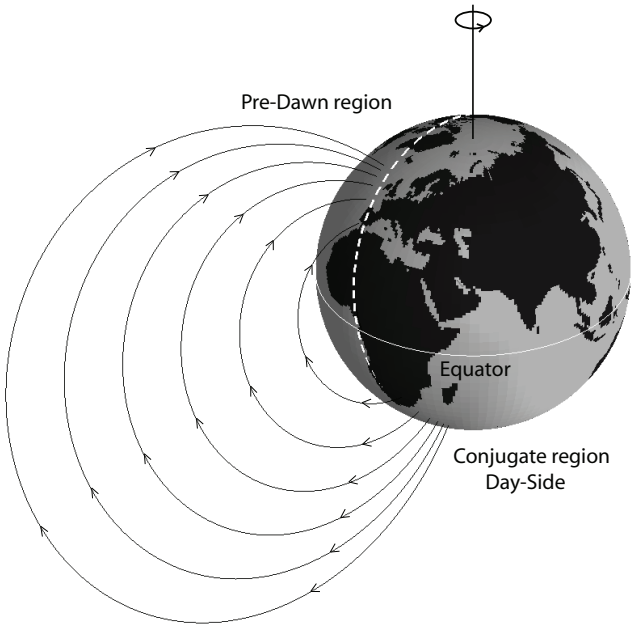

Figure 6.23: Geomagnetic conjugate region heating by photoelectrons at predawn region (dashed curve is boundary between day and night sides).

in figure 6.23.

It should be emphasized that ionization by the conjugate region heating phenomena is not large enough to affect significantly the nighttime ionosphere (Nagy et al., 1973). It is believed that due to the conjugate region heating and the complicated physics of ionosphere formation during sunrise, the simple Chapman production function cannot describe to a sufficient degree the actual rate of ion production during sunrise. Consequently we have to expect the large model error during predawn and sunrise. The large deviation of the model from real VTEC during sunrise is known as the *Predawn VTEC anomaly*.

The model can also be used as a forecasting model. In this case, to compute VTEC value $N(t_i)$ from the model, The VTEC value at previous epoch t_{i-1} (i.e. $N(t_{i-1})$) the observed VTEC is used. The graph of the forecasting VTEC is shown by dotted curve in figure 6.21.

For day 210 of 2004, the recombination coefficient of the F-region and the solar photon radiation intensity were estimated as $\beta = 2.21 \times 10^{-5}$ in $[s^{-1}]$ and $\phi_\infty = 9.50 \times 10^{12}$ in $[\frac{photon}{m^2\ s}]$, respectively. The lifetime of electrons for the ionospheric point in the F-region corresponding to the estimated β, is computed as $\tau_{_F} = 12.57$ hours.

In figure 6.24, the provided VTEC from GIM maps and the modeled VTEC for day 210 of 2004 are shown. The VTEC time series of day 210 of 2004 has two peaks during day time. This phenomenon is known as *daytime double maxima* (DDM) which happens often at the middle latitudes and lower latitudes as *twin peak* or *midday bite-outs* in the pattern

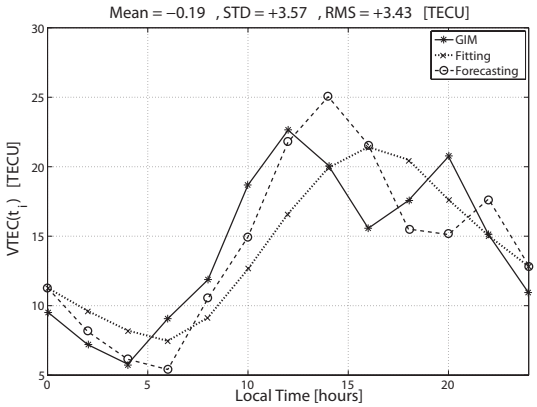

Figure 6.24: Provided daily VTEC from GIM maps (solid curve) and the modeled VTEC (dashed curve) in a ionospheric point ($\varphi = 35^\circ$ and $\lambda = 50^\circ$) for day 210 of 2004 (the dotted curve is for forecasting VTEC at each epoch using the model).

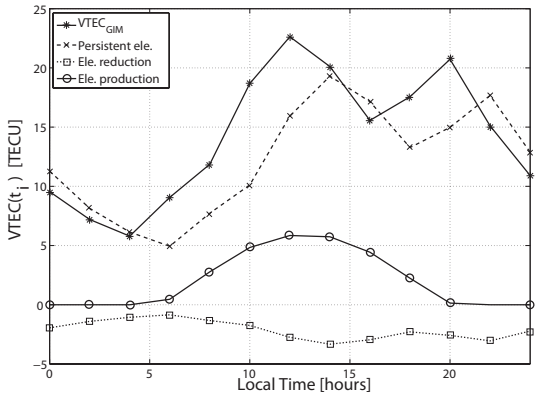

Figure 6.25: Electron production and reduction processes and persistence electrons of the VTEC in a ionospheric point ($\varphi = 35^\circ$ and $\lambda = 50^\circ$) for day 210 of 2004.

of VTEC. The DDM phenomenon is a result of the a combined effect of $\vec{E} \times \vec{B}$ drift and altitude-dependent F-region recombination coefficient β (Pi et al., 1993). Since the model of VTEC mathematically consists of only one peak in a day, the DDM can not be captured by the model and results in an increase of the model error and RMS value.

In figures 6.22 and 6.25, the *persistent* electrons $e^{-\beta\ \Delta t} N(t_{i-1})$ (which survive after the recombination), the electron production (summation of the second and third terms in the model), and recombination of electrons (electron reduction) versus time are plotted separately at the ionospheric point during days 130 and 210 of 2004, respectively. The amount of electron reduction at each epoch is computed as $N(t_{i-1}) - e^{-\beta\ \Delta t} N(t_{i-1})$. It can be seen that during night time, the ion production is stopped and only the recombination process is active.

6.3.2 Global test of the VTEC model

In this subsection, the model performance is checked globally and the VTEC model parameters are estimated using all GIM grid points. To get an insight of the global performance of the model, we need first to consider an appropriate strategy for the spatial dependency of the model parameters. Three strategies can be taken to estimate the model parameters over the world:

1. β and ϕ_∞ are both spatially dependent and they should be estimated individually for each grid point (both are *local parameters*)

2. ϕ_∞ depends on location of grid point and it should be individually estimated for each grid point (local parameter) but β is constant over the world and estimated as a *global parameter*

3. ϕ_∞ is considered as a *global parameter* and it is not spatial dependent but β should be estimated for each grid point (local parameter)

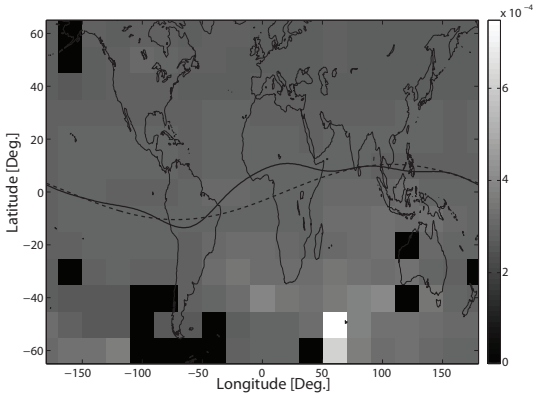

Figure 6.26: Spatial distribution of daily β in $[s^{-1}]$ over the world for day 130 of 2004 (the black pixels belong to the grid points that estimation process was not converged to the acceptable solution), the solid curve denotes the dip equator and the dashed curve denotes the geomagnetic equator.

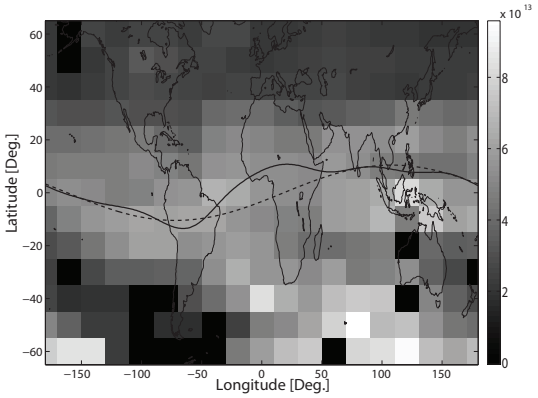

Figure 6.27: Spatial distribution of daily ϕ_∞ in $[\frac{photon}{m^2\ s}]$ over the world for day 130 of 2004 (the black pixels belong to the grid points that estimation process was not converged to the acceptable solution), the solid curve denotes the dip equator and the dashed curve denotes the geomagnetic equator.

It should be mentioned that the initial VTEC value N_0 is also spatial dependent and is considered as a local parameter in all three strategies. In the following the results obtained by each of these strategies are described. We did not use all GIM grid points but processing was carried out only for grid points of every 10° in latitude and 20° in longitude.

First strategy In the first strategy, both β and ϕ_∞ are taken as local parameters and the parameters were estimated for each selected grid point of the GIM maps independently. Therefore, the parameter estimation process was carried out exactly same as the local test processing in the previous subsection. In this strategy, the degree of freedom is independent of the number of grid points and it is 10. The parameters were estimated for the grid points (every 10° in latitude and 20° in longitude) for day 130 of 2004. Note that for the global test of the model in all three strategies, higher latitudes ($\phi > 60^\circ$) were not processed. The total number of the selected grid points was 234. The estimated values for β and ϕ_∞ for the selected grid points over the world are shown in the figures 6.26 and 6.27, respectively. These figures show spatial dependency of the estimated daily model parameters over the world. The observation equation 6.60 contains the ratio of $\frac{\phi_\infty}{\beta}$ in the second term, therefore iterations in estimating both of β and ϕ_∞ as local parameters sometimes becomes unstable and did not converge to a certain solution. This problem occurred for the grid points associated to the black pixels in the figures. Sometimes the solution converged to an abnormal solution (for instance the white pixel in figure 6.26). Because of this, solutions for some grid points differ significantly with the solutions in the vicinity of those grid points. This is the main disadvantage of the first strategy for model parameter estimation.

Second strategy It is assumed that β is not spatially dependent and its daily value is the same for all the grid points. Because of this, the degree of freedom of the estimation process is increased. This results in iterations always converging quickly to a certain solution. In second strategy, the observation vector Y and unknown vector x are as follows

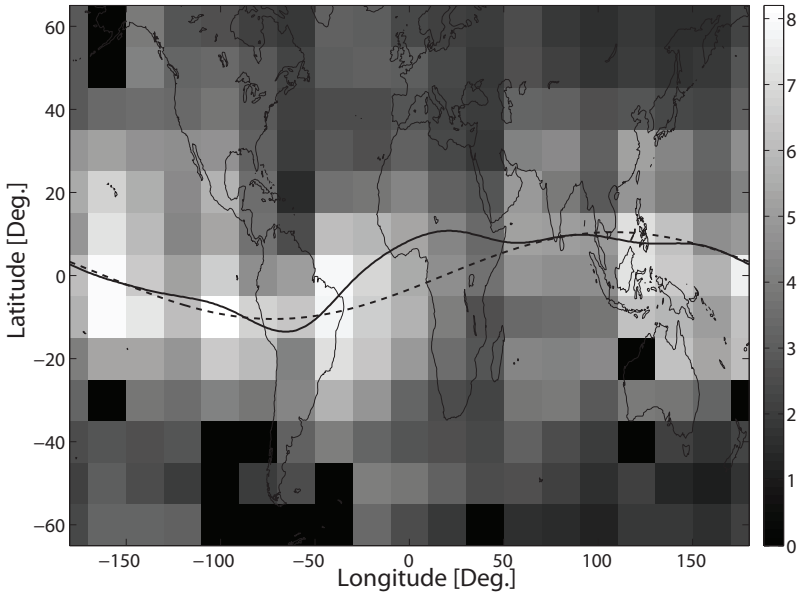

Figure 6.28: Computed RMS values of the model error (computed VTEC by model minus provided corresponding VTEC from GIM maps) for the selected grid points over the world for day 130 of 2004, total RMS of the model errors is $+4.90$ [TECU], the first strategy (β and ϕ_∞ as local parameters) was adapted in the model parameters estimation.

The estimated value for daily global β for day 130 of 2004 was $5.6\times10^{-5}\ s^{-1}$ with standard deviation of $\sigma_\beta = \pm2.9\times10^{-7}$. The corresponding lifetime of electrons for day 130 of 2004 is computed as $\tau_{_F} \approx 5$ hours. The spatial distribution of the estimated solar radiation flux is plotted in figure 6.29. It is clear that ϕ_∞ has a strongly latitudinal dependency or, to be more specific, it is depends on the modified Dip-latitude. Regardless of a slight longitudinal dependency, it could be useful to express the spatial dependency of ϕ_∞ by a simple function like a second order polynomial with respect to modified Dip-latitude.

The root-mean-squared (RMS) values of the *model errors* (computed VTEC by model minus provided corresponding VTEC from GIM maps) are given in figure 6.30. In the middle latitude region and higher latitudes, RMS values are a few TECU, which implies that the model fits well to the time series of VTEC provided from GIM maps. The average RMS value is equal to 4.24 in TECU. As expected, the larger RMS values are for the grid points around geomagnetic equator. In equatorial region, the ionospheric electron density is strongly suffered by the transport process. Since the transport term is not involved in the formulation of the model, the model does not fit particularly well in the equatorial region.

Third strategy For parameterization of the model, in the third strategy, β is considered as a local parameter and ϕ_∞ is as a global parameter. The least squares processing was carried out for day 130 of 2004 using all 234 selected grid points of the GIM maps. Daily global value of ϕ_∞ was obtained 3.55×10^{13} in $[\frac{photon}{m^2\ s}]$. The corresponding lifetime of electrons in the E-region is computed as $\tau_{_E} = 1.2\times10^{-4}$ second. The estimated effective β values for the grid points are shown in figure 6.31. The spatial distribution of β shows that it significantly depends on geomagnetic latitude for which estimated values of β for equator region is lower than the values for higher latitudes. It is because of the fact that VTEC value is higher in the equator and since ϕ_∞ estimated as a global parameter therefore the

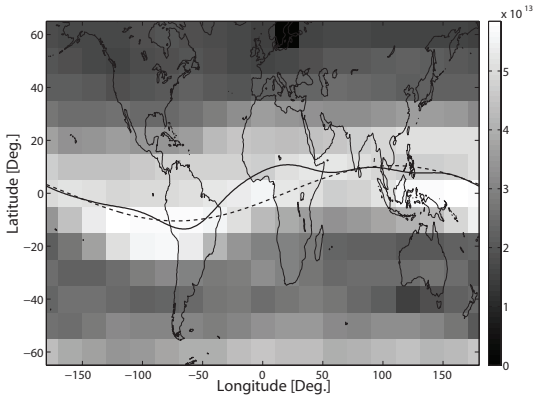

Figure 6.29: Spatial distribution of the effective daily solar radiation flux ϕ_∞ in $[\frac{photon}{m^2\ s}]$ over the world estimated by the second strategy (β as a global parameter estimated $5.6 \times 10^{-5}\ s^{-1}$) for day 130 of 2004, the solid curve denotes the dip equator and the dashed curve denotes the geomagnetic equator.

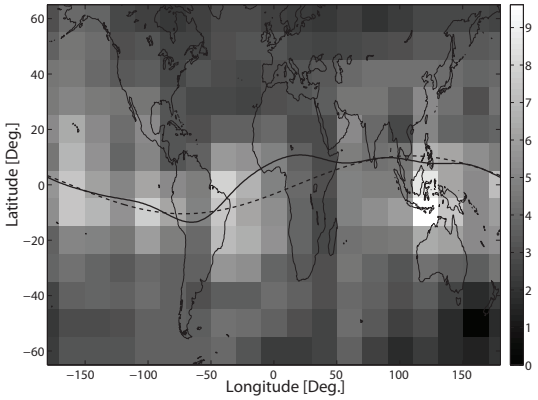

Figure 6.30: Computed RMS values of the model error (computed VTEC by model minus provided corresponding VTEC from GIM maps) for the selected grid points over the world for day 130 of 2004, total RMS of the model errors is $+4.24$ [TECU], the second strategy (β as a global and ϕ_∞ as a local parameters) was adapted in the model parameter estimation.

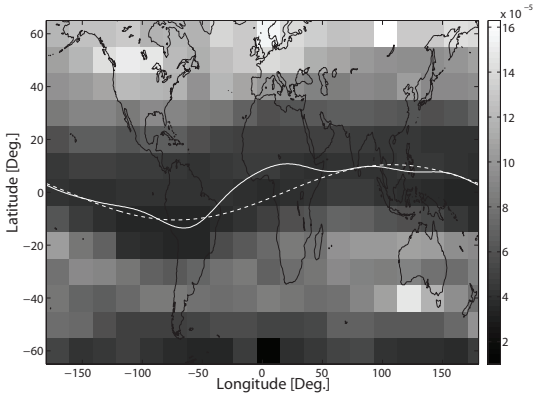

Figure 6.31: Spatial distribution of the effective daily β in $[s^{-1}]$ over the world estimated by the third strategy (ϕ_∞ as a global parameter estimated $3.55 \times 10^{13}\ [\frac{photon}{m^2\ s}]$) for day 130 of 2004, the solid curve denotes the dip equator and the dashed curve denotes the geomagnetic equator.

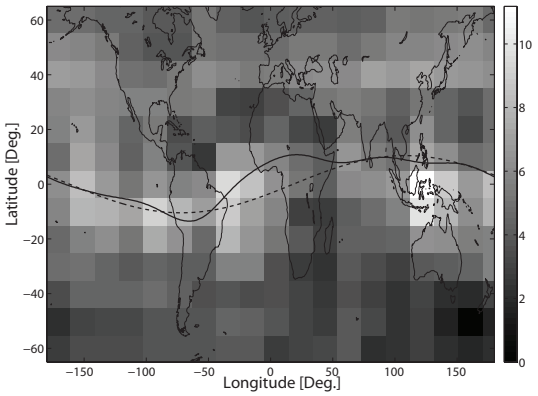

Figure 6.32: Computed RMS values of the model error (computed VTEC by model minus provided corresponding VTEC from GIM maps) for the selected grid points over the world for day 130 of 2004, total RMS of the model errors is $+5.15$ [TECU], the third strategy (β as a local and ϕ_∞ as a global parameters) was adapted in the model parameter estimation.

model can fit itself well to the data only for small value of β. The RMS values of the model errors computed for the grid points are depicted in figure 6.32. The larger RMS values are for the low latitudes. The total RMS of model error for the world in day 130 of 2004 was 5.15 [TECU]. Comparing this value to that of obtained for second strategy implies that the second strategy for model parameterization fits better to the observed data.

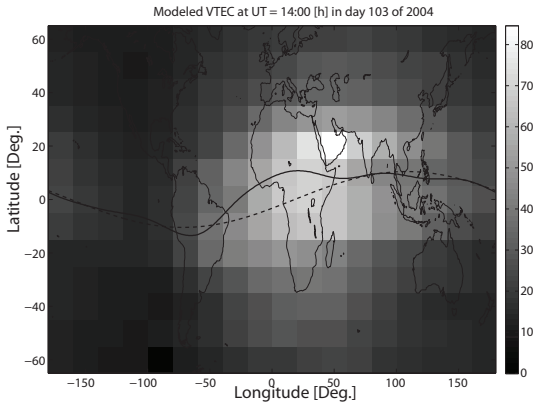

Figure 6.33: Global VTEC grid-map provided by the model for 14:00 UT in day 103 of 2004.

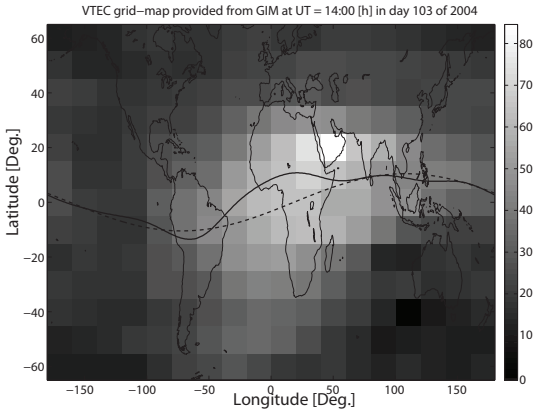

Figure 6.34: Global VTEC grid-map provided by the GIM for 14:00 UT in day 103 of 2004.

6.3.3 Applications of the VTEC model

The recursive model of the VTEC in equation 6.55 gives a good overall representation of the VTEC as a function of the solar zenith angle χ. The model can be used in many applications for which temporal interpolation or prediction of VTEC is needed. For instance, the model can be applied for temporal interpolation between the 2 hourly VTEC maps of GIM. Since the model has recursive form it can be also applied as the forecasting model for dynamic ionospheric modeling in the Network-RTK.

The model also can be used as a global ionospheric model that derived based on the ionospheric physics. The daily model parameters can be estimated from either the daily GNSS observations or daily VTEC maps of GIM. In figure 6.33,as an example the global grid-map of the VTEC provided by the model at 14:00 UT for day 103 of 2004 is shown. For comparison, the corresponding grid-map of VTEC provided from GIM is shown in figure 6.34.

As mentioned VTEC plays crucial role in the satellite communication field that its regular variation in time and space mostly related to changing in the geometry between the Sun and the Earth's ionosphere. The additional temporal and spatial fluctuations in the ionosphere are strongly related to the space weather. The model can be applied for monitoring response of the ionosphere to geomagnetic activities using the GNSS data. For this purpose, it is useful to apply the model to describe regular variations in the VTEC data and then the residuals could be interpreted as the geomagnetic effects on the ionosphere.

6.4 Summary

In this chapter a recursive model for temporal variation of vertical total electron content (VTEC) versus the solar zenith angle has been developed. The regular variations in VTEC by time and space that mostly related to changing in the geometry between the Sun and the ionosphere are predominantly captured by the model. To derive the model of VTEC, a normal ionosphere with two ionospheric regions (E- and F-regions) was defined with respectively quadratic and linear recombination coefficients for the normal E- and F-

regions. The model was developed based on the ionospheric plasma continuity or balance equation in absence of the transport processes. Therefore, the model only can describe regular variation of the VTEC under geomagnetically quiet conditions of the ionosphere. The Chapman theory was utilized for the rate of ion production during day time in which the Chapman grazing incidence function was applied for the absorption of ionizing solar radiation. The model is based on the ionospheric formation physics in which the linear recombination coefficient of the normal F-region β (in $[s^{-1}]$) and the solar ionizing radiation flux ϕ_∞ (in $[\frac{photon}{m^2\ s}]$) are estimated as daily model parameters. These model parameters are estimated in concept of the least squares adjustment by fitting the model to daily VTEC observations at a given point.

The performance of the model were checked at local and global scales using VTEC values from GIM maps. For the local test of the model, the model parameters are estimated for a given ionospheric point by least squares processing using daily VTEC values at the point provided from GIM maps. In the our case, the RMS of model error was a few TECU. For the global test of the model, different parameterization strategies were used. The best strategy turned out to be to estimate β as a global parameter and ϕ_∞ as a spatially dependent (local parameter). In general, initial results obtained from the cases analyzed for local and global tests of the model show that the our recursive model of VTEC gives a good overall representation of the VTEC in the middle latitude region under geomagnetically quite ionosphere.

The recursive characteristic of the model allows for application of the model not only for absolute VTEC computation at a given point and time but also for forecasting VTEC value for future epochs using a measured VTEC at the current epoch. This feature will be used in the next chapter for TID modeling in network RTK.

7 Real-Time Modeling for Medium-Scale TID

7.1 Introduction

In the two previous chapters different models for the regular variation of the total electron content in the ionosphere were discussed. In the current chapter a model for the irregular variations of the total electron content that particularly associated to the medium-scale *Traveling Ionospheric Disturbances* (TID) is given. The medium-scale TID is a wavelike fluctuation in electron density of ionosphere, accordingly in the TEC, which has wavelength less than 1000 km. It is the main obstacle for accurate spatial interpolation of ionospheric induced delays in a medium-scale GPS reference network (baseline length less than a few hundred kilometers). This is due to the fact that most interpolation methods either use spatial linear (or quadratic) interpolation or fit a lower-order surface. The result is that these methods are not capable to model the phase-offset, in the periodic fluctuations caused by TIDs, at the different ionospheric pierce points of the network.

In order to achieve short initialization times in the Network-RTK (Real-Time Kinematic) positioning system, it is necessary to have cm-level spatial prediction of the double differenced ionospheric delay. To achieve such a level of accuracy, the TID effects should be taken into account in the ionospheric spatial prediction. There are two major complications for the TID mitigation that should be addressed. Firstly, mitigation must be taken place at the double-difference level, which involves taking single differences between ionospheric delays for the same satellite between two receivers, followed by differencing their single differences for different satellites. This means that two different areas of the ionosphere are involved, each related to a different satellite, and each possibly associated with different TIDs. Secondly, for the Network-RTK a real-time strategy for TID detection and modeling is needed.

In the literature already several methods have been developed for TID detection and analysis. In all these methods the TID propagation parameters, e.g. propagation direction and phase velocity, are estimated using GPS derived un-differenced ionosphere delays mostly in *post-processing*. The main objective of this chapter is to develop an algorithm for estimation of the propagation parameters which can be applied to Network-RTK. A method is proposed to estimate the TID parameters using single-difference (between-receiver) ionospheric data obtained from a GPS network in *real-time*. The single-difference slant ionospheric delays for a baseline and for a satellite are computed epoch by epoch using both the code and carrier phase geometry-free linear combinations as explained in the chapter 4.

In the section 7.2 the properties of the different types of the TIDs are explained particularly

the medium scale TID. The mathematical equations for different types of mechanical waves are given in section 7.3. Although the geometry-dependent trend (associated to the solar zenith angle) is reduced to a large extent by the differencing, it is still necessary to detrend the data in order to provide the pure medium-scale TIDs observations. In section 7.4, the detrending of the single-difference data is explained based on fitting the physics-based model, developed in chapter 6, to the total electron content. The mathematical equation of the TID observations is derived in section 7.5 by taking into account the Doppler-shift due to the satellite motion. The derivation of the observation equation is based on a monochromatic longitudinal plane wave for TID wavelike oscillations which propagates horizontally in the thin layer ionosphere. The least-squares estimation of the TID wave parameters is discussed in section 7.6. In section 7.7 an algorithm for the real-time TID detection and modeling is developed. The performance of the algorithm is tested using real TID observations for two different satellites. Finally, section 7.9 ends the chapter with some concluding remarks.

7.2 Medium-Scale Traveling Ionospheric Disturbances

The Medium-Scale (MS) TIDs are wavelike disturbances in the electron density of the ionosphere, particularly in the F-region, which have wavelength less than $1000\ km$ and their period being in range of 12 minutes to 1 hours. This implies that the horizontal phase velocity of the MS-TIDs are usually smaller than velocity of sound in the ionosphere (about $300\ m/s$). Despite the small amplitude of the MS-TIDs, typically tenths of $TECU$, they are a main obstacle for accurate spatial interpolation of ionospheric induced delays in a medium-scale reference GPS network (baseline length less than a few hundred kilometers). This is due to the fact that most interpolation methods either use spatial linear (or quadratic) interpolation or fit a lower-order surface, which do not model the phase-offset caused by the MS-TIDs, at the different ionospheric pierce points of the network.

MS-TIDs are mostly a consequence of atmospheric gravity waves (AGW) propagating in the ionosphere via collisions of the neutral and charged particles (Hernandez-Pajares et al., 2006a; Afraimovich et al., 2002; Hines, 1960). This is in contrast to Large-scale TIDs that are generated mostly from specific geophysical events like ionospheric substorms (in the auroral zone) and are correlated with geomagnetic activity. It follows that the MS-TIDs have internal sources, because the AGWs are generated by the meteorological phenomena like *neutral winds*, *eclipses*, or *solar terminator*.

The morphology of the MS-TIDs, such as frequency of occurrence, time and location, have recently been studied by several authors. According to (Hernandez-Pajares et al., 2006a) the MS-TIDs are associated with the solar terminator and occur at daytime in local winter, and nighttime in local summer, and amplitude is correlated with the solar cycle. In (Candido et al., 2008), the frequency of occurrence of MS-TIDs, originating at high and mid latitudes, was analyzed statistically revealing an inverse dependency between the MS-TID occurrence rate and solar activity, with increased rate during the summer solstice.

MS-TIDs introduce wavelike and periodic variations in the TEC which are the result of changes in the electron density anywhere along the line of sight. In this chapter, the TEC observations from a network of dual-frequency GPS receivers are used for the MS-TID

analysis. It is assumed that the changes in TEC (due to the MS-TID) occur within a two dimensional thin ionosphere layer located at a fixed altitude of 350 km (approximately the altitude of the peak ionospheric electron density). The TIDs waves are modeled as mechanical waves. The general definition and mathematical expression for mechanical waves are presented in the next section.

7.3 Mechanical longitudinal wave equation

A *wave* is a disturbance or variation that propagates through space and time, which transferring energy. A *mechanical wave* is a type of wave that requires a medium for propagation (does not propagate in the vacuum). The medium itself is at 'rest' and material within the medium is not moved from one point to the other. Energy is transfered by local oscillations of the material only. When an initial energy input is added to the medium, a mechanical wave is created and travels through the medium until all the energy has been transferred. When the oscillations or vibrations are perpendicular to the the propagation direction the wave is called the *Transverse wave*. In case that the oscillation is in the propagation direction the wave called the *Longitudinal wave*. Mechanical longitudinal waves are also referred to as *compressional waves* or *compression waves*. A mechanical wave may take the form of a variation of pressure (density) or temperature or of an elastic deformation. The locus of points in the medium which oscillate with the same phase is called *wavefront*.

Plane wave The simplest form of a wavefront is the *plane wave* (or planar wave), where the surfaces of constant phase are infinite parallel planes of constant amplitude normal to the propagation direction. The term is also used to describe waves that are approximately plane waves in a localized region of space. For example, when the source of wave is far from observation region of the wave, then the wave is approximately a plane wave. In the following subsection, the mathematical expression for a traveling mechanical longitudinal plane wave is given.

7.3.1 Traveling plane wave

From a mathematical point of view, the most fundamental wave is a *harmonic* (sinusoidal) wave which is described in three dimensional space $\vec{r}$ and time t by following equation

$$y(t,\vec{r}) = A\ \sin\left(\omega t - \vec{K}\cdot\vec{r}\right) \tag{7.1}$$

where A is the amplitude of a wave - a measure of the maximum disturbance in the medium during one wave cycle-, $\omega = 2\pi f$ is the *angular frequency* and f the wave frequency (the period of the wave is $T = \frac{1}{f}$) and $\vec{K}$ is the *wave vector*. The amplitude and frequency in equation 7.1 do not depend on time and on space. In case that the amplitude varies with time and/or position, the wave is called the *envelope* of the wave.

The magnitude of the wave vector $\vec{K}$ is known as the *wave number* $K = \frac{2\pi}{\lambda}$ (λ stands for the wavelength). The propagation direction of the wave is $\vec{K} = \frac{2\pi}{\lambda}\frac{\vec{V}_{ph}}{V_{ph}}$ where $\vec{V}_{ph}$ stands for the *phase velocity* vector of the wave and V_{ph} the phase velocity of $\vec{V}_{ph}$. Note that in

case of the plane wave, the direction of the wave vector is constant in space. The phase velocity V_{ph} can be associated with the angular frequency ω by the relation $V_{ph} = \frac{\omega}{K}$.

Superposition Once two (or more) waves travel through the same medium at the same time (superposition), the net displacement of the medium at any point in space or time, is simply the sum of the individual wave displacements. This is true of waves which are finite in length (wave pulses) or which are continuous sine waves. The superposition of waves leads to either *Traveling* or *Standing* waves.

When two mechanical waves (with the same amplitude and frequency) are traveling with different phase velocities in the same direction in a medium, a new traveling wave is created with amplitude changing in space but constant in time. When the two waves are in-phase (phase difference equal to zero), they interfere constructively and the result has twice the amplitude of the individual waves. When the two waves have opposite-phase (phase difference equal to π), they interfere destructively and cancel each other out. In case that two waves of equal amplitude with different phase velocity have different frequency, the amplitude of the resulting traveling wave not only is time dependent but also is spatial dependent that leads to an envelope wave.

7.3.2 Standing plane wave

A standing wave, also known as a *stationary wave*, is a wave that remains in a constant position. As an example, a standing wave is created when the medium is moving in the *opposite direction* to the wave. As another example, in a stationary medium, a standing wave is formed by interference between two waves of the same amplitude and frequency traveling in *opposite* directions with the same velocity. The sum of these two plane waves is

$$\begin{aligned} y(t,\vec{r}) &= A\left[\sin\left(\omega t - \vec{K}\cdot\vec{r}\right) + \sin\left(\omega t + \vec{K}\cdot\vec{r}\right)\right] \\ &= \underbrace{2A\sin\left(\vec{K}\cdot\vec{r}\right)}_{\text{amplitude}}\ \cos\left(\omega t\right) \end{aligned} \tag{7.2}$$

which in the equation for a standing wave, the position and time dependent phases separated. The amplitude of the resultant wave is spatial-dependent.

As can be seen from equation 7.2, due to the interference of the two waves, there are certain points ($\vec{K}\cdot\vec{r} = 0, \frac{\lambda}{2}, \frac{3\lambda}{2}, \ldots$), called *nodes*, at which the total wave is zero at all times. The distance between two consecutive nodes is exactly half the wavelength, i.e. $\frac{\lambda}{2}$. The points at the middle between consecutive nodes ($\vec{K}\cdot\vec{r} = \frac{\lambda}{4}, \frac{3\lambda}{4}, \ldots$) are called *anti-nodes*. At the anti-nodes the total wave oscillates with maximum amplitude, equal to twice the amplitude of each wave. Anti-nodes are also half a wavelength apart.

7.4 GPS-provided TID observation

In this section a method is described to compute TIDs from measured slant ionospheric delay between a dual frequency receiver on the ground and GPS satellite. The slant ionospheric delay on the $L1$ phase observation $\tilde{I}_r^s(t)$ between receiver r and satellite s can be measured epoch by epoch using data of a dual frequency receiver, see section 4.6.

The measured slant ionospheric delay is time dependent because of satellite motion - changes in elevation angle and ionospheric piercing point - and regular and irregular variations in the ionospheric total electron density. The regular variation is due to temporal changes in the Earth-Sun geometry (changing in solar zenith angle $\chi(t)$), the irregular variations in the slant ionospheric delay are mainly caused by different types of TIDs.

Slant ionospheric delay decomposition The measured slant ionospheric delay $\tilde{I}_r^s(t, \chi(t), \vec{\rho}_r{}^s(t))$ (in meter) can be decomposed as follows

$$\tilde{I}_r^s(t, \chi(t), \vec{\rho}_r{}^s(t)) = \underbrace{G\left(\chi(t), \vec{\rho}_r{}^s(t)\right)}_{\text{Geometry-dependent Trend}} + TID(t) + \underbrace{\epsilon(t)}_{\text{Residual}} \tag{7.3}$$

with $\vec{\rho}_r{}^s(t)$ the geometric range vector between receiver r and satellite s, $\chi(t)$ is the solar zenith angle at the ionospheric piercing (IP) point, $G\left(\chi(t), \vec{\rho}_r^s(t)\right)$ a function that expresses *geometry-dependent trend* of the measured slant ionospheric delay, $TID(t)$ the effects due to both large- and medium-scale TIDs, and $\epsilon(t)$ effects of the small-scale TID and measurement (random) error.

To visualize and analyze the ionospheric disturbances, the geometry dependent trend $G\left(\chi(t), \vec{\rho}_r^s(t)\right)$, which is much larger than the TID, must be removed from the observation.

7.4.1 Geometry-dependent trend of slant ionospheric delay

The geometry-dependent ionospheric variations can be removed from the measured slant ionospheric delay in different ways. For instance (Hernandez-Pajares et al., 2006a) makes this detrending by subtracting from each value an average value of the previous and a posterior measurements. In Tsugawa et al. (2004), detrending is done by looking for vertical TEC perturbations from a dense GPS network that it is not applicable for a single-receiver measurements. Other methods for instance use a high-pass filtering or use a polynomial fitting to data over a certain period.

In this research, the goal is to analyze the irregular ionospheric variations in real-time for dynamic MS-TID modeling. Therefore, we will take advantage of the recursive physical model for VTEC discussed in subsection 6.2.1 to do detrending of the measured time series of slant ionospheric delay in real-time.

Under a thin layer assumption for the ionosphere, the geometry-dependent trend on the slant ionospheric delay can be expressed in terms of the vertical total electron content and a mapping function

$$G_r^s(t) = G\left(\chi(t), \vec{\rho}_r{}^s(t)\right) = \alpha \cdot m\left(\xi_{IP}(t)\right) \cdot N\left(\chi_{IP}(t), \varphi_{IP}(t), \lambda_{IP}(t)\right) \tag{7.4}$$

where $G_r^s(t)$ is in meter,N denotes the vertical total electron content at IP point at time t in $[TECU]$ as a function of solar zenith angle $\chi_{IP}(t)$ and geographic latitude $\varphi_{IP}(t)$ and longitude $\lambda_{IP}(t)$, $m\left(\xi_{IP}(t)\right)$ is a mapping function converting slant to vertical ionospheric delay with $\xi_{IP}(t)$ as satellite zenith angle at the IP point, and α is a factor that converts $TECU$ to m ($\alpha = 0.1624\ \frac{m}{TECU}$ for delay on L1).

For the N the following recursive relation was derived in section 6.2 that for simplicity the third term of equation 6.55 (associated to the E-region) was ignored.

$$N(t_i) = e^{-\beta\Delta t} N(t_{i-1}) + \frac{1}{e\ \beta\ H_F\ 10^{16}}\ A(\bar{\chi})\left(1 - e^{-\beta\Delta t}\right)\ \phi_\infty^{photon} \tag{7.5}$$

where $N(t)$ (in $TECU$) denotes $N\left(\chi_{IP}(t), \varphi_{IP}(t), \lambda_{IP}(t)\right)$. For an explanation of the other symbols see section 6.2.

In the following subsections, equation 7.4 is used as a basic model for detrending a given time series of measured slant ionospheric delays and the functional model of the observations and the least-squares solution for the model parameters will be discussed.

7.4.2 TID observation

Equation 7.4 can be fitted to a time series of slant ionospheric delays in order to estimate the geometric trend. For simplicity in the least-squares fitting, three assumptions are considered:

- the horizontal gradient of the vertical TEC at previous epoch is equal zero (i.e. $N(t_{i-1}, \varphi_{IP}(t_{i-1}), \lambda_{IP}(t_{i-1})) = N(t_{i-1}, \varphi_{IP}(t_i), \lambda_{IP}(t_i))$),
- the satellite zenith angles at two consecutive IP points are approximately the same $\xi_{IP}(t_{i-1}) \approx \xi_{IP}(t_i)$,
- the linear recombination coefficient β is not spatial-dependent.

Although IP point between ground receiver and satellite is moving, and each IP for which the slant ionospheric delay is measured represents a different area of the ionosphere, the changes are rather small and these three assumptions do not pose particular practical problems. Under these assumptions, equation 7.4 and equation 7.5 can be combined to get the following recursive equation

$$G_r^s(t_i) = e^{-\beta\Delta t}\ G_r^s(t_{i-1}) +\ \alpha\ \cdot m\left(\xi_{IP}(t_i)\right) \cdot \frac{A(\bar{\chi})\left(1 - e^{-\beta\Delta t}\right)}{e\ \beta\ H_F\ 10^{16}}\ \phi_\infty^{photon} \tag{7.6}$$

This model is fitted to the time series of slant ionospheric delay $\tilde{I}_r^s(t_i)$ with $i = 1, \ldots, k$ from t_1 to t_k measured by the geometry-free linear combination, three constant parameters β, ϕ_∞^{photon} and $G_r^s(t_0)$ must be estimated. For the least-squares fitting, 3-vector of unknown model parameters x and k-vector of observation y are considered as follows

$$x = \begin{bmatrix} \beta \\ \phi_\infty \\ G_0 \end{bmatrix} \quad ; \quad y = \begin{bmatrix} \tilde{I}^s_r(t_1) \\ \tilde{I}^s_r(t_2) \\ \vdots \\ \tilde{I}^s_r(t_k) \end{bmatrix} \quad ; \quad Q_y = \sigma^2_{\tilde{I}^s_r} \begin{bmatrix} 1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1 \end{bmatrix} \tag{7.7}$$

where G_0 denotes $G^s_r(t_0)$ and Q_y stands for $k \times k$ uncorrelated diagonal vc-matrix of the observations with $\sigma^2_{\tilde{I}^s_r}$ as variance factor (of unit weight) that gives precision of a single slant ionospheric delay observable.

The functional model of observations for estimating the three model parameters are

$$f(x,y) = \begin{bmatrix} f_1(x, \tilde{I}^s_r(t_0), \tilde{I}^s_r(t_1)) \\ f_2(x, \tilde{I}^s_r(t_1), \tilde{I}^s_r(t_2)) \\ \vdots \\ f_k(x, \tilde{I}^s_r(t_{k-1}), \tilde{I}^s_r(t_k)) \end{bmatrix} \approx 0 \tag{7.8}$$

where $E\{f(x,y)\} = 0$ and $f(x,y)$ is a k-vector and $f_i(x, \tilde{I}^s_r(t_{i-1}), \tilde{I}^s_r(t_i))$ condition equation based on equation 7.6 with $i = 1, \ldots, k$,

$$f_i(x, \tilde{I}^s_r(t_{i-1}), \tilde{I}^s_r(t_i)) \; = \tilde{I}^s_r(t_i) \; - e^{-\beta \Delta t} \tilde{I}^s_r(t_{i-1}) \; - \alpha \; m\left(\xi_{IP}(t_i)\right) \frac{A(\bar{\chi}) \left(1 - e^{-\beta \Delta t}\right)}{e \; \beta \; H_F \; 10^{16}} \; \phi_\infty \tag{7.9}$$

where $\Delta t = t_i - t_{i-1}$ and the averaged solar zenith angle is $\bar{\chi} = \frac{\chi(t_i) + \chi(t_{i-1})}{2}$.

Least-square estimation of model parameters The functional model is a *non-linear* and *non-explicit* with respect to x and y. The observation model must be linearized with respect to both observation and the unknown vectors about given initial values for unknown vector (denoted as x^0) and the observation vector ($y^0 = y$). The linearization can be carried out as it is explained in subsection 6.2.4 with $k \times 3$ design matrix of $\partial_{x^T} f(x^0, y^0)$ and $k \times k$ design matrix of $\partial_{y^T} f(x^0, y^0)$. The partial derivatives in the design matrices are

$$\frac{\partial}{\partial \beta} f_i \; = \; \Delta t \; e^{-\beta \Delta t} \tilde{I}^s_r(t_{i-1}) \; - \alpha \; m\left(\xi_{IP}(t_i)\right) \frac{A(\bar{\chi})}{e \; H_F \; \beta^2 \; 10^{16}} \left[e^{-\beta \Delta t}(\beta \Delta t + 1) - 1\right] \phi_\infty \tag{7.10}$$

$$\frac{\partial}{\partial \phi_\infty} f_i \; = \; \alpha \; m\left(\xi_{IP}(t_i)\right) \; \frac{A(\bar{\chi})}{e \; H_F \; \beta \; 10^{16}} (e^{-\beta \Delta t} - 1) \tag{7.11}$$

$$\frac{\partial}{\partial G_0} f_i \; = \; \begin{cases} -e^{-\beta \Delta t} & \text{if} \;\; i = 1 \\ \\ 0 & \text{if} \;\; i \neq 1 \end{cases} \tag{7.12}$$

$$\frac{\partial}{\partial \tilde{I}_{i-1}} f_i \; = -e^{-\beta \Delta t} \tag{7.13}$$

where $\tilde{I}_{i-1}$ stands for $\tilde{I}^s_r(t_{i-1})$. The unknown parameters can be estimated in an iterative process as it is discussed in subsection 6.2.5.

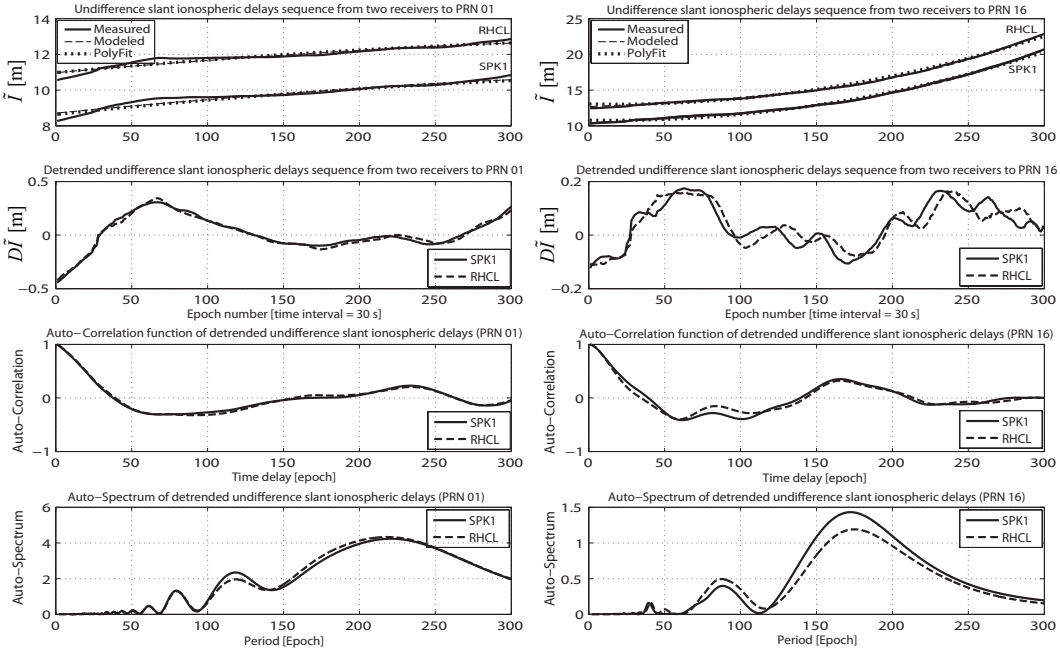

Figure 7.1: Measured and modeled time series of ionospheric delay during 2.5 h time span (UT=16:00 - 18:30) for two stations (SPK1 ans RHCL) using PRN 01 (left) and PRN 16 (right). The 'dotted line' in the first row is for modeled series by fitting 2-order polynomial and 'dashed line' is the modeled series by fitting the model of equation 7.6. The graphs in the second row show the detrended series (measured series minus modeled series). The graphs in third and forth rows are showing normalized auto-correlation function and corresponding auto-spectral density function of the detrended series, respectively.

TID observation When the model parameters are provided by the least-squares fitting of equation 7.6 to the time series of slant ionospheric delay $\tilde{I}_r^s(t_i)$ with $i = 1, \ldots, k$, the geometry-dependent trend $G_r^s(t_i)$ can be computed for all observation epochs t_i , $i = 1, \ldots, k$. Subtracting the computed geometry-dependent trend from the time series gives the *TID observation* (denoted by $D\tilde{I}_r^s(t)$)

$$D\tilde{I}_r^s(t) = \tilde{I}_r^s(t) - G_r^s(t) \;=\; TID(t) \;+\; \epsilon(t) \tag{7.14}$$

As an example, figure 7.1 shows time series of slant ionospheric delay measured from two permanent GPS stations SPK1 and RHCL in Southern California Integrated GPS Network (SCIGN) to two satellites PRN 01 and PRN 16 for the time span from UT=16.00 to 18.30 (300 epochs with 30 seconds time interval) dated 2003.10.28. The locations of the corresponding IP points are shown in figure 7.2 and the associated satellites ionospheric elevation angles and velocity of the IP points (average of two receivers) are depicted in figure 7.3.

In figure 7.1, the geometry-dependent trend of each time series is computed in two ways: 1) by least-square fitting a second order polynomial to the series (dotted curve) and 2) by least-square fitting of the model in equation 7.6 to the series (dashed curve).

The RMS values of the residuals (measured minus computed) of the fit are shown in table 7.1. In case of PRN 01 there is no significant difference between the two fitting methods but in case of PRN 16 the model of equation 7.6 fitted better to the time series than the 2nd order polynomial.

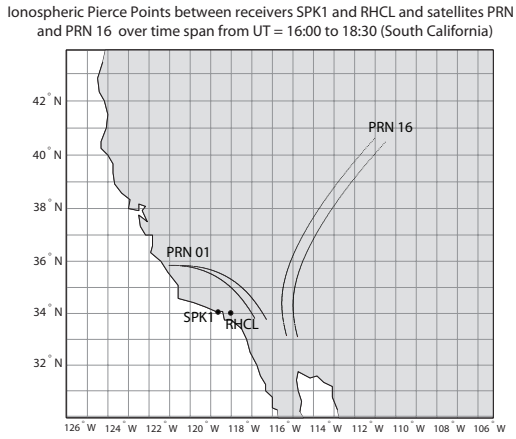

Figure 7.2: Ionospheric pierce points (IP) between receivers SPK1 and RHCL and satellites PRN 01 and PRN 16 from UT=16:00 to 18:30 in 2003.10.28.

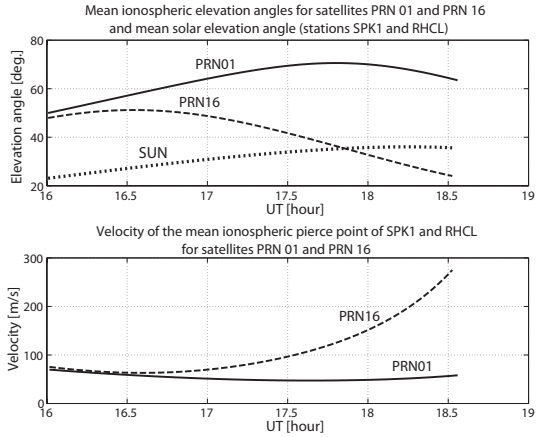

Figure 7.3: Ionospheric elevation angles from two receivers SPK1 and RHCL to satellites PRN 01 and PRN 16 from UT=16:00 to 18:30 in 2003.10.28 (top), corresponding velocity of the ionospheric pierce points (bottom).

Table 7.1: The RMS values of the residuals (measured minus computed) of fitting the model in equation 7.6 and a 2nd order polynomial to the time series of two stations SPK1 and RHCL for two satellites PRN 01 and PRN 16 in $[m]$.

	SPK1		RHCL	
	Equation 7.6	2nd order poly.	Equation 7.6	2nd order poly.
PRN 01	0.158	0.153	0.155	0.152
PRN 16	0.090	0.177	0.090	0.184

It can be resulted that a 2nd order polynomial is not always good mathematical function for detrending a time series of slant ionospheric delay. Because selection of the mathematical function depends on length and time of the time series. The model in equation 7.6 was developed on the basis of the geometry and physics involved in formation of the ionosphere, therefore it describes the geometry-dependent trend of the time series of slant ionospheric delay with any size at any time. Furthermore, it has a recursive form that provide opportunity for forecasting in case of dynamic modeling for the ionosphere.

The detrended slant ionospheric delay, or TID observation $D\tilde{I}_r^s(t)$, is plotted in the 2nd row of figure 7.1. The auto-correlation and The *auto-spectral density* functions are shown in the 3rd and 4th rows. The auto-spectral density function of a time series describes the general frequency composition of the time series in terms of the spectral density of its mean square value (Bendat and Piersol, 1971) that, specifically for stationary data, it is related to the auto-correlation function by *Fourier transform*. The global maximum of the auto-spectral density functions correspond a period which is larger than 120 epochs (1 hour). It means that the time series of PRN 01 and PRN 16 are dominated by the LS-TID.

7.4.3 Single-difference TID observation

The un-differenced ionospheric delay contained also residual of the geometric trend, effects from LS-TID and MS-TID. As we are mainly interested in MS-TID the effects of LS-TID need to be removed. In a medium scale GPS network (baseline length $\leq 200\ km$),

between-receiver single-differencing of the measured slant ionospheric delays is a simple way to remove the LS-TID. This is because of the fact that wavelength of the LS-TID is large ($\geq 1000\ km$), compared to the distance between receivers. Consequently, between-receiver single-differencing leads to data in which the effect of LS-TID is significantly reduced. The single-difference slant ionospheric delay, between two receivers r_1 and r_2, for satellite s is denoted by $sd\tilde{I}^s_{r_1,r_2}(t) \ = \ \tilde{I}^s_{r_2}(t) - \tilde{I}^s_{r_1}(t)$.

Although between receiver single-differencing also eliminates to some extend the geometry-dependent trend, the remaining part is considerable and it is therefore essential to also detrend the single difference data. This is because of the horizontal VTEC gradient and the curvature of the Earth. To do, one can either use single-difference of the modeled geometry-dependent trend ($G^s_{r_2}(t) - G^s_{r_1}(t)$), or form first the single-difference of the model in equation 7.6, and then fit it to time series of $sd\tilde{I}^s_{r_1,r_2}(t)$.

The latter is more confidence way to make detrending because on the one hand *geomagnetic condition* is spatially correlated, the single-differencing cancels out the effects of the geomagnetic on the data. On the other hand, referring to the chapter 6, the model of VTEC in equation 6.55 has been developed with disregarding the geomagnetic dependency of the ionosphere (*mis-modeling*). Thus, the single-difference model does not have mis-modeling (missing the geomagnetic field) could describe well the behavior of the single-difference slant ionospheric delay $sd\tilde{I}^s_{r_1,r_2}(t)$. The single-difference slant ionospheric delay may be decomposed as follows

$$sd\tilde{I}^s_{r_1,r_2}(t) = \underbrace{sdG^s_{r_1,r_2}(t)}_{\text{Geometry-dependent Trend}} + sdTID(t) \ + \underbrace{\epsilon(t)}_{\text{Residual}} \tag{7.15}$$

where $sdG^s_{r_1,r_2}(t)$ stands for the single-difference geometry-dependent trend, $sdTID$ denotes the single-difference MS-TID and ϵ denotes both measurement random errors and single-difference SS-TID with $E\{\epsilon\} = 0$. $sdG^s_{r_1,r_2}(t)$ is obtained by between-receiver single-differencing of equation 7.6 as follows

$$\begin{aligned} sdG^s_{r_1,r_2}(t_i) \ &= \ G^s_{r_2}(t_i) \ - \ G^s_{r_1}(t_i) \\ &\cong \ e^{-\beta\Delta t}\ sdG^s_{r_1,r_2}(t_{i-1}) + \ \alpha \cdot m\left(\bar{\xi}_{IP}(t_i)\right) \cdot \frac{A(\bar{\chi}_{r_1,r_2})\left(1-e^{-\beta\Delta t}\right)}{e\ \beta\ H_F\ 10^{16}}\ \phi^{photon}_{\infty} \end{aligned} \tag{7.16}$$

where $sdG^s_{r_1,r_2}(t_{i-1})$ is the between-receiver single-difference slant ionospheric delay from the previous epoch, the averaged satellite zenith angle is $\bar{\xi}_{IP}(t_i) = \frac{1}{2}\left(\xi_{IP_1}(t_i) + \xi_{IP_2}(t_i)\right)$ and averaged solar zenith angle is $\bar{\chi}_{r_1,r_2} = \frac{1}{2}(\bar{\chi}_1 + \bar{\chi}_2)$. Note that for the sake of simplicity for estimating model parameters, the averaged geometry from two receivers (r_1 and r_2) to satellite s and to the Sun are used (see figure 7.4).

Equation 7.16 consists three unknown parameters of $sdG^s_{r_1,r_2}(t_0)$, β and ϕ^{photon}_{∞} that can be estimated by fitting the equation to a time series of measured single-difference slant ionospheric delays $sd\tilde{I}^s_{r_1,r_2}(t)$ for a given baseline and satellite. The functional model of the least-squares estimation of the three parameters and its linearized version can be obtained using an approach similar to the previous subsection.

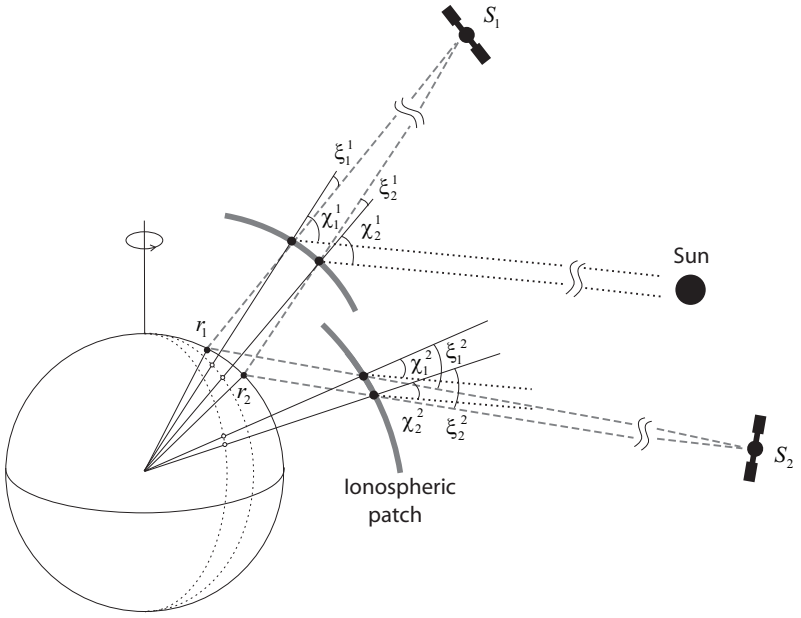

Figure 7.4: Illustration of the geometry between single baseline and two satellite and the Sun.

Single-difference TID observation Subtracting equation 7.16 from equation 7.15 leads to detrended single-differenced slant ionospheric delay that it can be known as the single-difference TID observation (denoted by $Dsd\tilde{I}$),

$$Dsd\tilde{I}^s_{r_1,r_2}(t) = sd\tilde{I}^s_{r_1,r_2}(t) - sdG^s_{r_1,r_2}(t) = sdTID(t) + \epsilon(t) \tag{7.17}$$

Figure 7.5 shows time series of the between-receiver single-difference slant ionospheric delays measured from baseline of SPK1-RHCL to two satellites PRN 01 and PRN 16, using the same data is in the example of the previous subsection. In first row of the figure, the solid curve is for the measured single-difference slant ionospheric delay $sd\tilde{I}(t)$. The geometry-dependent trend of the time series $sd\tilde{I}(t)$ is computed in two ways: 1) by least-square fitting a second order polynomial to the series (dotted curve) and 2) by least-square fitting of the model in equation 7.16 to the series (dashed curve). The RMS values of the residuals are the same for both fitting methods (0.019 for PRN 01 and 0.036 m for PRN 16).

The second row of figure 7.5 shows the single-difference TID observation $Dsd\tilde{I}^s_r(t)$. The bold-solid curve is corresponding to $Dsd\tilde{I}^s_r(t)$ detrended by fitting equation 7.16 and the dotted curve is for the case that detrending was done at the un-differenced level (taking difference between two stations in figure 7.1). The RMS values corresponding to the dotted curve is 0.020 for PRN 01 and 0.037 m for PRN 16 which are slightly larger than the residuals obtained from fitting equation 7.16. The third and forth rows in figure 7.5 gives the normalized auto-correlation and auto-spectral density functions of $Dsd\tilde{I}^s_r(t)$ for both satellites.

The periods corresponding to the global maximum of the both auto-spectral density functions are in the typical range of period for MS-TID (less than 1 hour). It means that the MS-TID is dominated in the single-differenced time series for PRN 01 and PRN 16. This

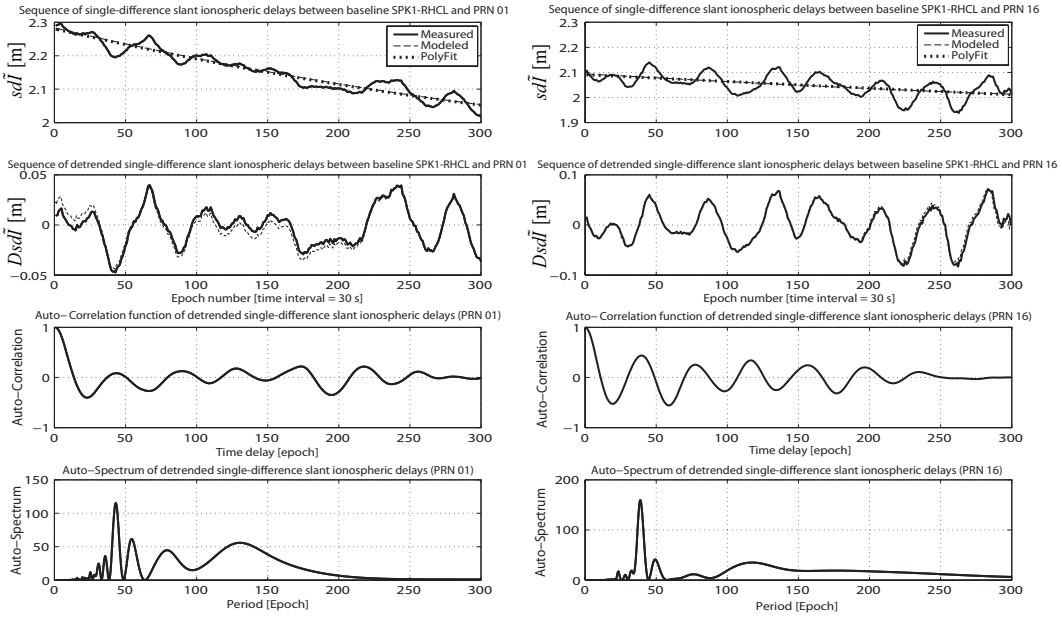

Figure 7.5: Measured and modeled series of single-difference ionospheric delays during 2.5 h timespan (UT=16:00 - 18:30) between baseline SPK1-RHCL (54 km) and PRN 01 (left) and PRN 16 (right). The 'dotted line' in the first row is a 2nd order polynomial fit the 'dashed line' is the modeled series by fitting the physical model of equation 7.16. The second row shows the detrended series (measured series minus modeled series). The third and forth rows show the normalized auto-correlation function and corresponding auto-spectral density function of the detrended series, respectively.

is an important result that single differencing can be used for detecting and analyzing the MS-TID.

The single difference is the starting point for this research. For completeness, also the double difference is discussed, although it will not be used further in this research.

7.4.4 Double-difference TID observation

As with the single-difference slant ionospheric delay decompositions, the double-difference slant ionospheric delay can be written as follows

$$dd\tilde{I}^{s_1,s_2}_{r_1,r_2}(t) = \underbrace{ddG^{s_1,s_2}_{r_1,r_2}(t)}_{\text{Geometry-dependent Trend}} + ddTID(t) + \underbrace{\epsilon(t)}_{\text{Residual}} \tag{7.18}$$

with $dd\tilde{I}^{s_1,s_2}_{r_1,r_2}(t) = sd\tilde{I}^{s_2}_{r_1,r_2}(t) - sd\tilde{I}^{s_1}_{r_1,r_2}(t)$ the double-difference slant ionospheric delay between two receivers r_1 and r_2 and two satellites s_1 and s_2 at epoch t, and where ϵ denotes both measurement random error and double-difference SS-TID with $E\{\epsilon\} = 0$. The $ddTID$ expresses the combination of two single-difference MS-TIDs for two different ionospheric patches. In equation 7.18, $ddG^{s_1,s_2}_{r_1,r_2}(t)$ expresses the double-difference geometry-dependent trend (and horizontal ionospheric gradients in the two ionospheric patches) that can be obtained from subtracting two associated single-difference geometry-

dependent trend,

$$
\begin{aligned}
ddG^{s_1,s_2}_{r_1,r_2}(t_i) \;&=\; sdG^{s_2}_{r_1,r_2}(t_i) \;-\; sdG^{s_1}_{r_1,r_2}(t_i) \\
&\doteq e^{-\beta_2 \Delta t}\; G^{s_2}_{r_1,r_2}(t_{i-1}) + \alpha \cdot m\left(\bar{\xi}^{\,s_2}_{IP}(t_i)\right) \cdot A(\bar{\chi}^{\,s_2}_{r_1,r_2})\; \frac{\left(1-e^{-\beta_2 \Delta t}\right)}{e\;\beta_2\;H_F\;10^{16}}\; \phi^{photon}_{\infty,2} \\
&-\; e^{-\beta_1 \Delta t}\; G^{s_1}_{r_1,r_2}(t_{i-1}) + \alpha \cdot m\left(\bar{\xi}^{\,s_1}_{IP}(t_i)\right) \cdot A(\bar{\chi}^{\,s_1}_{r_1,r_2})\; \frac{\left(1-e^{-\beta_1 \Delta t}\right)}{e\;\beta_1\;H_F\;10^{16}}\; \phi^{photon}_{\infty,1}
\end{aligned}
\tag{7.19}
$$

where β and ϕ^{photon}_{∞} for two different ionospheric patches of satellites s_1 and s_2 are specified by sub-indexes 1 and 2, respectively. The β and ϕ^{photon}_{∞} are not the same for both patches because the ionospheric patches associated to two satellites may be located for apart and their properties can be totally differed. Figure 7.4 shows the geometry between a baseline and two satellites and the Sun. Since $\bar{\xi}_{IP}$ and $\bar{\chi}_{r_1,r_2}$ are different for two satellites, unfortunately it is not possible to simplify equation 7.19.

In order to estimate the model parameters $x = \left[\beta_1, \phi_{\infty,1}, G^{s_1}_{r_1,r_2}(t_0), \beta_2, \phi_{\infty,2}, G^{s_2}_{r_1,r_2}(t_0)\right]^T$, one should apply the least-squares fitting of the single-difference model individually for the data of the two ionospheric patches.

Subtracting equation 7.19 from 7.18 results in the *double-difference TID observation* (or detrended double-difference slant ionospheric delay)

$$
Ddd\tilde{I}^{s_1,s_2}_{r_1,r_2}(t) \;=\; dd\tilde{I}^{s_1,s_2}_{r_1,r_2}(t) \;-\; ddG^{s_1,s_2}_{r_1,r_2}(t) \;=\; ddTID(t) \;+\; \epsilon(t) \tag{7.20}
$$

where $Ddd\tilde{I}$ stands for the double-difference TID observation.

The time series of the double-difference TID observations, using the same data in the previous subsection, for baseline SPK1-RHCL and two satellites PRN 01 and PRN 16, is shown in figure 7.6. In the first row, the measured (solid curve) and modeled values (dashed and dotted curves) of $dd\tilde{I}(t)$ are shown. The second, third, and forth rows show, respectively, the double-difference TID observations and corresponding normalized auto-correlation function and auto-spectral density function. As mentioned $Ddd\tilde{I}$, in fact, represents the combination of two MS-TIDs that propagating in two different ionospheric patches. Because of this, the behavior of $Ddd\tilde{I}$ becomes more complicated. In the example, periods (or frequencies) of the two MS-TIDs are close to each other and consequently the combined spectrum of their combination looks very much the same as the individually obtained auto-spectral density functions for the two single-difference TID observations. This is rather a coincidence and it is not always that both MS-TIDs have the same frequencies.

7.5 TID observation equation

In the previous section, the double-, single-, and un-differenced TID observations were derived using both the code and phase geometry-free linear combinations. In this section, a mathematical model for the TID observations is derived. For this purpose, it is supposed that a TID wavelike oscillations of the ionosphere propagates continuously in a *horizontal*

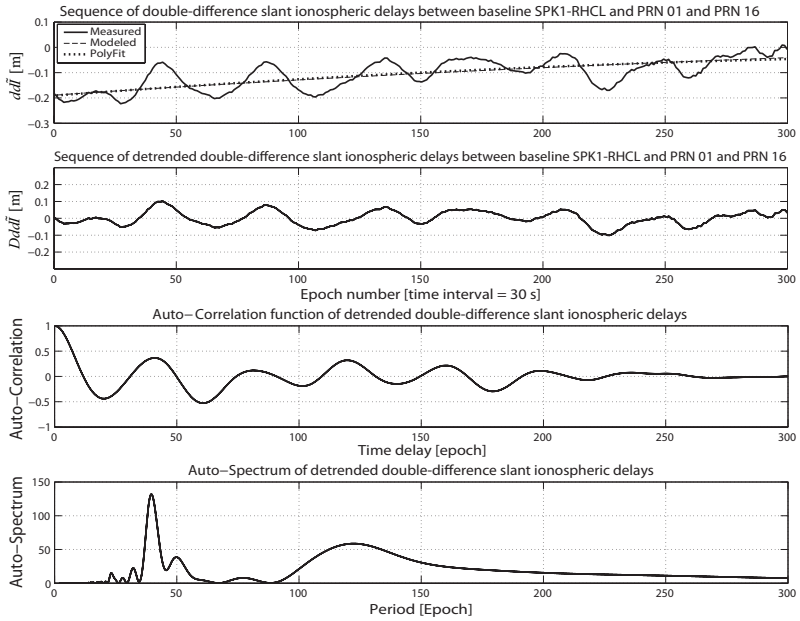

Figure 7.6: Measured and modeled series of double-difference ionospheric delays during 2.5 h timespan (UT=16:00 - 18:30) between baseline SPK1-RHCL (54 km) and PRN 01 and PRN 16. The 'dotted line' in the top figure is 2nd order polynomial fitted to the data and the 'dashed line' is the modeled series by fitting the physical model of equation 7.19. The second plot shows the detrended series (measured minus modeled). The third and forth plots show the normalized auto-correlation function and corresponding auto-spectral density function of the detrended series, respectively.

ionospheric patch as a *monochromatic and perfect periodic* longitudinal plane wave. Taking satellite motion into account, the spatio-temporal characteristics of the TID wave, in a patch between receiver r and satellite s, can be modeled by the traveling plane wave equation 7.1 as follows

$$D\tilde{I}_r^s(t,\vec{r}) = A\ \sin\left(\omega t - \vec{K}\cdot\vec{r}(t)\right) \tag{7.21}$$

where $\omega = 2\pi f$ and A are respectively the angular frequency and amplitude of the TID wave and $\vec{K} = \frac{2\pi}{\lambda}\frac{\vec{V}_{ph}}{V_{ph}} = \omega\frac{\vec{V}_{ph}}{V_{ph}^2}$ is the *TID wave vector* with λ as wavelength and $\vec{V}_{ph}$ is vector of phase velocity of the TID wave. In this equation, $\vec{r}(t)$ is the horizontal position vector of the IP point in the ionospheric thin layer with origin located at the point of TID source (see figure 7.7).

7.5.1 Doppler-shift on TID observation

The motion of GPS satellites in space causes the IP point between receiver and satellite to move as well and therefore the position vector of IP becomes time-dependent. This means that the observation point of the TID wave is moving with respect to the source of the TID which results in a "*Doppler-shift*" in the TID observations. If the position vector at

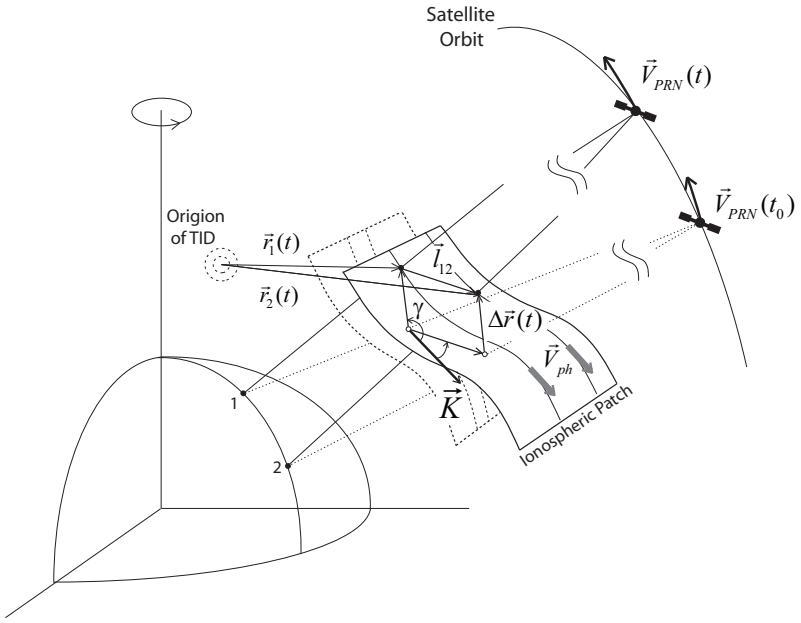

Figure 7.7: Illustration of the satellite motion and the TID propagation in the ionospheric patch.

each epoch is written as $\vec{r}(t) = \vec{r}(t_0) + \Delta\vec{r}(t)$ then equation 7.21 yields

$$D\tilde{I}_r^s(t,\vec{r}) = A\ \sin\left(\omega t - \underbrace{\vec{K}\cdot\Delta\vec{r}(t)}_{\text{Doppler-shift}}\ -\ \varphi_0\right) \tag{7.22}$$

where $\varphi_0 = \vec{K}\cdot\vec{r}(t_0)$ is *initial phase* of the TID wave at reference time t_0 and $\vec{K}\cdot\Delta\vec{r}(t)$ stands for the Doppler-shift. If velocity of IP point is denoted as $\vec{V}_{IP}(t) = \frac{\Delta\vec{r}(t)}{t-t_0}$ then, using $\vec{K} = \omega\frac{\vec{V}_{ph}}{V_{ph}^2}$, the inner product in the Doppler-shift term becomes $\frac{V_{IP}}{V_{ph}}\ \omega(t-t_0)\ \cos(\gamma(t))$ that leads to (with $t_0 = 0$)

$$D\tilde{I}_r^s(t,\vec{r}) = A\ \sin\left(\left(1 - \frac{V_{IP}(t)\ \cos(\gamma(t))}{V_{ph}}\right)\omega t\ -\ \varphi_0\right) \tag{7.23}$$

where $\gamma(t)$ denotes time dependent angle between TID propagation direction and moving direction of IP point (see figure 7.7). It is seen from equation 7.23, the observed frequency of the TID wave depends on the projection of $\vec{V}_{IP}$ on TID wave propagation direction. There are two special cases: (1) If $V_{ph} = V_{IP}(t)\ \cos(\gamma(t))$ then TID observations becomes *constant* over time (2) in case that $V_{ph} = -V_{IP}(t)\ \cos(\gamma(t))$ results in a *standing wave* form of TID in the observations.

Equation 7.23 can be considered as the *observation equation* for the least-squares estimation of the TID wave parameters.

7.5.2 Single-difference TID observation equation

Suppose two receivers on the ground are tracking a GPS satellite and a monochromatic plane wave TID propagates in the associated ionospheric patch (see figure 7.7). The

position vectors of the two associated IP points, with origin at the TID source, can be written as follows

$$\begin{aligned} \vec{r}_1(t) &= \vec{r}_1(t_0) + \Delta\vec{r}_1(t) \\ \vec{r}_2(t) &= \vec{r}_2(t_0) + \Delta\vec{r}_2(t) = \vec{r}_1(t_0) + \vec{l}_{12}(t_0) + \Delta\vec{r}_2(t) \end{aligned} \tag{7.24}$$

where $\vec{l}_{12}$ is baseline vector between two IP points on the thin ionosphere layer. Then, using equation 7.21, the TID observation equations for two IP points at time t are (with $t_0 = 0$)

$$\begin{aligned} D\tilde{I}_1^s(t,\vec{r}_1) &= A\ \sin\left(\omega t - \varphi_0 - \vec{K}\cdot\Delta\vec{r}_1(t)\right) \\ D\tilde{I}_2^s(t,\vec{r}_2) &= A\ \sin\left(\omega t - \varphi_0 - \vec{K}\cdot\Delta\vec{r}_2(t) - \underbrace{\vec{K}\cdot\vec{l}_{12}(t)}_{\text{Phase-Offset}}\right) \end{aligned} \tag{7.25}$$

where $\varphi_0 = \vec{K}\cdot\vec{r}_1(t_0)$ is an initial phase at the first IP point (as reference IP) and $\vec{K}\cdot\vec{l}_{12}(t)$ an additional *phase-offset* caused by distance $\vec{l}_{12}$ between two IP points on the ionosphere thin shell. We may assume for short period of time that $\Delta\vec{r}(t) = \Delta\vec{r}_1(t) \approx \Delta\vec{r}_2(t)$ and accordingly $\vec{l}_{12}(t) = \vec{l}_{12}$. Subtracting two TID observation equations in 7.25 gives the single-difference TID observation equation

$$Dsd\tilde{I}_{12}^s(t,\vec{l}_{12}) = 2A\ \sin\left(-\frac{1}{2}\vec{K}\cdot\vec{l}_{12}\right)\cos\left(\omega t - \varphi_0 - \vec{K}\cdot\Delta\vec{r}(t) - \frac{1}{2}\vec{K}\cdot\vec{l}_{12}\right) \tag{7.26}$$

Substituting $\vec{K}\cdot\Delta\vec{r}(t) = \frac{V_{IP}(t)\cos(\gamma(t))}{V_{ph}}\ \omega t$ into equation 7.26 yields

$$Dsd\tilde{I}_{12}^s(t,\vec{l}_{12}) = 2A\ \sin\left(-\frac{1}{2}\vec{K}\cdot\vec{l}_{12}\right)\cos\left(\left(1 - \underbrace{\frac{V_{IP}(t)\cos(\gamma(t))}{V_{ph}}}_{\text{Doppler-effect}}\right)\omega t - \varphi_0 - \underbrace{\frac{1}{2}\vec{K}\cdot\vec{l}_{12}}_{\text{Phase-Offset}}\right) \tag{7.27}$$

The between-receiver differencing of the TID observation produces a *new monochromatic wave*. It is clear from equation 7.27 that frequency and propagation direction of the new wave are the same as those for the original TID wave in equation 7.23. The term of $2A\ \sin\left(-\frac{1}{2}\vec{K}\cdot\vec{l}_{12}\right)$ in equation 7.27 is the amplitude of the new wave, the amplitude of the new wave is amplified by factor $2\sin\left(-\frac{1}{2}\vec{K}\cdot\vec{l}_{12}\right)$ compared to amplitude of TID observation (A) and it is also a function of baseline vector $\vec{l}$. Another important point is that the phase-offset for the new wave is halved. This implies that the phase velocity (or wavelength) is amplified by single differencing by a factor two.

Since the satellite velocity vector $\vec{V}_{PRN}(t)$ in equation 7.27 is time-dependent, not only the velocity of the IP point becomes as a function of time $V_{IP}(t)$, but also the azimuth of $\Delta\vec{r}$ is changing in time by the motion of IP point (the *convergence of meridians*), resulting in a time-dependent angle $\gamma(t)$ between the TID wave vector $\vec{K}$ and $\Delta\vec{r}$.

So far, we derived the mathematical equations for the un-differenced and single-differenced TID observations. The parameters in the observation equation are the TID wave parameters, i.e. *frequency* f, *wavelength* λ, *phase velocity* V_{ph}, *amplitude* A, *initial phase* φ_0, and azimuth of propagation Az. The next section deals with methods for the least-squares estimation of these TID wave parameters.

7.6 Estimation of TID wave parameters

This section deals with a method for estimating the MS-TID wave parameters using the single difference observations. The main reason to use the single differenced data is that between-receiver single-differencing leads to data in which the effect of LS-TID and geometry-dependent trend in data are significantly reduced therefore the detrending of data can be done better than the case using un-differenced data. All previously developed methods for estimation of the TID wave parameters use the un-differenced TID observations, e.g. (Lejeune and Warnant, 2008; Garrison et al., 2007; Hernandez-Pajares et al., 2006a; Calais et al., 2003; Afraimovich et al., 2002). In our method for estimating TID wave parameters, the single-difference TID observations are used and the method takes the high temporal resolution of GPS data into account. The period of the TID is the first parameter that should be determined. Once the period has been determined it will be possible to determine the wave vector and the amplitude of TID.

7.6.1 Period determination

Consider a *stationary* (*ergodic*) *random process* $\underline{x}(t)$ that it is defined by summation of a sine wave with circular frequency f_0 and *wide-band random noise* $\underline{e}(t)$ with a uniform auto-spectral density function $S_{\underline{ee}}$ over bandwidth f_c (*cut-off frequency*), i.e. $S_{\underline{ee}}(f) = S$ for $0 \leq f \leq f_c$ and elsewhere $S_{\underline{ee}}(f) = 0$, as follows (randomness indicated by underscore)

$$\underline{x}(t) = a\sin(2\pi f_0 t + \varphi) + \underline{e}(t) \tag{7.28}$$

The auto-correlation function $R_{\underline{xx}}(\tau)$ of the random process $\underline{x}(t)$ is the sum of the auto-correlations for the sine wave and the wide-band random noise

$$\begin{aligned} R_{\underline{xx}}(\tau) &= \lim_{T\to\infty} \tfrac{1}{T}\int_0^T \underline{x}(t)\underline{x}(t-\tau)dt \\ &= \tfrac{a^2}{2}\cos(2\pi f_0\tau) \;+\; Sf_c\left(\tfrac{\sin(2\pi f_c\tau)}{2\pi f_c\tau}\right) \end{aligned} \tag{7.29}$$

where the first term is the auto-correlation function of the sine wave and the second term the auto-correlation function of the wide-band noise. The auto-correlation function of the sine wave is a cosine with frequency f_0 and an amplitude equal to the *mean square value* of the original sine wave ($\frac{a^2}{2}$). Because of the amplification in the amplitude, periodic fluctuations of the original wave can be recognized more clearly in the pattern of the auto-correlation function than the original time series. The auto-correlation function of the wide-band random noise is a *Sinc* function that decreases very rapidly after a first crossing

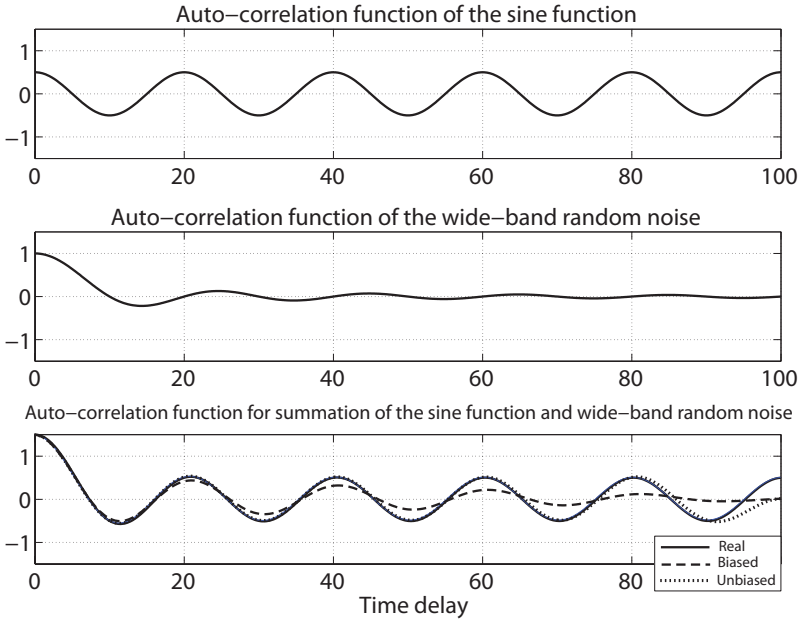

Figure 7.8: A typical auto-correlation function for a sine function (top), for wide-band random noise (middle), and for a sine function plus wide-band random noise (bottom). The 'solid', 'dashed' and 'dotted' curves are respectively for real, biased and unbiased auto-correlation function.

at $\tau = \frac{1}{2f_c}$. Because of this, $R_{xx}(\tau)$ decays quickly to the cosine term that describes the sinusoidal portion of the random process (see figure 7.8). This *key property* suggests that the auto-correlation function of $\underline{x}(t)$ (in 7.28) can be used to measure the *period* of original sine wave $\frac{1}{f_0}$.

Now, suppose $x(t_k),\ k = 1, \ldots, n$ is a time series of length T with n *equally spaced* measured values from the random process $\underline{x}(t)$. The sampling interval (*time-resolution*) is $\Delta t = \frac{T}{n}$. Because of the limited duration of the series (*truncation*), it is not possible to see signature of all lower frequencies in the series. Hence, the *minimum frequency-resolution bandwidth* available from the series is $\frac{1}{T}$, also known as the *fundamental frequency*. Due to *discretization*, the *maximum frequency* that will appear in the series is well-known *Nyquist cutoff frequency* $f_c = \frac{1}{2\Delta t} = \frac{n}{2\,T}$. Hence, the sinusoidal portion of random process $\underline{x}(t)$ with frequency of f_0, can only be recognized in the series if $n \geq 2T f_0$ (where f_0 is in Hertz if T is in second).

An *unbiased* estimate of the auto-correlation function of a *stationary ergodic* random process $\underline{x}(t)$ based on a single time series $x(t)\ ,\ 0 \leq t \leq T$, is given by (Bendat and Piersol, 1971)

$$R_{xx}(\tau) = \frac{1}{T-\tau} \int_0^{T-\tau} x(t)x(t-\tau)dt \tag{7.30}$$

If $\frac{1}{T-\tau}$ is replaced by $\frac{1}{T}$ this results in a *biased* estimate of the auto-correlation function of $R_{\underline{xx}}(\tau)$. In figure 7.8, typical biased and unbiased auto-correlation functions together with the the auto-correlation function of a pure sine function (real one) are shown. The unbiased auto-correlation function starts to deviate from the real one at about $\tau = T - \frac{1}{f_0}$, while the biased auto-correlation function is deviated right from the start. As a consequence,

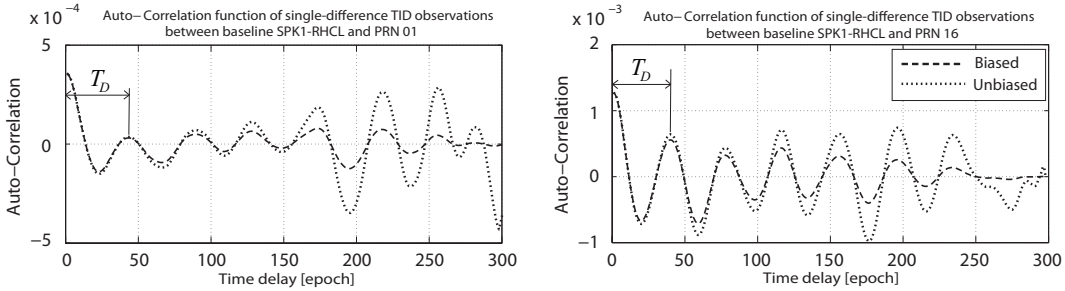

Figure 7.9: Biased (dashed curve) and unbiased (dotted curve) auto-correlation functions for series of single-difference TID observations for a 2.5 h timespan (UT=16:00 - 18:30) between baseline SPK1-RHCL (54 km) and PRN 01 (left) and PRN 16 (right).

the absolute values of the extermum points for the unbiased auto-correlation function are larger than those for the biased auto-correlation function. Because of this to measure the period of the original sine wave ($\frac{1}{f_0}$), the unbiased auto-correlation function is likely to give the best results.

In this research, the time series of single-difference TID observations, of subsection 7.4.3, $Dsd\tilde{I}^s_{r_1,r_2}(t)$ for $k = 1, \ldots, n$ is considered to be a *realization* of the stationary (ergodic) random function defined in equation 7.28. In the other words, the single-difference MS-TID wave is modeled as a sine wave over $t_1, \ldots, t_k$. The short period TID, residuals associated to the LS-TID and to the geometry-dependent trend are considered to represent themselves like a wide-band random noise with a uniform auto-spectral density function over bandwidth $0 \leq f \leq f_c$. It is supposed that the mean square value of the which in bandwidth $0 \leq f \leq f_c$ is not significant.

As an example, figure 7.9 shows both the biased (dashed curve) and unbiased (dotted curve) auto-correlation functions for the single-difference TID observations between baseline SPK1-RHCL and PRN 01 (left subfigure) and PRN 16 (right subfigure) over a 2.5 h timespan (UT=16:00 - 18:30) in 2003.10.28. We should accept that a TID wave is not a perfect periodic wave like the sine function. This is why the computed unbiased auto-correlation functions, in figure 7.9, don't behave like the cosine function. If the TID behaves more like as a perfect wave then the auto-correlation function tends to be more like the cosine function. As observed earlier in figure 7.5 (2nd row), the TID observations for PRN 16 more close by a sinusoidal perfect wave than those for PRN 01. Consequently the corresponding auto-correlation function for PRN 16 tends to be more like a cosine function. The amplitude of the auto-correlation function is in lower for PRN 01 than for PRN 16.

The period of the TID wave is measured by the distance between the two first consecutive maximum points of the auto-correlation function. In case of PRN 01, the measured period is $T_{_D} = 44$ epochs (22 minutes). The measured TID period for PRN 16 is $T_{_D} = 40$ epochs (20 minutes). It is important to note that the period measured from the auto-correlation function is, in fact, a period which is affected by the Doppler-shift. This is the so-called *Dopplered-period* and indicated by $T_{_D}$. Accordingly, the obtained TID frequency $f_{_D} = \frac{1}{T_{_D}}$ is called the *Dopplered-frequency* (it is in Hertz if $T_{_D}$ is in seconds).

7.6.2 TID wave vector determination

The goal here is to estimate propagation parameters of the MS-TID wave such as *phase velocity* V_{ph} and azimuth of *propagation direction* $Az_{_{TID}}$ using between-receiver single-difference TID observations ($Dsd\tilde{I}$) provided in subsection 7.4.3. Once these two parameters are estimated the corresponding TID wavelength is computed by $\lambda_{_{TID}} = \frac{V_{ph}}{f}$. Accordingly the TID wave vector is determined by $\vec{K} = \frac{2\pi}{\lambda_{_{TID}}} \frac{\vec{V}_{ph}}{V_{ph}}$. Based on the time and spatial resolutions of the TID observations, the propagation parameters can be estimated by two methods.

- **High spatial resolution data** When the spatial resolution of the TID observations is high the spatial gradient of the TID, computed from the network data, can be used to estimate the propagation parameters (Afraimovich et al., 1998; Mercier, 1986). In this case the separation between the network stations should be much smaller than the typical wavelength of MS-TID.
- **High time resolution data** In case that the spatial resolution of data is low but the high temporal resolution data is available, the propagation parameters can be estimated based on the measured time delays by the cross-correlation between data for the network stations, (Hernandez-Pajares et al., 2006a; Garrison et al., 2007).

The method to be presented in this chapter uses the cross-correlation between many pairs of time series of the single-difference TID observations provided from receivers in the network. This is because of the fact that the network of permanent reference stations in the Network-RTK technique is always a medium-scale network (baseline length less than a few hundred kilometers) and density of the stations in space is not enough to determine precisely the spatial gradient of the data.

Single-difference TID wave vector estimation Let us consider a network with n receivers on the ground that are tracking GPS satellite s during a certain timespan. Suppose during the tracking the satellite a MS-TID propagates in the ionospheric patch associated to the IP points of the network. Therefore, $n-1$ time series of the single-difference TID observations between a selected reference receiver and the other receivers can be provided. The $n-1$ measured time series are spatially correlated by the traveling TID wave. Then, it follows that $n-2$ time delays can be measured by the cross-correlation between the time series of a selected reference baseline and other baselines in the network. These time delays are constrained by the geometry of the network of IP points and the TID wave vector.

In order to estimate the TID propagation parameters, a mathematical equation for a time delay measured by the cross-correlation is needed. For a simple case, the geometry between four moving IP points (with velocity of $\vec{V}_{_{IP}}$) and phase velocity of a monochromatic plane wave TID ($\vec{V}_{_{ph}}$) is shown in figure 7.10 on the ionospheric thin layer. The first IP point is considered as a reference point for three baselines (on the thin layer) between pairs of the IP points. For the sake of simplicity, the origin of the coordinate system is located at the reference IP point.

A time delay measured by cross-correlation between the first baseline (as a selected reference baseline) and another baseline in the network is denoted by $\tau_{_{2i}}$ where subscript 2 is for

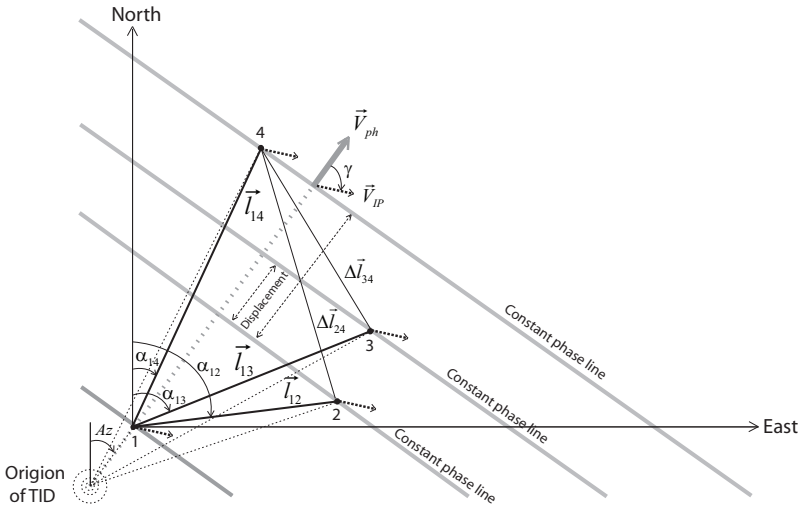

Figure 7.10: Geometry between a monochromatic plane wave TID and a network of IP points moving over the ionospheric thin layer.

second receiver of the reference baseline $\vec{l}_{12}$ and subscript $i = 3, \ldots, n$ indicates the second receiver of the other baselines in the network (the reference receiver 1 is as the first receiver for the baselines). In fact, the time delay τ_{2i} gives the traveling time of the single-difference TID wave between the moving IP points 2 and i.

To derive a relationship between the measured time delay τ_{2i} and the TID phase velocity, it is necessary to have the displacement of the wavefront during the time delay τ_{2i}. As mentioned in subsection 7.5.2, between-receiver differencing of the TID wave (with frequency f and wave number K) results in a new wave with the same frequency but with its wave number is halved $\frac{K}{2}$. The phase velocity of the new wave is therefore amplified by factor 2. Consequently, by taking motion of the IP points into account, the velocity of the wavefront with respect to the IP points is obtained by $\vec{V} = 2\vec{V}_{ph} + \vec{V}_{IP} \cdot \frac{\vec{V}_{ph}}{V_{ph}}$. The time delay is then

$$\tau_{2i}(t) = \Delta\vec{l}_{2i}(t) \cdot \frac{\vec{V}(t)}{V(t)^2} \tag{7.31}$$

with $\Delta\vec{l}_{2i} = \vec{l}_{1i} - \vec{l}_{12}$. The corresponding scalar expression for equation 7.31 can be written as follows (subscripts $_E$ and $_N$ denote, respectively, the Easting and Northing of the vectors)

$$\tau_{2i}(t) = \Delta l_{2i,E}(t)\ v(t)\ +\ \Delta l_{2i,N}(t)\ w(t) \tag{7.32}$$

where $v = \frac{V_E}{V^2}$ and $w = \frac{V_N}{V^2}$ with $V^2 = V_E^2 + V_N^2$ and two components of $\Delta\vec{l}_{2i}(t)$ are obtained as follows

$$\begin{aligned} \Delta l_{2i,E}(t) &= \vec{l}_{1i} \sin\left(\alpha_{1i}(t)\right)\ -\ \vec{l}_{12} \sin\left(\alpha_{12}(t)\right) \\ \Delta l_{2i,N}(t) &= \vec{l}_{1i} \cos\left(\alpha_{1i}(t)\right)\ -\ \vec{l}_{12} \cos\left(\alpha_{12}(t)\right) \end{aligned} \tag{7.33}$$

where α denotes azimuth of ionospheric baseline which is a function of time.

Equation 7.32 is a linear observation equation for the time delay, measured by cross-correlation, with two unknown parameters v and w. For each epoch the parameters v and

w can be estimated by the least-squares using the following functional model

$$\underbrace{\begin{bmatrix} \tau_{23}(t) \\ \tau_{24}(t) \\ \vdots \\ \tau_{2n}(t) \end{bmatrix}}_{(n-2)\times 1} = \underbrace{\begin{bmatrix} \Delta l_{23,E}(t) & \Delta l_{23,N}(t) \\ \Delta l_{24,E}(t) & \Delta l_{24,N}(t) \\ \vdots & \vdots \\ \Delta l_{2n,E}(t) & \Delta l_{2n,N}(t) \end{bmatrix}}_{(n-2)\times 2} \cdot \underbrace{\begin{bmatrix} v(t) \\ w(t) \end{bmatrix}}_{2\times 1} \tag{7.34}$$

where at least two time delay measurements ($n = 4$) are needed to obtain the unknown vector of $[v(t)\ ,\ w(t)]^T$.

Once v and w are estimated the azimuth of the TID propagation direction can be determined from $Az_{TID}(t) = \arctan\left(\frac{v(t)}{w(t)}\right)$. The E-W and N-S components of the velocity vector of the wave with respect to the moving IP points $\vec{V}$ are given as follows

$$\begin{aligned} V_E(t) &= \tfrac{v(t)}{v(t)^2+w(t)^2} \\ V_N(t) &= \tfrac{w(t)}{v(t)^2+w(t)^2} \end{aligned} \tag{7.35}$$

It is important to know that the TID propagation direction is invariant with respect to the motion of the IP points. The E-W and N-S components of $\vec{V}_{ph}$ are then

$$\begin{aligned} V_{ph,E}(t) &= \tfrac{1}{2}\,(\ V_E(t)\ -\ V_{IP}(t)\cos(\gamma(t))\sin(Az_{TID}(t))\ ) \\ V_{ph,N}(t) &= \tfrac{1}{2}\,(\ V_N(t)\ -\ V_{IP}(t)\cos(\gamma(t))\cos(Az_{TID}(t))\ ) \end{aligned} \tag{7.36}$$

where $\gamma(t) = \alpha(t) - Az_{TID}(t)$ denotes angle between moving direction of the IP points $\alpha(t)$ and direction of the TID propagation (see figure 7.10) and $V_{ph}(t) = \sqrt{V_{ph,E}^2(t) + V_{ph,N}^2(t)}$ is phase velocity of the TID wave. Now, the actual frequency f_{TID} and wavelength λ_{TID} of the TID wave can be computed as follows

$$\begin{aligned} f_{TID}(t) &= f_D\left(\tfrac{V_{ph}(t)}{V_{ph}(t) - V_{IP}(t)\cos(\gamma(t))}\right) \\ \lambda_{TID}(t) &= T_{TID}\ V_{ph}(t) \end{aligned} \tag{7.37}$$

where $T_{TID} = \frac{1}{f_{TID}}$ is actual period of the TID wave and f_D is the Dopplered-frequency. Finally, the TID wave vector is determined by $\vec{K} = \frac{2\pi}{\lambda_{TID}}\frac{\vec{V}_{ph}}{V_{ph}}$.

7.6.3 TID wave amplitude determination

Once the period and propagation parameters of the TID wave are estimated then the amplitude of the TID wave can be computed. For this purpose, the amplitude of the TID

wave, equation 7.27 is written as follows

$$Dsd\tilde{I}^{s}_{1i}(t,\vec{l}_{1i}) = \underbrace{2A(t)\ \sin\left(-\frac{1}{2}\vec{K}\cdot\vec{l}_{1i}\right)}_{A_{sd}(t,\vec{l}_{1i})}\cos\left(\psi(t,\vec{l}_{1i})-\varphi_0(t)\right) \tag{7.38}$$

with

$$\psi(t,\vec{l}_{1i}) = \left(1-\frac{V_{IP}(t)\cos(\gamma(t))}{V_{ph}}\right)\omega t - \frac{1}{2}\vec{K}\cdot\vec{l}_{12} \tag{7.39}$$

where $i = 2,\ldots,n$ stands for the different ionospheric baselines in the network (first receiver is selected as a reference). The single-difference TID wave amplitude is denoted by $A_{sd}(t,\vec{l}_{1i})$ to emphasize its temporal and spatial dependency. Both the unknown initial phase $\varphi_0(t)$ (at reference receiver) and amplitude $A(t)$ of the TID wave are time-dependent. To estimate these two unknown parameters, equation 7.38 can be re-parameterized as follows

$$\begin{aligned} Dsd\tilde{I}^{s}_{1i}(t,\vec{l}_{1i}) = a(t)&\left[2\cos\left(\psi(t,\vec{l}_{1i})\right)\sin\left(-\tfrac{1}{2}\vec{K}\cdot\vec{l}_{1i}\right)\right] \\ +\ b(t)&\left[2\sin\left(\psi(t,\vec{l}_{1i})\right)\sin\left(-\tfrac{1}{2}\vec{K}\cdot\vec{l}_{1i}\right)\right] \end{aligned} \tag{7.40}$$

where this equation is a linear observation equation with respect to new unknown parameters $a(t) = A(t)\cos(\varphi_0(t))$ and $b(t) = A(t)\sin(\varphi_0(t))$. These parameters can be estimated in the normal least-squares adjustment process either in the *time domain* or in the *space domain*. For epoch by epoch estimation of the parameters, it is necessary to set up $(n-1)$-vector of the single-difference TID observations in the space domain

$$y(t) = \begin{bmatrix} Dsd\tilde{I}^{s}_{12}(t) \\ Dsd\tilde{I}^{s}_{13}(t) \\ \vdots \\ Dsd\tilde{I}^{s}_{1n}(t) \end{bmatrix}_{(n-1)\times 1} \tag{7.41}$$

To estimate the parameters in the time domain, the observation vector has to consist of data from the initial to current epochs. By taking into account the data from earlier epochs the redundancy is significantly increased. In general, the observation vector for epochs $1,\ldots,k$ is denoted as follows

$$Y = \begin{bmatrix} y(t_1) \\ y(t_2) \\ \vdots \\ y(t_k) \end{bmatrix}_{k(n-1)\times 1} \tag{7.42}$$

with $y(t)$ from equation 7.41 (for $n-1$ baselines in the network with n receivers). Once the least-squares estimates of $a(t)$ and $b(t)$ are obtained the amplitude and initial phase of the TID wave are computed by

$$\begin{aligned} A(t) &= \sqrt{a(t)^2+b(t)^2} \\ \varphi_0(t) &= \arctan\left(\tfrac{b(t)}{a(t)}\right) \end{aligned} \tag{7.43}$$

This completes the estimation of the TID wave parameters using the between-receiver single-difference TID observations in the network. In the next section, we concentrate on a procedure for the *real-time* modeling of the MS-TID waves.

7.7 Real-Time Medium-scale TID modeling

Most results on TID analysis in the literatures (e.g. (Hernandez-Pajares et al., 2006a; Afraimovich et al., 2002; Calais et al., 2003)) have been obtained using post-processing of data. In Hernandez-Pajares et al. (2006b), a method proposed for the real-time modeling of the medium-scale TID effect on precise GNSS navigation techniques that takes into account a climatic model of propagation parameters previously determined in their above mentioned dedicated study. In this section we will present a new algorithm for the real-time medium-scale TID modeling that the detection and the parameter estimation of the TID are implemented in real-time. The algorithm will take advantage of the high temporal resolution GPS data that is available from a medium-scale network. The single-difference TID observations time series, which provided by equation 7.17, are the main input for the algorithm.

The real-time detection and monitoring of MS-TID should essentially consists of three steps: *initialization*, TID *detection*, and TID *ending* that can be carried out using individual baseline data. After the detection of TIDs the TID wave parameters have to be estimated. This can only be done using a network of the local GPS receivers. Therefore between the TID detection and ending three additional steps are added to estimate the wave parameters: scraping, cross-correlation and estimation of wave parameters.

Our algorithm for real-time TID monitoring and modeling consists therefore of the six individual steps: *initialization*, *detection*, *scraping*, *cross-correlation*, *estimation*, and *ending* which will be explained in detail in the following subsections.

We will consider a medium-scale GPS network with n dual frequency receivers that simultaneously track a satellite s over a period of time. Between-receiver single-difference TID observations (epoch by epoch) are provided with respect to a selected reference receiver using the procedure of subsection 7.4.3. The minimum number of unique baselines is $n-1$ and number of all possible baselines is $\frac{n(n-1)}{2}$. The algorithm presented here is based on the use of the minimum number of baselines. As a result, $n-1$ time series of the single-difference TID observations $\check{I}_{1i}(t_k)$ are provided for each baseline l_{1i} with $i = 2, \ldots, n$, '*check*' in $\check{I}_{1i}(t_k)$ emphasizes the *equally spaced* time series of the single-difference TID observations in a time span from t_1 to t_k.

7.7.1 Initialization step

The initialization step is needed in order to estimate the geometric dependent trend of the data, and subsequently remove the trend from the data. The MS-TID have typical periods from 10 to 60 minutes. Because the TID detection step requires a full-period of a wave in data, the minimum period of the MS-TID (10 minutes) is considered as a minimum length for the time series. The cycle-slips can extend the duration of the initialization step.

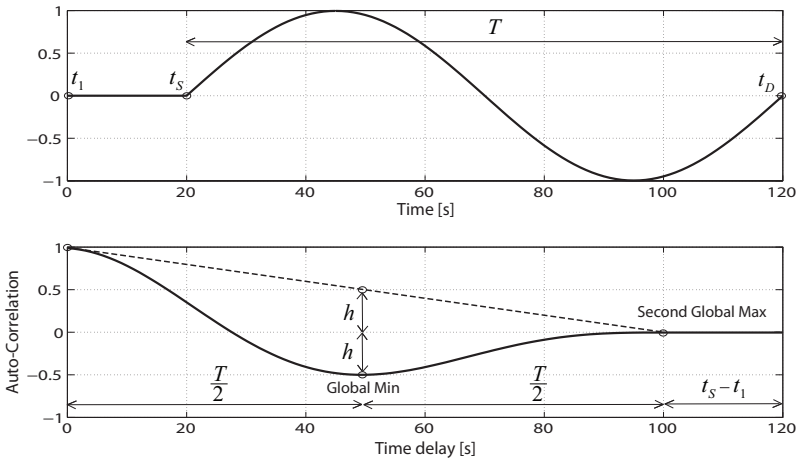

Figure 7.11: Illustration a single-cycle sine function $y(t)$ with period $T = 100\ ^s$ started at t_S and ended at t_D (top) and the corresponding auto-correlation function $R_{yy}(\tau)$ (bottom).

7.7.2 TID detection and scraping steps

After the initialization step a detrended time series $Dsd\tilde{I}^s_{ij}(t)$ is available for $t_1, \ldots, t_k$, which is handed over for the TID detection step. Algorithm for TID detection can be described as follows. Let consider a continuous function $y(t)$ that is defined in a time domain $[t_1, t_D]$ as follows

$$y(t) = \begin{cases} 0 & t_1 \leq t \leq t_S \\ \\ \sin(2\pi f t + \varphi_0) & t_S \leq t \leq t_D \end{cases} \tag{7.44}$$

where t_S and t_D denote respectively the stating and ending times of the sine function with $t_1 \leq t_S \leq t_D$ and $T = \frac{1}{f} = t_D - t_S$ stands for the period of the sine function. Using equation 7.30, the normalized unbiased auto-correlation of $y(t)$ for case that $\varphi_0 = 0$ yields

$$R_{yy}(\tau) = \frac{T - \tau}{T} \cos(2\pi f \tau) \tag{7.45}$$

where the amplitude of the cosine function is reduced linearly with respect to time delay τ.

The function in equation 7.44 and the corresponding auto-correlation function of equation 7.45 are depicted in figure 7.11 for the case that $T = 100\ ^s$. The geometrical properties of $R_{yy}(\tau)$ are of interest for the TID detection. As can be seen from the plot of $R_{yy}(\tau)$, in case of $\varphi_0 = 0$, the *global-minimum* in the auto-correlation is at $\tau = \frac{T}{2}$, which the second *global-maximum* in the auto-correlation occurs at $\tau = T$. Note that the first global-maximum is always at $\tau = 0$.

The geometrical relationships between the first and second global-maximum points and the global-minimum point in $R_{yy}(\tau)$ are *invariant* not only with respect to the initial phase φ_0 but also with respect to the length of the *prelude* associated to the non-sinousidal portion (from t_1 to t_S). This is a very useful property. That is important to know that the geometrical relationships are not sensitive to status that equation 7.44 is contaminated

by a wide-band random noise if mean squared value of the bandwidth of the noise is not considerable. Otherwise the geometry of equation 7.45 will be governed by the noise. The geometrical relationships of the interest are

if the function $y(t)$ is contaminated by a wide-band random noise with non-considerable mean squared value associated to its bandwidth. The geometrical relationships of the interest are

$$\tau_{_{Gmin}} = \frac{1}{2}\,\tau_{_{SGmax}} \tag{7.46}$$

$$|R_{yy}\,(\tau_{_{Gmin}})| = \frac{R_{yy}(0) - R_{yy}\,(\tau_{_{SGmax}})}{2} + R_{yy}\,(\tau_{_{SGmax}}) \tag{7.47}$$

where subscripts $_{Gmin}$, $_{Gmax}$, and $_{SGmax}$ are indicating the global-minimum, first and second global-maximum points, respectively. Two equations 7.46 and 7.47 are the principal conditions that can be exploited for detecting the MS-TID occurrence in a given time series of single-difference TID observations.

Once both of the conditions 7.46 and 7.47 are fulfilled by the normalized unbiased auto-correlation function of a given time series $\check{I}_{1i}(t_k)$ at current epoch t_k (for baseline l_{1i} in the network) and $\tau_{_{SGmax}} \geq 10$ minutes and also the correlation function is smooth enough (no other local extremum points before $\tau_{_{SGmax}}$), one can say that MS-TID is detected. This time event is marked as $t_{_D}$.

Although a TID wave is usually not a perfect sine wave, once a wavelike TID fluctuation appears in a given time series with a completed cycle, the associated normalized unbiased auto-correlation function gets the shape of figure 7.11 and both of the conditions 7.46 and 7.47 will work.

The TID is detected at $t_{_D}$, but this does not provide any information in the start time $t_{_S}$ yet. The period from t_1 tp $t_{_S}$ is called the prelude. Although the geometry of the auto-correlation function is not affected by the prelude, but it is essential to remove the prelude from the series for the cross-correlation step.

Removal of the prelude from a time series in which TID detected should be carried out epoch by epoch. The scraping starts at the first epoch and completes subsequentially others epochs until the conditions 7.46 and 7.47 are fulfilled. This will then be $t_{_S}$. This process is known as *scraping*. The Dopplered-period of TID is then $t_{_D} - t_{_S}$.

In the detection step, it may be occurred that TID is detected in several baselines in the same time. Each of them can be selected as a reference baseline. Among them, the best candidate is a baseline which the period of the TID is larger and the conditions in the equations 7.46 and 7.47 are fulfilled better than those for the others baselines.

7.7.3 Cross correlation step

In this step the time delay between the baselines is computed. It is necessary to exclude at first the baselines which are totally uncorrelated. Therefore, once the MS-TID is detected and the scraping step is completed the question arises that if there are some baselines for which their time series are not spatially correlated to the time series of the reference

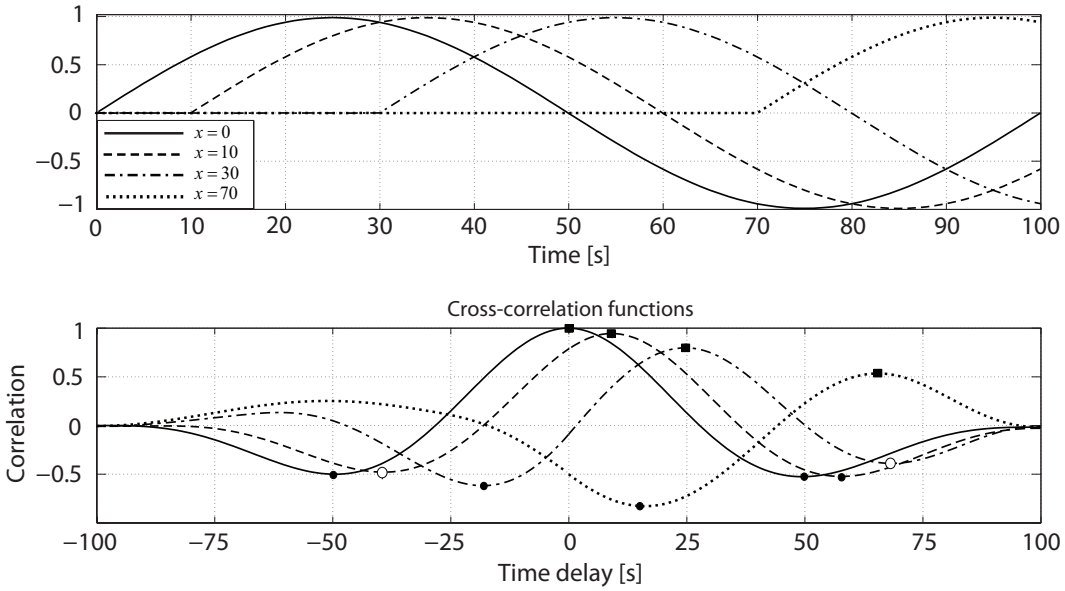

Figure 7.12: Illustration of a simple traveling sine wave $y(t,x)$ with period $T = 100\ s$ and wavelength $\lambda = 100\ m$ (top) started at $t_S = 0$ and ended at $t_D = 100\ s$ for $x = 0$ (solid), $x = 10\ m$ (dashed), $x = 30\ m$ (dash-dotted), $x = 70\ m$ (dotted), the cross-correlation functions between the functions and the reference function $y_0(t)$ are depicted in the bottom for $\tau \in [-T\ ,\ T]$ ('circle' and 'solid circle' denote respectively *local* and *global* minimum points and 'solid square' denotes global maximum point).

baseline. To answer this question, it is necessary to develop an algorithm to check *epoch by epoch* for spatial correlation between the time series in the network. For this, we will take advantage of the following simulation.

Let consider a simple traveling wave that varies both with time t and distance x as follows

$$y(t,x) = \sin(2\pi f t + Kx) \tag{7.48}$$

where $K = \frac{2\pi}{\lambda} = \frac{2\pi f}{V_{ph}}$ stands for wave number with wavelength λ and phase velocity V_{ph}. Lets focus on a particular distance, x, and have a look at the wave $y_x(t)$ passing at distance x. If t_S indicates the starting time of the wave, the continuously observed disturbances (associated to the wave) from t_S to t_D at any point $x \geq 0$ with period $T = t_D - t_S$ can be mathematically expressed as follows

$$y_x(t) = \begin{cases} 0 & t_S \leq t \leq t_S + \tau_x \\ \\ \sin\left(2\pi f(t+\tau_x)\right) & \tau_x + t_S \leq t \leq t_D \end{cases} \tag{7.49}$$

where $\tau_x = \frac{x}{V_{ph}}$ denotes traveling time of the wave between the origin $x = 0$ and a given point and subscript x denotes distance from origin of the wave. This function is depicted in figure 7.12 (top) for $x = 0$, $x = 10$, $x = 30$ and $x = 70\ m$ with $T = 100\ s$ and $\lambda = 100\ m$. In figure 7.12, in the bottom the normalized unbiased cross-correlation functions between reference function $y_0(t)$ and the other functions (i.e. $y_0(t)$, $y_{10}(t)$, $y_{30}(t)$, $y_{70}(t)$) are shown for $\tau \in [-T\ ,\ T]$. The simulation above does not include all the cases that one may deal with in the practice. For instance, the cases for $x \leq 0$ are not shown in the figure.

In the bottom plots of figure 7.12, general features of the cross-correlation patterns can be identified. In the patterns, a 'square' denotes *global* maximum point and 'circle' and 'bold

circle' denote respectively *local* and *global* minimum points. The time delay at the global maximum point τ_{Gmax} gives approximately traveling time of the wave from reference point to a given point in space domain. As expected, an increase in the distance leads to decrease in correlation. This is because the contribution of the sine function is get small when x gets larger. The main common feature for all the patterns, is their smoothness. As second feature is that if $\tau_{Gmax} \geq \frac{T}{2}$ then $\tau_{Gmin} \geq 0$. The other features are that $\tau_{Gmax} \leq \frac{2}{3}T$ results in $R(\tau_{Gmax}) \geq 0.5$ and $R(\tau_{Gmin}) \leq -0.5$ where R stands for the cross-correlation.

There are also some approximate relationships between the global and local extermum points that are fulfilled for all the patterns. For instance in this simulation, $\tau_{Gmax} - \tau_{Gmin} \approx \frac{T}{2}$.

If the above properties are not fulfilled anymore the associated baseline can be excluded in the least-squares estimation of the TID wave parameters at the current epoch. Note that checking for the cross-correlation is implemented epoch by epoch and it may happens that in one epoch a baseline is specified as an uncorrelated baseline and in the next epochs it becomes as a correlated one.

It is important to note that an uncorrelated baseline either indicates that the amplitude of the TID is damped toward second IP point of the baseline, or the TID wavefront deviates strongly from the underlying planar assumption.

7.7.4 TID parameter estimation

This step uses as input a the list of spatially correlated baselines and also the vector of measured time delays between the reference and other correlated baselines from the cross-correlation step. The increasing length of the time series results in for ever new epoch the time delays computed with cross-correlation step to become more reliable. The TID wave parameters are estimated using the algorithms discussed in section 7.6. As mentioned before, for the wave parameters estimation, it is necessary to have at least three correlated baselines (two measured time delays). Having more correlated baselines results in more redundancy in the estimation process.

A weighted least-squares processing is carried out using the weight matrix defined by $W = Q_\tau^{-1}$ with Q_τ indicates the *dispersion* matrix of the measured time delays. The (time delay) measurements are weighted using the global maximum correlation provided from the cross-correlation function that leads to the Weighted Least Squares Estimator (WLSE) of the TID parameters as a Linear Unbiased Estimator (LUE).

7.7.5 TID ending

TID have only a limited life time due to damping of the amplitude of the TID during propagation. Therefore, it is necessary compute the ending time of the detected TID. In this subsection, the goal is to develop an algorithm to find the ending time of TID. We assume that the detected TID decays first in the reference baseline therefore, only data from the reference baseline is used for computing the ending time of TID.

In order to develop an algorithm to detect the ending of TID, we again take advantage of

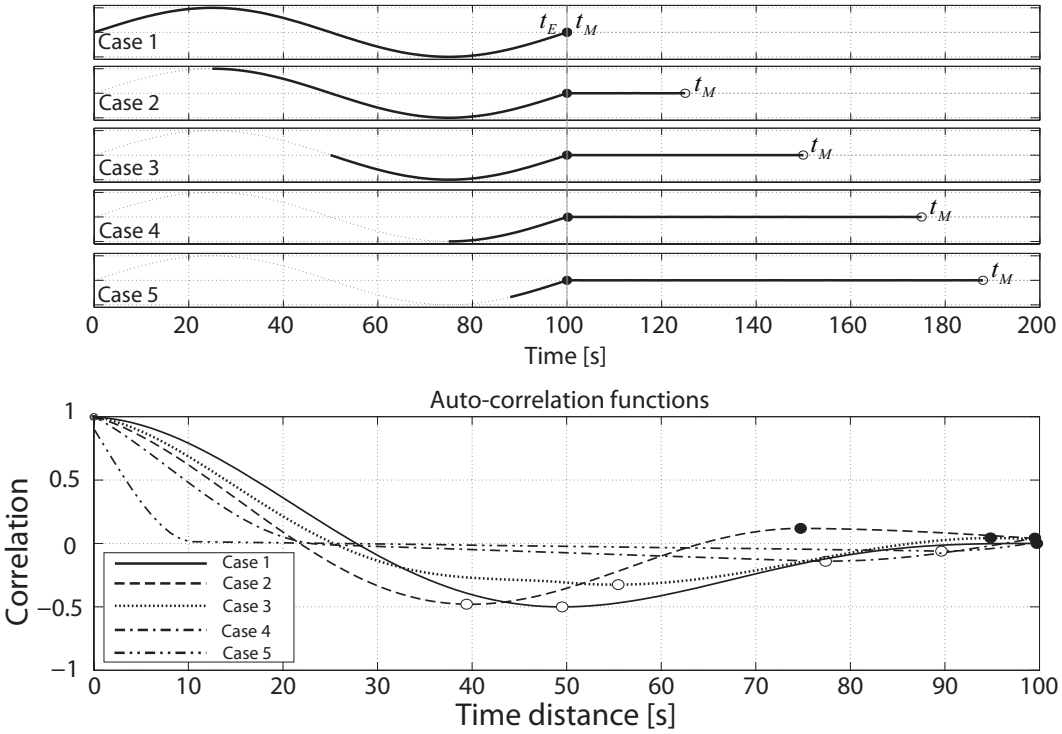

Figure 7.13: Time series includes only a portion of cycle of a sine wave for 5 cases (top) and the corresponding normalized unbiased auto-correlation functions (bottom), in these functions 'solid circle' and 'circle' indicating respectively the global maximum and minimum points.

a simulation. Let consider a simple traveling sine wave in equation 7.48 that is propagated from $t_{_S}$ to $t_{_E}$ (unknown ending time) and detected at $t_{_D} \leq t_{_E}$. The wave equation for the reference point $x = 0$ from $t_{_M} - T$ to $t_{_M} \in [t_{_E}\ ,\ t_{_E} + T]$ (subscript $_M$ indicates stopping time of the TID modeling) is as follows

$$y_0(t) = \begin{cases} \sin(2\pi f t) & t \leq t_{_E} \\ \\ 0 & t_{_E} \leq t \end{cases} \tag{7.50}$$

Figure 7.13 (top) shows the graphs of equation 7.50 for 5 cases with different values for $t_{_M}$. The cases are for $t_{_M} = t_{_E}\ ,\ t_{_E} + \frac{T}{4}\ ,\ t_{_E} + \frac{T}{2}\ ,\ t_{_E} + \frac{3\ T}{4}\ ,\ t_{_E} + \frac{8\ T}{7}$ where $T = \frac{1}{f}$ stands for the period. In the bottom of figure 7.13, the corresponding auto-correlation functions are plotted. The global maximum and minimum points are indicated by a 'solid circle' and 'circle', respectively. It is clearly seen from the patterns of the auto-correlation functions that, for $\frac{T}{2} \geq t_{_M} - t_{_E}$, an increase in $t_{_M} - t_{_E}$ leads to not only decreasing in the time distance between the global minimum and maximum points but also the minimum correlation tends to zero value. These two geometrical variations in the patterns is continued till when that the counterpart of the sine wave in the series becomes almost shorter than $t_{_M} - t_{_E} \geq \frac{7\ T}{8}$. For the cases that $t_{_M} - t_{_E} \geq \frac{7\ T}{8}$, the pattern of the auto-correlation function is totally governed by the counterpart of the series from $t_{_E}$ to $t_{_M}$.

The algorithm for ending detected TID is based on the above mentioned two geometrical variations in the pattern of the auto-correlation function. In practice, to specify the TID

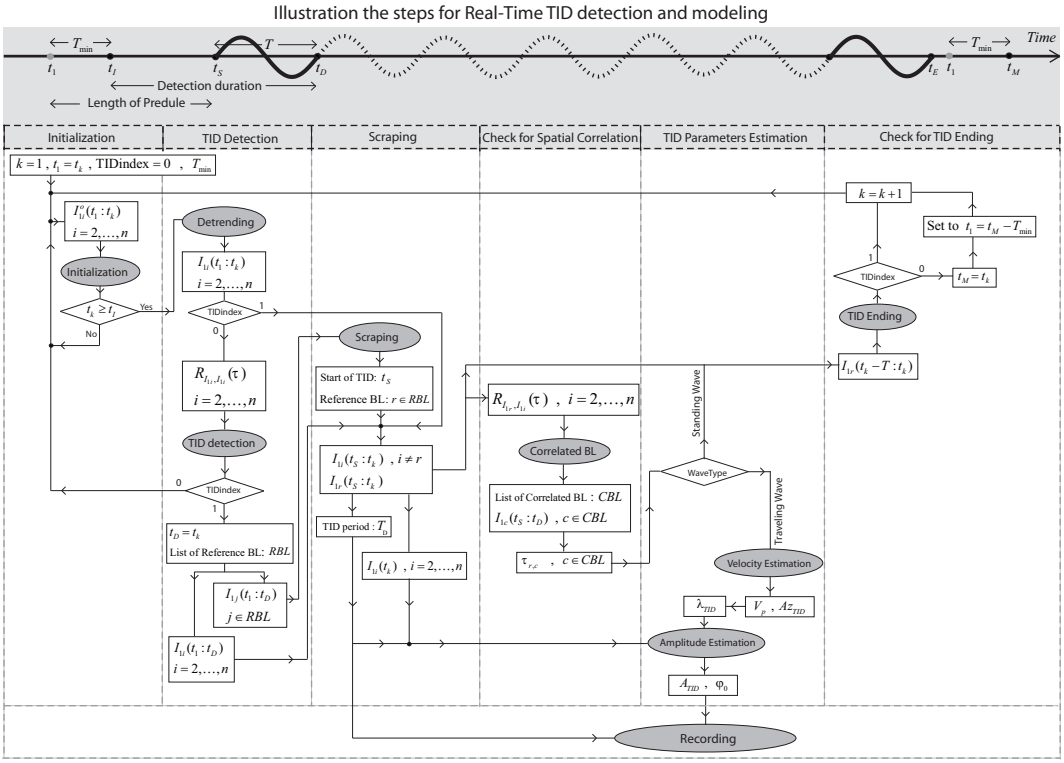

Figure 7.14: Schematic illustration of the algorithm for Real-Time Medium-scale TID modeling.

ending time, one can apply the following conditions

$$\tau_{Gmax} - \tau_{Gmin} \leq \frac{T}{4} \tag{7.51}$$

$$|R(\tau_{Gmin})| \leq \epsilon \tag{7.52}$$

where R stands for the normalized and unbiased auto-correlation function and ϵ is an arbitrary value of the smallest correlation for TID termination. According to our experience, a good value is $\epsilon = 0.2$. These two conditions should be applied epoch by epoch for the pattern of the auto-correlation function of a provided time series with length of T_{TID} (period of detected TID) from t_k to $t_k - T_{TID}$ in the data of the reference baseline (t_k stands for current epoch). Once these condition are fulfilled in the pattern of the auto-correlation function, the time for stopping TID modeling is $t_{_M} = t_k$.

After the ending of a TID has been detected the algorithm will start again looking for new TID using the detection algorithm. The reference epoch for the detection algorithm is set to $t_{_1} = t_{_M} - \ 10$ minutes.

7.7.6 Flowchart of the Real-Time TID modeling algorithm

The flowchart for the algorithm for real-time MS-TID modeling is schematically depicted in figure 7.14. In this figure, the time series of single-differenced geometry-free GPS observations from initial epoch t_1 to current epoch t_k for n baselines in a network are denoted as ${I_{1i}}^o(t_1 : t_k)$ with $i = 2, \ldots, n$ (first receiver as a reference). The superscript o is

removed when the time series are detrended. $TIDindex$ is set to 1 when a TID is detected otherwise to 0. At the starting time t_1, $TIDindex = 0$. Duration of the initialization step is $t_I - t_1 = T_{min}$ where $T_{min} = 10$ minutes is the shortest period of MS-TID. When the length of the time series exceeds T_{min} (i.e. $t_I \leq t_k$), first the detrended time series are provided and then the auto-correlation (denoted by $R(\tau)$) for each time series are computed for the TID detection process. Note that the detrending process is carried out every epoch.

In case a MS-TID is detected, the list of baselines which are candidate reference baselines, is computed (denoted by RBL). The epoch that a TID detected is t_D, $t_D - t_1$ is the *duration of detection*. In the scraping process, the best reference baseline and start time t_S of the detected TID are computed as well as the Dopplered-period of the TID, denoted by T_D. The spatial correlation between the reference baseline and the others baselines is checked to provide a list of the correlated baselines (denoted by CBL). Then, by excluding the uncorrelated baselines, the time delays of the correlated baselines with respect to the reference baseline are computed by cross-correlation. In case there are more than 3 correlated baselines the measured time delays are used to estimate the wave parameters of the TID. In case that all of the time delays are zero, the TID wave will be a standing wave and the estimation step is stopped. The check for TID ending is carried out to compute the end time of TID in the reference baseline (denoted by t_E) and to specify the time for stopping estimation of the TID parameters (denoted by t_M). It should be noted that, the check for TID ending is done only for time series of the reference baseline with length equal to T_D. In case that the detected TID is ended then $TIDindex$ is set to 0 and the initial epoch t_1 is updated for detecting of the next MS-TID.

7.7.7 Dependency on reference baseline

The algorithm for TID modeling utilizes the cross-correlation function for the estimation of the propagation parameters. In this subsection, it is shown that the least squares estimates of the propagation parameters is not always *invariant* against the selected reference point (or baseline).

Let consider four IP points in an ionospheric patch with a planar TID wave (see figure 7.15). The amplitude of the TID is depends on location. The wave equation at the points $(i = 1, \ldots, 4)$ is

$$y_i(t) = A(t, \vec{r}_i) \sin\left(\omega t - \varphi(t, \vec{r}_i)\right) \tag{7.53}$$

The single-difference wave equation with respect to an arbitrary reference point can be written as

$$y_{ji}(t) = y_i(t) - y_j(t) = A(t, \vec{l}_{ji}) \cos\Big(\omega t - \varphi(t, \vec{l}_{ji})\Big) \tag{7.54}$$

with subscript $j \in [1, \ldots, 4]$ for the reference point and $\vec{l}_{ji}$ the baseline vector between reference point and other points $(i = 1, \ldots, 4)$. The cross-correlation function of $y_{ji}(t)$ and $y_{jk}(t)$ can be expressed as

$$R_{y_{ji}, y_{jk}}(\tau) = \frac{A(t, \vec{l}_{ji}) A(t, \vec{l}_{jk})}{2} \cos\Big(\omega\tau + \Delta\varphi(t, \vec{l}_{ji}, \vec{l}_{jk})\Big) \tag{7.55}$$

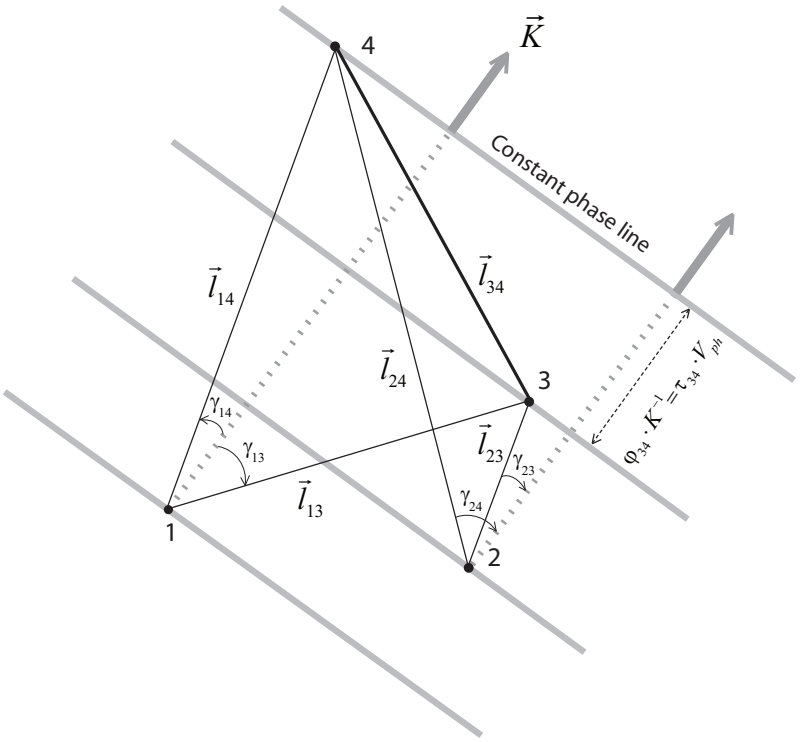

Figure 7.15: Illustration of the phase-offset between two IP points measured from two different reference points.

with $\Delta\varphi(t, \vec{l}_{ji}, \vec{l}_{jk}) = \varphi(t, \vec{l}_{ji}) - \varphi(t, \vec{l}_{jk})$ the initial phase of the cross-correlation pattern. Since the amplitude and the initial phase do not depend on time delay τ, equation 7.55 can be written as follows with $\tau = 0$

$$R_{y_{ji},y_{jk}}(0) = \frac{A(t, \vec{l}_{ji})A(t, \vec{l}_{jk})}{2} \cos\left(\Delta\varphi(t, \vec{l}_{ji}, \vec{l}_{jk})\right) \tag{7.56}$$

It should be noted that the time delay between two given points i and k, measured by cross-correlation, is related to the initial phase by $\tau_{ik} = \Delta\varphi \cdot V_{ph}$, where V_{ph} denotes the phase velocity of the wave. Using propagation law for the cross-covariance function, it can be shown that

$$R_{y_{ji},y_{jk}}(0) = R_{y_{ji},y_{ji}}(0) - R_{y_{ij},y_{ik}}(0) \tag{7.57}$$

where $R_{y_{ji},y_{ji}}(0) = \frac{A^2(t,\vec{l}_{ji})}{2}$ stands for the auto-correlation function of y_{ji} (at $\tau = 0$). Substituting equation 7.57 into 7.56 yields

$$\frac{A^2(t, \vec{l}_{ji})}{2} - \frac{A(t, \vec{l}_{ij})A(t, \vec{l}_{ik})}{2} \cos\left(\Delta\varphi(t, \vec{l}_{ij}, \vec{l}_{ik})\right) = \frac{A(t, \vec{l}_{ji})A(t, \vec{l}_{jk})}{2} \cos\left(\Delta\varphi(t, \vec{l}_{ji}, \vec{l}_{jk})\right) \tag{7.58}$$

In order to have the same time delays between two points 3 and 4 in figure 7.15, measured using two different reference points (e.g. $j = 1$ and $j = 2$), it is necessary to have the

following equality fulfilled

$$\frac{A^2(t,\vec{l}_{13}) - A(t,\vec{l}_{31})A(t,\vec{l}_{34})\cos\Big(\Delta\varphi(t,\vec{l}_{31},\vec{l}_{34})\Big)}{A^2(t,\vec{l}_{23}) - A(t,\vec{l}_{32})A(t,\vec{l}_{34})\cos\Big(\Delta\varphi(t,\vec{l}_{32},\vec{l}_{34})\Big)} = \frac{A(t,\vec{l}_{13})A(t,\vec{l}_{14})}{A(t,\vec{l}_{23})A(t,\vec{l}_{24})} \tag{7.59}$$

It is only guarantied to have equation 7.59 fulfilled if first the amplitude of wave (i.e. $A(t,\vec{r}_i) = A(t)$) is not spatially dependent and secondly the reference points (e.g. $j = 1$ and $j = 2$) are in a line perpendicular to the propagation direction of the wave.

Because in the practice, the amplitude of TID is not spatially independent it follows that the measured time delay from cross-correlation is not always invariant against the reference baseline. Therefore, TID modeling based on the cross-correlation is the reference point dependent. It implies that it is necessary either to develop a strategy for selection of the best reference point or to overcome this problem all possible baselines in the network at same time is used for estimation of the wave propagation parameters which gives an averaged solution.

7.7.8 Sensitivity to temporal resolution

The propagation parameters are estimated using time delay measurements from cross-correlation. The delay measurements are sensitive to the temporal resolution of the time series. Let consider three IP points (denoted by i, j, and k) in an ionospheric patch over which a TID moving. The time resolution of TID observations (sampling rate) is Δt then the time delay measured from cross-correlation between time series of two baselines $\vec{l}_{ij}$ and $\vec{l}_{ik}$ will be zero in the following three cases: (1) $\vec{l}_{jk} \perp \vec{V}_{ph}$ ($\vec{V}_{ph}$ stands for phase velocity) (2) $\vec{V}_{ph}$ tends to infinity (Standing wave) (3) Time delay is smaller than the time resolution $\tau_{_{jk}} \leq \Delta t$. The first case depends on the propagation direction of the TID with respect to the baselines and the second case depends on the velocity of the TID.

The last case is the case that is relevant to the time resolution of the data. Because the resolution of the cross-correlation is the same as the time series, $\tau_{_{jk}}$ will be obtained zero if $\tau_{_{jk}} \leq \Delta t$. This occurs in case that the traveling time of TID between the points of the network is smaller than the time resolution of data. Therefore when a network with short baselines is used for the TID analysis it needs to have high temporal resolution data.

7.8 Implementation of the Real-Time TID modeling

In this section the performance of the real-time MS-TID modeling is demonstrated for two test cases. A network including 7 stations of the California SCIGN network simulates the permanent GPS network. Station SPK1 is selected as a reference to compute single-differenced ionospheric delays. Four other SCIGN stations inside the reference network and four stations outside the assigned network are selected for testing purposes. In figure 7.16 the configuration of all 15 stations is depicted. In this figure, the solid lines show baselines in the assigned permanent network which are used for the TID parameter estimation. The dashed lines are the baselines that are used to test the TID model. The baseline lengths

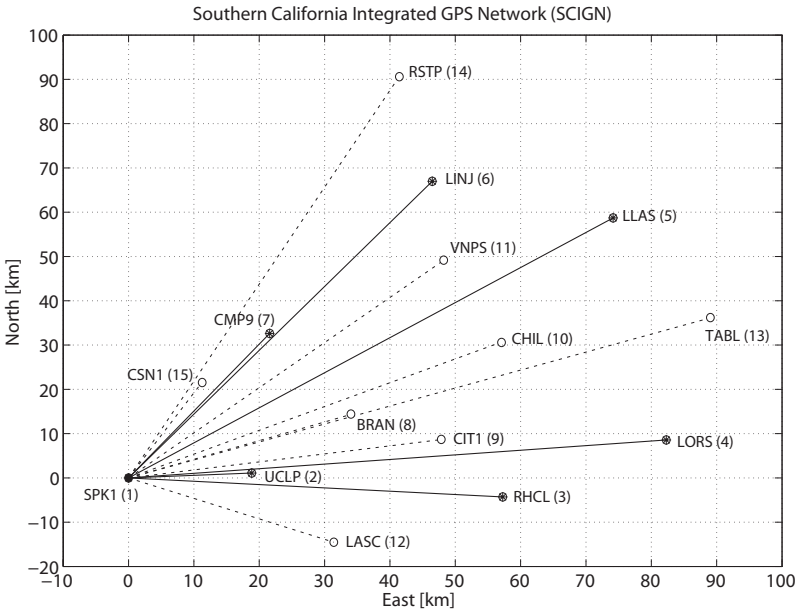

Figure 7.16: Selected 15 stations from SCIGN Network.

are less than 100 km. The shortest baseline length in the assigned permanent network is 20 km (for SPK1-UCLP) and the longest one is 95 km (for SPK1-LLAS).

Dual-frequency GPS phase and code data in July 07, 2000 from $06{:}00\ h$ to $11{:}00\ h\ \ UTC$ were used during which a strong TID occurred. The corresponding local time in California was from $23{:}00\ h$ on July 06, to $04{:}00\ h$ on July 07. The goal is here to detect and to model the TIDs using the algorithm described in the previous sections. The TID will be analyzed using the network GPS data for two satellites, PRN 02 and PRN 08.

The locations of ionospheric pierce points of the reference network with respect to two satellites PRN 02 and PRN 08 during the 5 h timespan is shown in figure 7.17. The ionospheric elevation angles and velocity of the IP points (average over the network) are depicted in figure 7.18. Note that the coordinates and the velocity of the IP points depend on the hight of the thin-layer used to model the TID. In the literatures different values are used, e.g. 200 km in (Hernandez-Pajares et al., 2006a) and 300 km in (Afraimovich et al., 1998) and 400 km in (Garrison et al., 2007). It has been shown that the estimated TID wave parameters are not very sensitive to the ionospheric height (Garrison et al., 2007). We assumed a height of 350 km for the thin-layer ionosphere in this chapter.

Based on the sampling interval of 30 seconds, a total of $5 \times 120 = 600$ epochs the single-difference TID observations were computed using the procedure described in subsection 7.4.3. To facilitate the drawing of these graphs, detrending was not made epoch by epoch but only once over the which time span (in a batch using post-processing). In order to visualize the TID, the time series of the single-difference TID observations of the all 14 baselines of the assigned network are shown in figure 7.19 for PRN 02 (top) and PRN 08 (bottom). In this figure, bold curves are for the permanent network baselines and thin curves and dashed curves are respectively for baselines with second receivers inside and outside the network (distance and azimuth between two ends IP points of the baselines are

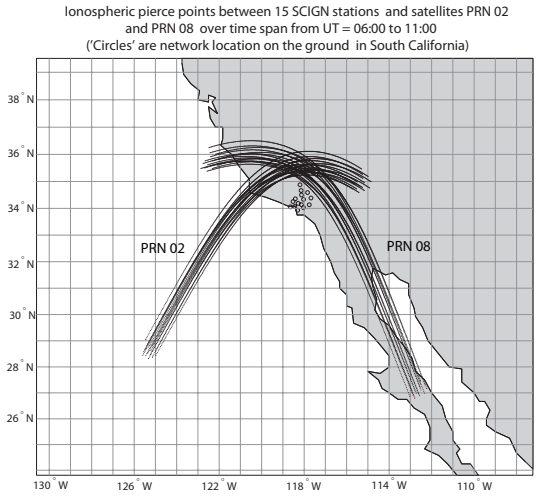

Figure 7.17: Ionospheric pierce points (IP) for satellites PRN 02 and PRN 08 from UT=06:00 to 11:00 in 2000.07.07.

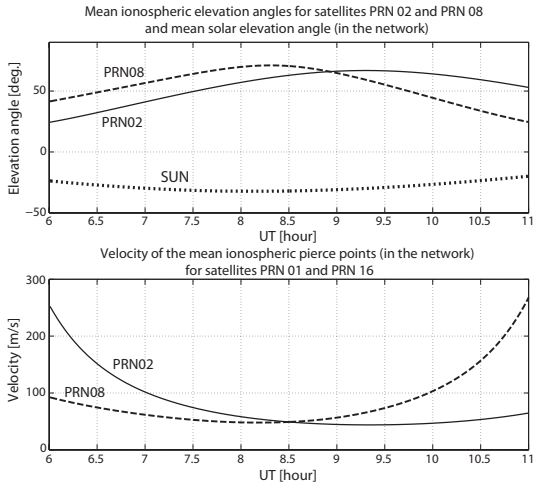

Figure 7.18: Ionospheric elevation angles for satellites PRN 02 and PRN 08 from UT=06:00 to 11:00 in 2000.07.07 (top), and velocity of the ionospheric pierce points (bottom).

given in the brackets). The signature of the TID can be clearly seen in the graphs for the both satellites. One can see that magnitude of the TID depends on baseline length and azimuth.

In the following subsections the TID wave parameters are estimated epoch by epoch for both satellites.

7.8.1 Case study: PRN 02

The first row of figure 7.20, gives the time series of single-difference TID observations of the 6 baselines in the assigned permanent network for satellite PRN02 (600 epochs). The TID detection algorithm was applied to these time series and several medium scale TIDs were detected. The strongest MS-TID was detected at epoch 145 in the time series of the baseline SPK1-LINJ. In the 2nd row of figure 7.20, the second receiver numbers of the baselines associated to each of the detected TIDs are shown, the doted line is the combined initialization and detection period, the solid line the period from detection time (t_D) to end of modeling (t_M).

The wave parameters for a TID were estimated epoch-by-epoch over interval t_D to t_M. The period of a TID wave was computed using the algorithm presented in subsection 7.6.1. The computed Dopplered-period (T_D) and actual period (T_{TID}) are denoted respectively by 'circles' and 'dots' in the 3rd row of figure 7.20. The gray parts in the 3rd row refer to epochs during which either the TID behaves as a standing wave or the estimated phase velocity of TID tends to infinity. The tendency to infinity is for epochs in which all the measured time delays in the network, from cross correlation, were smaller than the temporal resolution of data. As a consequence, it leads to zero values for the time delays. For those epochs which TID is as a standing wave or the phase velocity tends to infinity, only the Dopplered-period of TID can be measured and the other wave parameters are not estimated. The reason for some discontinuity in the measured Dopplered-periods (accordingly in the actual period) is

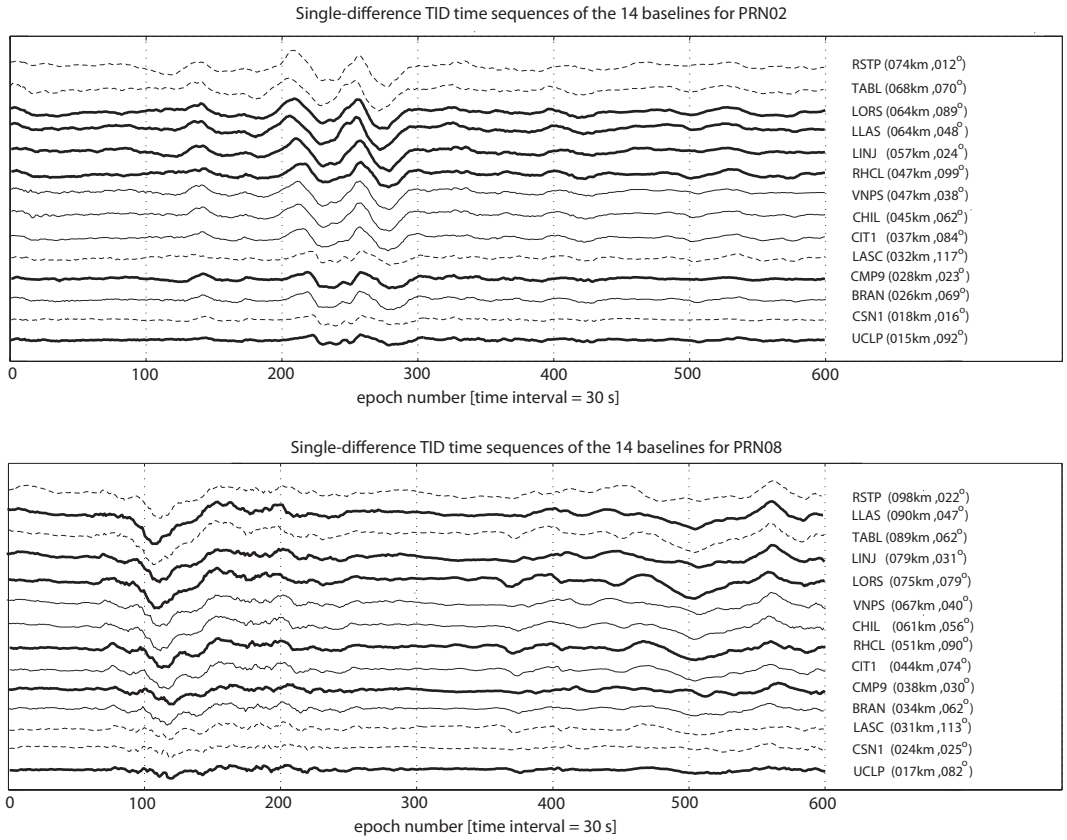

Figure 7.19: Time series of the single-difference TID observations for the 14 baselines in the SCIGN network, for PRN 02 (top) and for PRN 08 (bottom), bold curves are for the network baselines and thin and dashed curves are for baselines with second receivers, respectively, inside and outside the network (baseline length and azimuth are given in the brackets).

because of the fact that the Dopplered-period is computed from the distance between the first and second local maximum points of auto-correlation function, and because of noise in data, the second local maximum is sometimes replaced by another local maximum.

The 4th row in figure 7.20 gives number of the spatially correlated baselines for each epoch. The phase velocity and azimuth of the TID can be estimated if at least 2 baselines are correlated with the reference baseline. In the 5th row of figure 7.20, the estimated azimuth of detected TIDs (dots) and azimuth of the PRN 02 (red color solid curve) are depicted. In the 6th row of figure 7.20, the estimated phase velocity (dots) and the (averaged) velocity of the network IP points (red color solid curve) are shown. The strongest detected TID propagated in the south-west direction with phase velocity about $280\ [m/s]$. The 7th row of figure 7.20 gives the computed wavelengths corresponding to the estimated phase velocity for each detected TID. In the 8th and 9th rows of the figure 7.20, respectively, the estimated amplitude and initial phase for each TID are depicted. The estimated initial phase shows almost a linear behavior with time particularly for period which the estimated amplitude does not change considerably in time. It is because of the estimating of the amplitude and initial phase using a single epoch data that results in the estimated initial phase affected by the phase difference between two consecutive epochs (i.e. $2\pi f \Delta t$).

Despite detection of a few short-life TIDs in this case study, it can be seen from figure

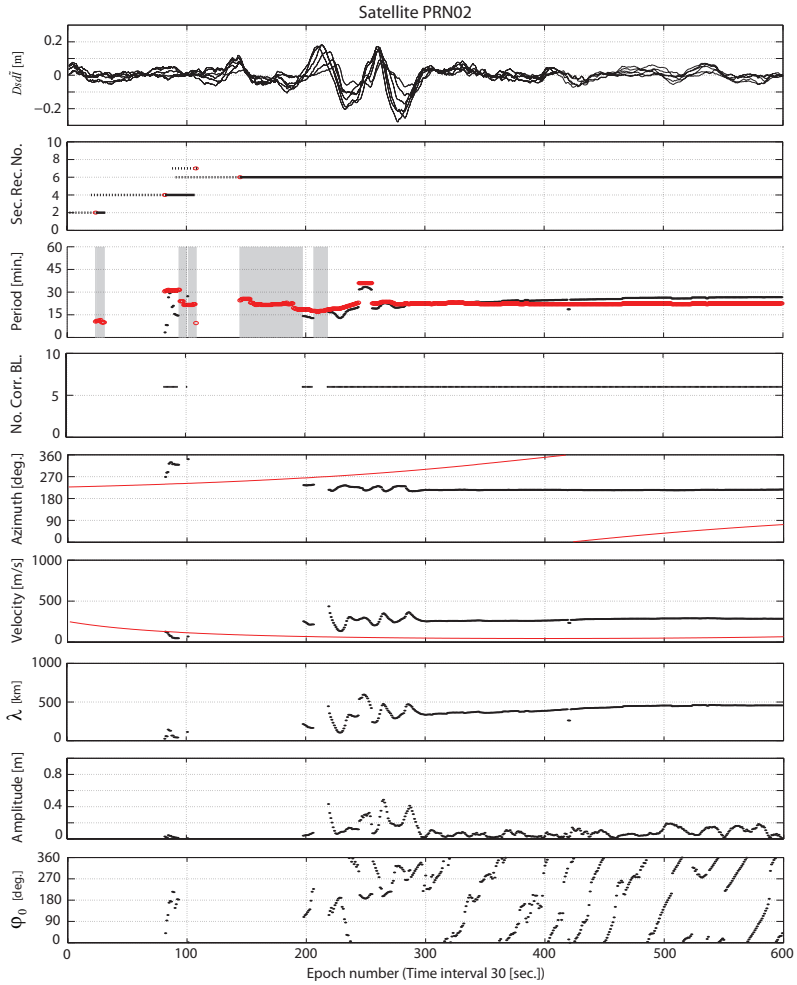

Figure 7.20: Estimated TID parameters (epoch-by-epoch) using the single-difference TID observations of the 6 baselines in the assigned network for satellite PRN 02 from UT=06:00 to 11:00 (600 epochs) on July 07, 2000 in the first row; second receiver number of the selected reference baselines in the 2nd row; Dopplered-period (circles) and actual TID period (dots) in the 3rd row; number of the spatially correlated baselines through TID in the 4th row; azimuth of TID (dots) and azimuth of PRN02 (red color solid curve) in the 5th row; estimated phase-velocity of TID (dots) and velocity of the network IP points (red color solid curve) in the 6th row; estimated wavelength and amplitude and initial-phase of TID respectively in the 7, 8 and 9th rows.

7.20 that the detection of TID and estimation of the parameters were successfully carried out epoch-by-epoch by the proposed algorithm. This approves preliminary the performance of the algorithm. Now, we should check how good the TID effects were modeled by the monochromatic traveling plane wave.

In order to assess the goodness of fit of the TID model, the single-difference TID values were computed from the model using the estimated parameters for the 6 baselines in the assigned permanent network. In figure 7.21, the observed time series (dotted curve) and the computed values (blue color solid curves) are shown for each baseline. The red color curves show the residuals (observed minus computed). It is clear that the computed TID values fit well to the TID observations which results in small residuals (comparing to the observed TID time series). But there is a phase-difference between the observed and the

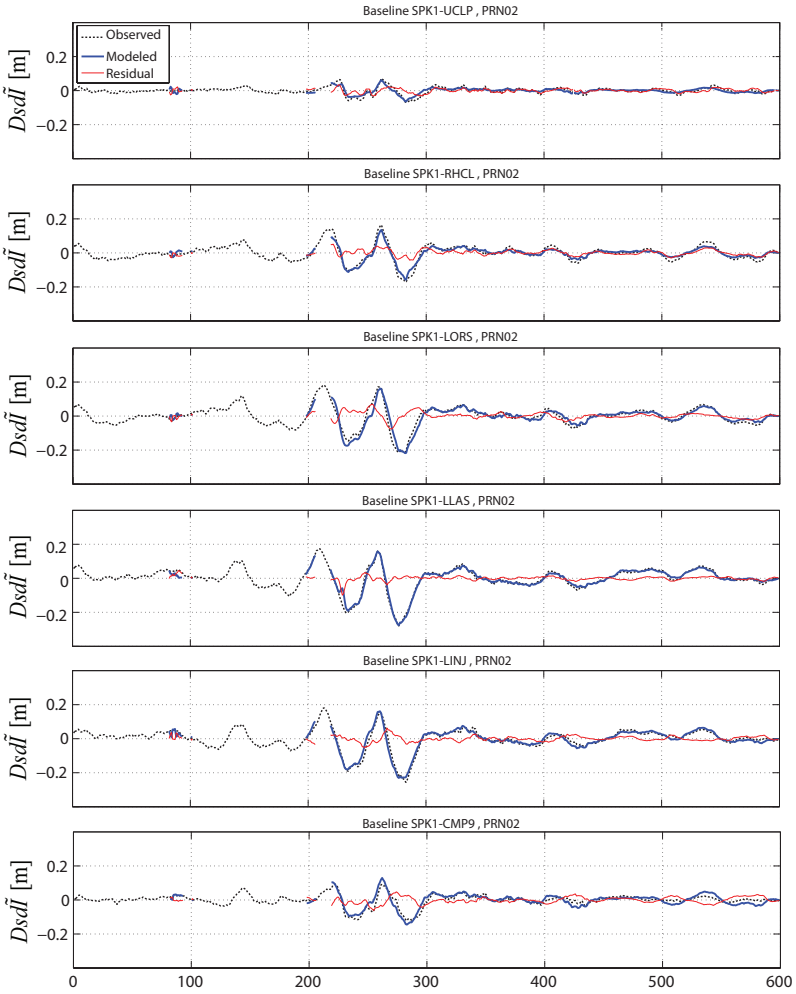

Figure 7.21: TID modeling residuals for the 6 baselines in the assigned permanent network for PRN02.

computed time series. This is due to the fact that the planar wave approximation was used for estimating the phase velocity of the TID which leads to an averaged phase velocity being estimated. Because of this, in a given epoch, the phase-differences are different for the different baselines. The large phase-difference is for an epoch that the TID wavefront considerably deviates from the planar wave.

To test the performance of the TID model for spatial prediction, the single-difference TID values were computed from the model for the other 8 baselines ('dashed' lines in figure 7.16). Note that four stations (BRAN, CIT1, CHIL and VNPS) are inside the assigned permanent network and four stations (LASC, LLAS, CMP9 and CSN1) are outside the network. The observed and computed single-difference TID time series are depicted in figure 7.22. The red color curves show the residuals (observed minus computed). The four rows on the top are for baselines inside the network and the four rows on the bottom are for the baselines outside the network. For comparison, the Kriging interpolation method was used to compute interpolated TID values for the four stations inside the network. Kriging is a linear spatial interpolation method that fully takes the spatial correlation between given points into account. In figure 7.22, the green curves show the residual from the Kriging interpolation.

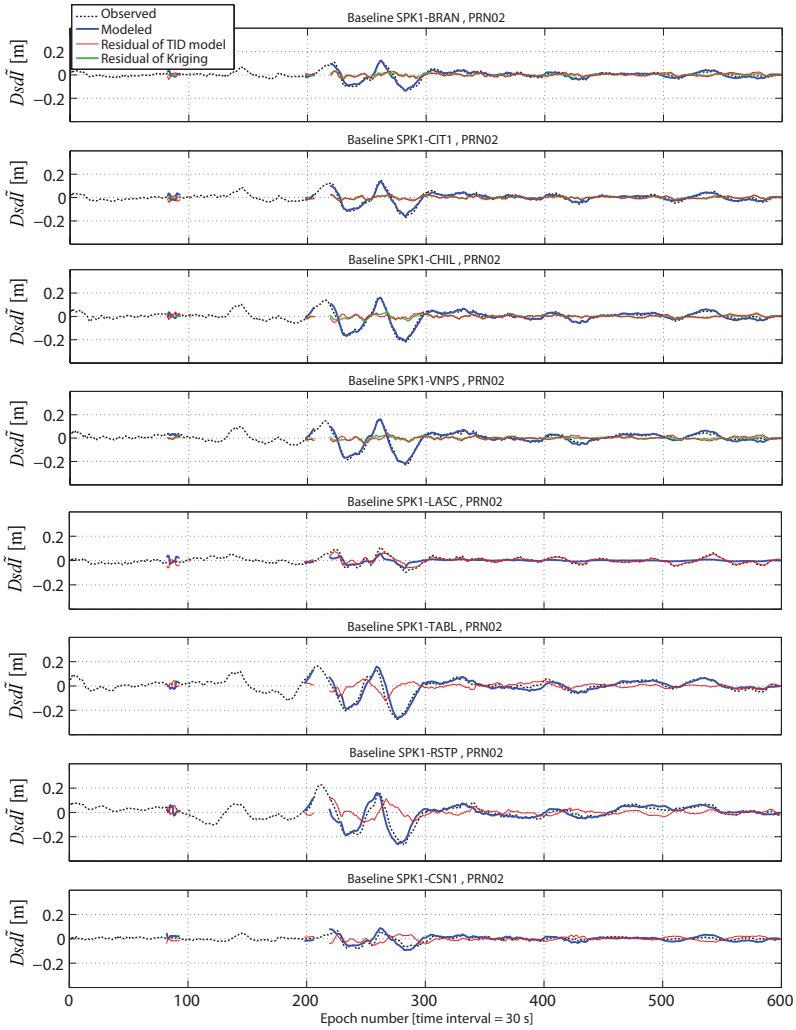

Figure 7.22: Spatial prediction of TID and the residuals, four rows on the top associated for spatial interpolation and four rows on the bottom for spatial extrapolation (PRN02).

From figure 7.22 it follows that the algorithm could spatially predict the TID fluctuations considerably in the assigned network. The interpolation was done with better accuracy than for the extrapolation. For comparison the performance of the algorithm for interpolation and for the extrapolation, maximum and Root-Mean-Square (RMS) values of the time series and the residuals for all the baselines are shown in figure 7.23. The maximum and RMS values are given for the time series of TID observations and for the residual from the TID modeling and from the Kriging. Note that the residual from Kriging is given only in case of interpolation. The maximum and RMS values for each time series was computed only over the epochs that the residuals were computed. From this figure it is clear that the maximum and the RMS values of the residuals obtained from the TID modeling are always significantly smaller than those for the TID observations. It implies that the algorithm could considerably model TID effects in the data. As it was expected, the RMS values of the residuals for the interpolation are smaller than those for the extrapolation.

Another result of figure 7.23 is that the RMS values of the residuals from the TID modeling and from Kriging are comparable. That is an important result because these two methods

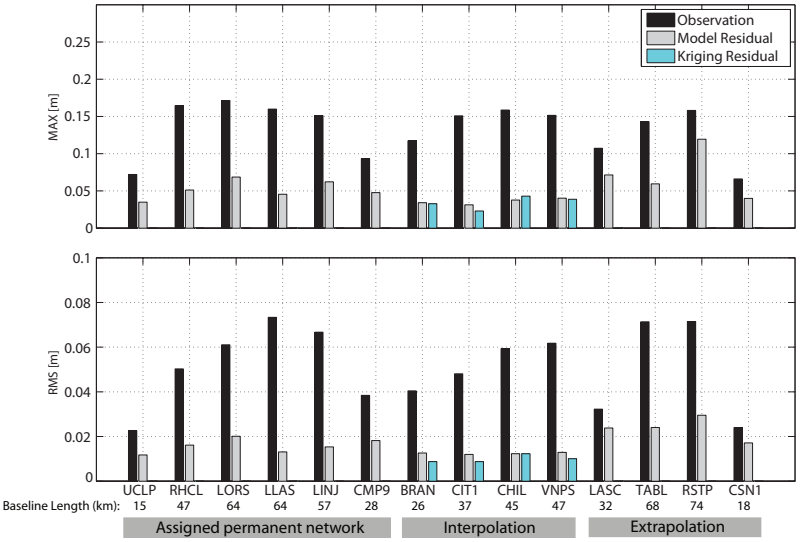

Figure 7.23: Maximum values of the single-difference TID observations for satellite PRN 02, with maximum values of the corresponding residual from Kriging and the model from the section 7.7 (top). Root-Mean-Squared (RMS) values for the same cases (bottom).

are developed using totally different principals, but they give similar result. The Kriging is basically a linear spatial interpolation method that takes spatial correlations into account. Our algorithm has been developed based on the physics of ionosphere using the planar wave approximation for the TID wave.

An advantage of the our algorithm is that there is a possibility to improve the algorithm for the phase-offset modeling by avoiding the planar wave approximation. While there is no any way to improve the Kriging for the phase-offset modeling between the IP points.

7.8.2 Case study: PRN 08

In this subsection another example of the epoch-by-epoch estimation of the wave parameters associated to the TIDs propagating over the network IP points of satellite PRN 08 is given. As in the previous example, the TID detection algorithm was applied for the 6 time series of single-difference TID observations in the assigned permanent network in the 5 h time span for satellite PRN 08. In this case study, the strongest TID was detected at epoch number 107 with phase velocity about $150\ [m/s]$ toward almost West direction in the baseline SPK1-RHCL.

The estimated parameters (epoch-by-epoch) of the TID waves are shown in figure 7.24. In this figure, the first row gives the time series of single-difference TID observations of the 6 baselines in the assigned permanent network for satellite PRN08 (600 epochs) and the 2nd row shows second receiver number of the reference baseline for each TID. The measured Dopllered-period (T_D) and actual period (T_{TID}) of the TIDs are depicted respectively by 'circles' and 'dots' in the 3th row. The gray parts in the 3rd row relates to the epochs that either the TID behaved as a standing wave or the estimated phase velocity tended to infinity. The 4th row shows the number of spatially correlated baselines for each epoch. The estimated azimuth of TIDs (dots) and azimuth of the PRN 08 (solid curve) are shown

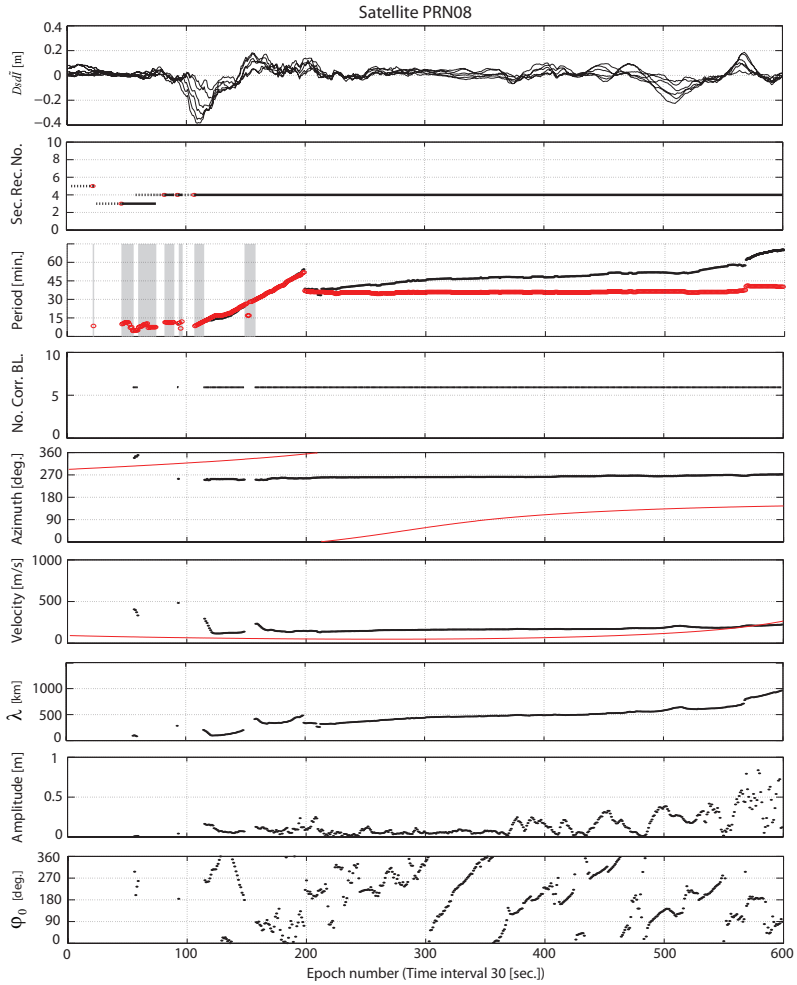

Figure 7.24: Estimated TID parameters (epoch-by-epoch) using the single-difference TID observations of the 6 baselines in the assigned network for satellite PRN 08 from UT=06:00 to 11:00 (600 epochs) on July 07, 2000 in the first row; second receiver number of the selected reference baselines in the 2nd row; Dopplered-period (circles) and actual TID period (dots) in the 3rd row; number of the spatially correlated baselines through TID in the 4th row; azimuth of TID (dots) and azimuth of PRN 08 (red color solid curve) in the 5th row; estimated phase-velocity of TID (dots) and velocity of the network IP points (red color solid curve) in the 6th row; estimated wavelength and amplitude and initial-phase of TID respectively in the 7, 8 and 9th rows.

in the 5th row. In the 6th row, the estimated phase velocity associated to the TIDs (dots) and the (averaged) velocity of the network IP points (solid curve) are given. The wavelength of the detected TIDs are shown in the 7th row and the estimated amplitude and initial phase of TIDs are depicted respectively in the 8th and 9th rows.

Although the TID effects, in this case study, were more complicated (less similar to periodic fluctuations) it can be seen from figure 7.24 that the algorithm was able to detect TIDs and estimate epoch-by-epoch the parameters. To show the goodness of fitting of the TID model, the time series of single-difference TID observations for 6 baselines in the assigned permanent network and the corresponding computed time series are given in figure 7.25. From this figure follows that the computed TID are fitted well to the TID observations, which results in small residuals. As in the previous case study, it is clear that there is a

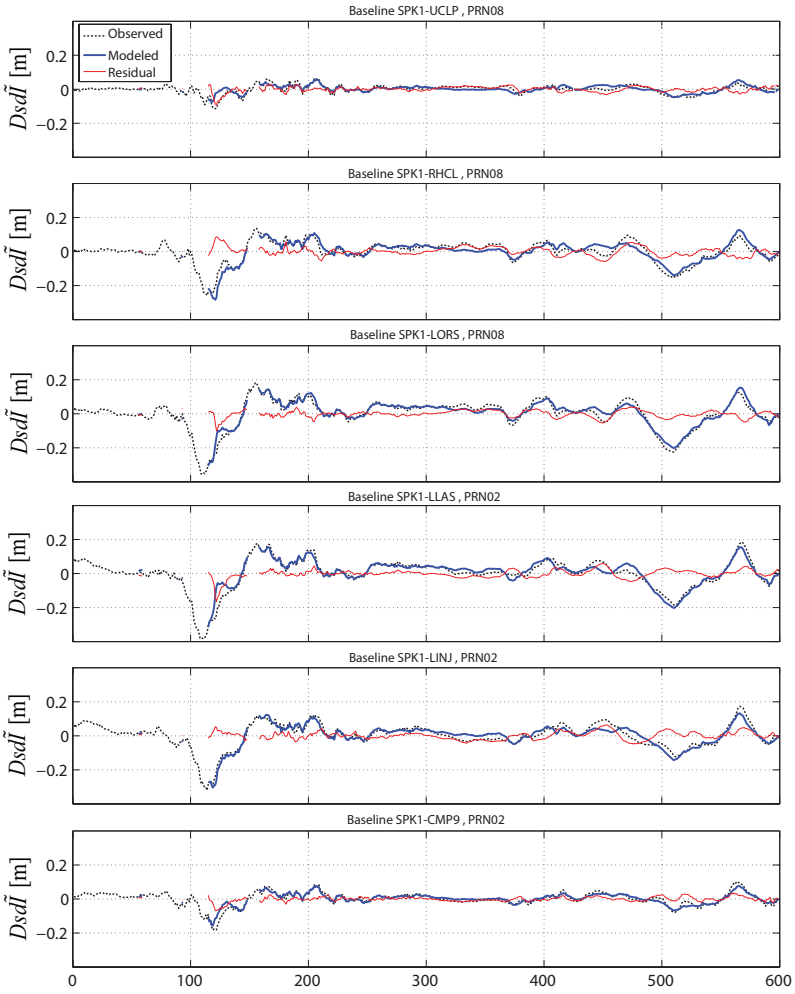

Figure 7.25: TID modeling residuals for the 6 baselines in the assigned permanent network for PRN08.

phase-difference between the observed and modeled time series that was caused by the planar wave approximation.

The modeled TID values are computed for the other 8 baselines, to test the performance of the algorithm for spatial prediction in the ionosphere. The observed and the computed time series of single-difference TID observations are shown in figure 7.22. The four sub-figures on the top are for baselines inside the network and the four sub-figures on the bottom are for the baselines outside the network. The time series of residuals provided from the TID modeling and from the Kriging interpolation are depicted for the 4 baselines inside the network. To compare the residuals, the maximum and RMS values of the residuals for all baselines are given in figure 7.27. It follows from this figure that the maximum and the RMS values of the residuals are always significantly smaller than those for the TID observations. In case of interpolation the RMS values are smaller than those for the case of extrapolation. It is again clear that the RMS values of the residual for the TID modeling and for Kriging are comparable.

From the two case studies (PRN 02 and PRN 08) the following results can be obtained:

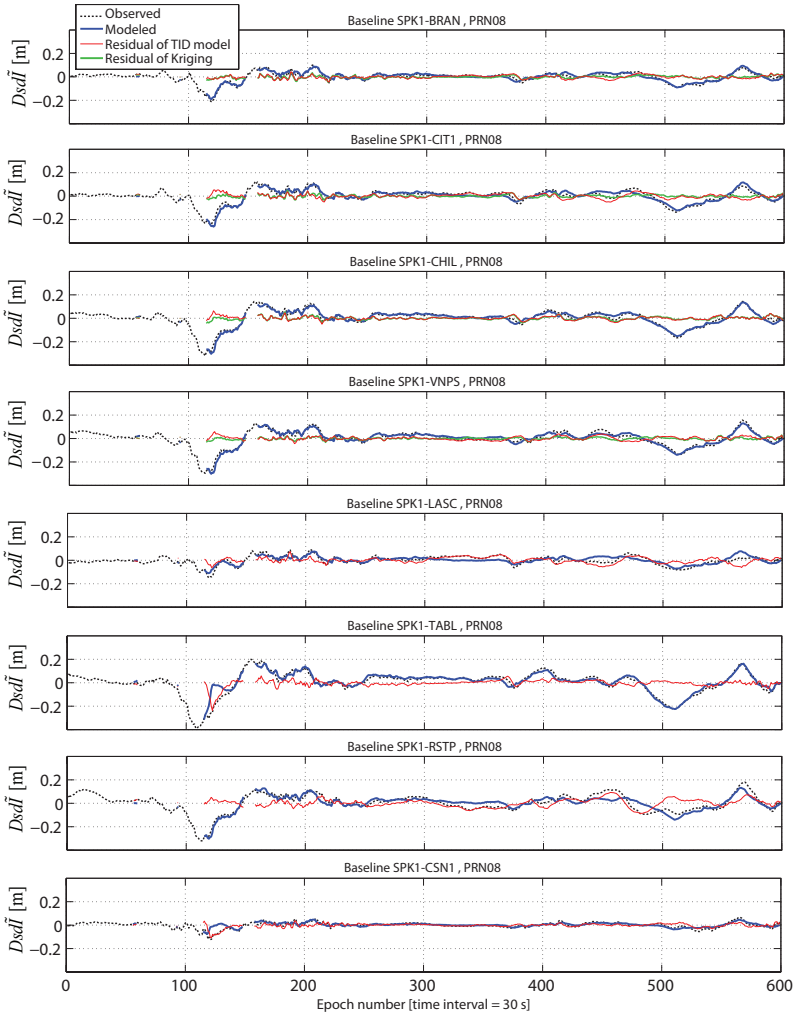

Figure 7.26: Spatial prediction of TID and the residuals, four rows on the top associated for spatial interpolation and four rows on the bottom for spatial extrapolation (PRN08).

- The TID effects on the GPS provided ionospheric time series don't behaves as a perfect periodic fluctuation. Due to the Doppler-effect caused by satellite motion, in special situation, the TID effects behave as a standing wave on data. In this situation the projected IP velocity along the phase velocity of TID is very close to the phase velocity but with opposite sign. As a consequence, the phase-offset between the time series for different ionospheric points becomes zero.

- Despite the TID effects on the time series were not a perfect periodic fluctuation, the algorithm could detect successfully medium-scale TIDs and estimate epoch-by-epoch the TID parameters. Although the plane wave approximation is used for TID modeling but the algorithm could significantly model the TID effects. The residual (model error) for the interpolation was smaller than those for the extrapolation.

- The spatial prediction of TID in our algorithm and in the Kriging interpolation are different in principle but the residuals obtained from these two methods are comparable. An advantage of our algorithm is that there is possibility to improve the algorithm in a way that the phase-offset between the IP points is correctly modeled.

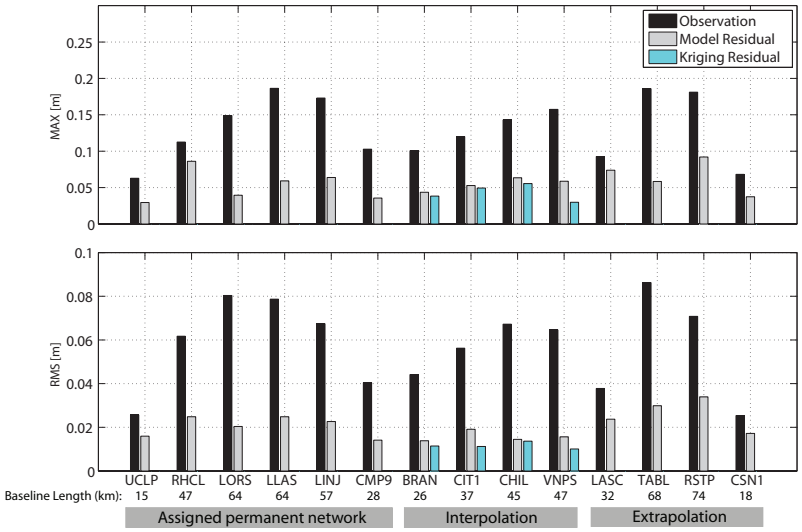

Figure 7.27: Maximum values of the single-difference TID observations for satellite PRN 08, with maximum values of the corresponding residual from Kriging and the model from the section 7.7 (top). Root-Mean-Squared (RMS) values for the same cases (bottom).

7.9 Conclusions and remarks

In this chapter an algorithm was developed for the real-time MS-TID detection and modeling in a ionospheric patch using data from a medium scale permanent GPS network. This algorithm is based on the time series of between-receiver single-difference ionospheric delays provided by the dual-frequency phase and code GPS observations collected at the GPS reference stations. The algorithm takes high temporal resolution in the GPS data into account in the cross-correlation function to model TID effects in the ionospheric data. A MS-TID wave is modeled as a planar longitudinal traveling wave with spatially independent amplitude that propagates on the ionospheric patch of the network of IP (Ionospheric Pierce) points associated to a given satellite. The characteristic parameters of a TID wave (e.g. period, phase velocity, propagation direction, amplitude), are considered to be time dependent and the Doppler shift caused by the satellite motion is taken into account.

The algorithm consists of 6 steps: Initialization, Detection, Scraping, Cross-Correlation, Parameter Estimation, and Ending. In all these steps, the detrended time series of the ionospheric observations is used. The geometry-dependent trend due to the regular ionospheric variations was removed using the physics-based TEC model developed in the previous chapter. The performance of the model was tested in a medium scale permanent GPS network over a time span of 5 hour for two satellites. The tracks of the detected TIDs show that a TID in the ionosphere is not a perfect planar wave and also its amplitude is space dependent. Despite this, the algorithm was able to model (in time and in space) the TID fluctuations well and its performance is comparable with the Kriging interpolation method. The results obtained from the two case studies show that due to the planar wave approximation in the estimation of the TID phase velocity, the algorithm is not capable to model correctly the phase-offset in the space domain. But there is a possibility to improve the algorithm in such a way that the phase-offset between the IP points can be modeled. For this purpose one needs to estimate the phase velocity of the TID as a function of the

azimuth of the baselines in the network. This is the advantage of this algorithm compared with all other ionospheric interpolation methods such as Kriging.

8 Conclusions and recommendations

The main objective of this thesis is the development of a procedure for modeling and predicting the Total Electron Content (TEC) of the Earth's ionosphere for high precision differential GNSS applications. This objective led to the development of a physics based model for the regular TEC variations and a procedure for modeling medium scale Traveling Ionospheric Disturbances (TID). The model for the regular TEC variation is a recursive model based on ionospheric physics, involving the production and recombination of ions and free electrons, that can be used in any recursive estimation procedure. The input to this recursive TEC model consists of the known solar zenith angle, and estimates of the effective solar radiation intensity and recombination coefficients for one or two ionospheric stratifications.

The research described in this thesis can roughly be divided into three parts. In the first part the performance of several empirical ionosphere models, such as the Klobuchar and NeQuick, and the IGS Global Ionosphere Maps (GIM) is studied. The main focus in this part is on the estimation and spatial variation of the so-called effective ionization parameter (Az) which is used by the Nequick model. In this thesis also an alternative method for estimating Az is proposed. This part corresponds to chapters 4 and 5. In the second part a novel physic-based recursive model for the regular variation of TEC is developed. The main motivation for developing this model is that the empirical models of the first part were either ill–suited or too complicated to model and predict the regular variation of TEC for high precision differential GNSS applications. The recursive model, described in chapter 6, is based on the ionospheric physics described in chapters 2 and 3 which are mainly the result of a literature study. In the third part, which corresponds to chapter 7, a new algorithm for the real-time modeling of the TID effects on the TEC observations is developed. The modeling of TIDs is essential to improve the precision and reliability of high precision GNSS applications such as Network-RTK. The main conclusions are summarized in the following sections.

8.1 Estimation of effective ionization for NeQuick

NeQuick is a semi-empirical ionospheric model which is proposed as a real-time ionospheric correction model for single frequency users of the European Galileo navigation system. The use of NeQuick is not restricted to Galileo and it can be used also – at least in post-process applications – for GPS and GLONASS. The key input parameter of NeQuick is called "the effective ionization level" (Az). This parameter will be provided in the

Galileo navigation message as a second order polynomial in modified dip-latitude. In the nominal approach, the three coefficients of the polynomial will be estimated on a daily basis from at least 20 globally distributed permanent Galileo monitoring stations. As Galileo is under development the coefficients of Az parameter are not yet available. Therefore, NeQuick cannot yet be used as ionospheric correction model, unless users compute Az themselves. The nominal approach, which requires data from the Galileo monitoring stations, cannot be used directly, but can be emulated using daily RINEX data of about 20 stations from the International GNNS Network (IGS).

In chapter 5 an alternative approach is proposed. The alternative approach uses Global Ionospheric Maps (GIM), from the International GNSS service (IGS), to estimate Az. This approach consists of three steps: (1) compute time series of daily vertical total electron content (VTEC) for each grid point of the GIM, (2) estimate Az for each grid point from the VTEC time series, and (3) fit a second order polynomial in modified dip-latitude to represent the spatial variation in Az similar to the proposed parameterization for Galileo. The alternative approach has been validated by comparing slant ionospheric delay at three IGS sites computed from the NeQuick model with Az parameters computed using the alternative approach and Az computed using the nominal approach. The main advantages of the alternative approach over the standard approach are: (1) the alternative approach is more reliable, because, each IGS GIM is based on data of up to 200 GNSS stations world-wide and each IGS GIM is the combination of results of up to four analysis centers, (2) the coefficients of the second order polynomial are more representative for all regions on the world because they are computed from a world-wide 5° and 2.5° grid instead of about 20 distinct locations, (3) with the alternative approach it is possible to provide Az in a different representation, for instance using a higher order polynomial, grid, or other function types, and (4) the computational effort is much smaller assuming the IGS GIMs have already been computed. A further advantage is that the Az can be computed from predicted GIMs that are made available by the CODE analysis center of the University of Berne. A small disadvantage of the alternative approach is that it relies on a product computed by another organization, IGS, and not on raw receiver data, and therefore depends to some extend on procedures and guideline followed by IGS and outside control of the user. However, this is more of an disadvantage for a GNSS system provider than for a regular user.

It can therefore be concluded that the alternative estimation approach for the effective ionization level Az using GIM from IGS is not only accurate and reliable, but also has many other advantages over the standard procedure. It is therefore recommended to set up an operational service to compute Az from IGS Global Ionospheric Maps.

8.2 Spatial and temporal variation of effective ionization level

The estimation of the NeQuick model parameter, Az, on a daily basis, on a world-wide grid, using IGS Global Ionospheric Maps, provides a valuable opportunity to study the spatial and temporal variability of this parameter, and the correlation between Az and the solar radiation intensity. The results in chapter 5 showed that a second order polynomial in modified dip-latitude (as proposed for Galileo) is not the best fitting function to express

the spatial variation of Az. Az behaves as a symmetric function with respect to the modified dip-latitude and there is a local maximum around the geomagnetic equatorial region. Because of this, a fourth-order polynomial would be a better function to express the spatial variation of Az. The NeQuick model was originally developed for the F10.7 index instead of the Az parameter. The Az serves as a replacement for F10.7, whereby F10.7 is determined by solar observatories or from the average sun spot index, and Az is estimated from GNSS data. Therefore, a high correlation between Az and F10.7 is expected. However, from a study on the correlation between Az and the F10.7 index, using a time series from 2003 to 2006, it was concluded that the correlation between Az and F10.7 is not very high: the maximum correlation coefficient was only 0.65. The correlation coefficient between Az and F10.7 depends also on the latitude, with larger correlations observed in the equatorial region. In these experiments is was also found that the lower limit in NeQuick for Az, which was 64 (the same as for F10.7), was exceeded many times. This lower limit for Az turned out to be very unrealistic and the lower limit for Az parameter was reset to a value of zero.

Daily maps of the Az parameter do not have a clear physical meaning. What these maps tell us is the ionospheric activity compared to a model, in this case NeQuick. It helps to identify regions and times with a relative higher or lower ionospheric activity than what was assumed by the model. Therefore, the Az maps can be used to characterize ionospheric activity and provide a local index on ionospheric activity. At the same time, a systematic analysis of Az maps will help to identify potential model improvements and may assist in the validation of model improvements. It is therefore recommended to compute Az maps from IGS Global Ionospheric Maps on a routine basis and to make these available to the scientific community.

In principle, two-hourly maps of Az and the NeQuick model, could be used as a replacement for IGS Global Ionospheric Maps (GIM). In this case the GIM would is "only" an intermediate product in the estimation of the two-hourly Az maps, with the same resolutions as the GIM, but the user would use the Az maps and the NeQuick model to compute the slant TEC. Actually, there is a one–to–one relationship (albeit complex) between the GIM and Az maps. The VTEC that results from both approaches is identical, but not the slant TEC (STEC). This is because the STEC computation in NeQuick is based on profiling, whereas for the GIM a very elementary mapping function is assumed. Although this idea was never tested on real data, we believe this could potentially result in an improved accuracy for the slant TEC. Another possibility is to use the NeQuick model in the estimation of Global Ionospheric Maps (or Az maps) as a "sort of" replacement for the ionospheric mapping function. In this case the best of both worlds would be combined, and would result in the direct estimation of Az-maps, together with satellite and receiver differential code biases, using the full IGS network.

8.3 Performance of global TEC models

Continuously operating GPS stations provide an excellent data source to study the performance of ionospheric models. The performance of the NeQuick and Klobuchar models, and the IGS Global Ionosphere Maps (GIM), was investigated under different ionospheric

conditions in the mid-latitude region, using slant ionospheric delay measured by three permanent GPS stations. The best accuracy was obtained with the GIM model under both quiet and severe ionospheric conditions. As expected, the bias of GIM was zero. The RMS value was at the level of a few decimeters, which corresponds to $1-2$ [TECU], which was better than the officially stated accuracy for the GIM of $2-8$ [TECU]. The GIM model was able to correct the absolute slant ionospheric delay to better than 80% under different geomagetic conditions of the ionosphere. The NeQuick model could correct about 60% of the absolute slant ionospheric delay. Although the slant ionospheric delay computed from the NeQuick model was less accurate than the GIM, it was considerably accurate than the slant delays computed from the Klobuchar model. The performance of the NeQuick model with Az provided from both standard and alternative procedures was comparable.

These results serve as an illustration of typical performances that can be obtained using global TEC models. The best performances that we obtained were at the level of $1-2$ [TECU], or a few decimeters, for the GIM. Although a systematic analysis, using multiple years of data and many stations, was outside the scope of this thesis, these results provide enough evidence that improvements in global TEC models to achieve centimeter accuracy will be very difficult, if not impossible. Certainly, none of the improvements in the estimation of Az that were investigated in this thesis, or the ideas we have on improving Global Ionospheric or Az maps, will result in substantial improvements to achieve centimeter level accuracy. This is because there is a fundamental limitation to TEC models: they assume all electrons are concentrated in a single layer. In reality the electron density varies with height, as discussed in chapter 3, with as result that it is impossible to map slant paths through the ionosphere to a single point in a single layer model. Therefore, increasing the spatial resolution in a single layer model does not result in an substantial improvement, and because of the limited spatial resolution, increasing the temporal resolution does not help either. A possible solution to this problem is to introduce a – data driven – multiple layer model or 3-D electron density model. Actually, the IONEX format, used for the IGS Global Ionospheric Maps, already have the possibility for using a multi-layer model, and could even be used to represent a 3-D electron density model, but the IGS products use a single layer only. The main challenge in estimating 3-D electron density models and multiple layer models from GNSS slant ionospheric delays is the inversion process. Several groups are currently investigating these techniques, using either tomography, 4-D variational assimilation or Kalman filtering, see (Schunk et al., 2004; Hajj et al., 2004; Pi et al., 2003; Kunitsyn and Tereshchenko, 2003; Lee et al., 2008; Allain and Mitchell, 2008). Especially in 4-D variational assimilation the physical modeling is complicated. In addition, tomography and 4-D variational assimilation suffer from estimability problems in the vertical direction when only ground based GNSS data is used. Therefore, most groups add other observations or space based GNSS measurements to the inversion.

In this thesis a different approach has been chosen, partly for lack of resources to develop 3-D density modeling from scratch, but mainly because we are interested in modeling the differential ionosphere for differential GNSS applications. It means we are mainly interested in temporal and spatial variations in the slant ionospheric delay. The main elements in this approach are the development of a physics based model for the regular TEC variations, modeling of much smaller regions of the ionosphere, and the development of a procedure to observe and model Traveling Ionospheric Disturbances (TID).

8.4 Model of temporal evolution of VTEC

In chapter 6, a normal ionosphere was defined to approximate the Earth's ionosphere by considering a spherical stratification for an isothermal atmosphere in which density varies exponentially with height. The normal ionosphere consists of a lower and upper region, respectively with quadratic and linear recombination coefficients. The Chapman production function, with the Chapman grazing incidence function for the absorption of ionizing solar radiation, is used to model the ion production rate at a given time and height in the two regions. The lower region is formed in a photochemical equilibrium, resulting in a so called Chapman layer. The upper region is formed in a diffusive equilibrium, whilst ignoring the geomagnetic field, resulting in a new *Chapman-based ionospheric layer*. The lower and upper region correspond to the E- and F- regions in the Earth's ionosphere respectively. Two differential equations are derived for the temporal evolution of the vertical total electron content (VTEC) by integrating the continuity equation over height, once for the lower, and once for the upper region. Particular solutions for the two differential equations were obtained by solving a boundary value problem (BVP) under certain boundary conditions. The boundary condition is that the VTEC after 24 hours should be the same. The solutions, each giving the temporal evolution of the VTEC for their region, depend on season, local time, and latitude of the point of interest. The solutions also take the recombination coefficient and the solar radiation intensity as input parameters. The diurnal variation in VTEC from these solutions has a minimum around sun-rise and a maximum just after local noon, as expected. Also expected, the level of VTEC at all latitudes is higher in the summer solstice and is lower in the winter solstice w.r.t. the level of VTEC in the equinoxes. Furthermore, VTEC decreases as latitude increases, but the equatorial anomaly is not present as the model basically ignores all transportation processes. For the same reason the model does not feature a second local maximum in the diurnal variation of VTEC, while this is often present in reality. Similar to the real Earth's ionosphere, the level of VTEC associated to the lower region in the model ionosphere is much smaller than the VTEC associated to the upper region.

The two differential equations and their solutions which have been derived in this thesis form a useful addition to the existing theory of the Chapman functions. The Chapman functions already provided, in the photochemical equilibrium, equations for the electron density as function of height, called Chapman-α and -β layers respectively. The new type of Chapman-based ionospheric layer is not formed under photochemical equilibrium. This new layer explains the F-region better than the Chapman-β layer because, instead of the photochemical equilibrium, the diffusive equilibrium is taken into account. To describe the evolution of VTEC over time for different locations, seasons, and solar activity two new functions have been derived. One of them is based on a Chapman-α layer and the other is based on the new Chapman-based ionospheric layer for respectively E- and F- regions. The discrete form lof these equations ed to a new recursive model for the temporal evolution of VTEC with respect to the solar zenith angle at any given point on the ground.

Because the recursive model of VTEC is derived by solving the continuity equation of the electron density in the ionosphere the model can be considered a physics based ionospheric model. The model consists of three unknown parameters: the linear recombination coefficient, the solar radiation intensity and VTEC at t_0. These parameters must be essentially

estimated by fitting the model to a time series of VTEC observations. The performance of the model is tested at local and global scales using GIM. The model can explain the daily pattern of the VTEC at a given point. The RMS of the model error was a few TECU. In the global test, it was found that the best parametrization strategy is a linear recombination coefficient which is constant in space, while the solar radiation intensity and VTEC at t_0 are spatial dependent. With this strategy the model fits better to the GIM maps than other strategies. The distribution of the estimated solar radiation intensity at the grid points revealed a dependency on the geomagnetic latitude. Despite the geomagnetic field was ignored in the model, all the test cases showed that the recursive model gives a good overall representation of the regular variation of VTEC in the mid-latitude region under geomagnetically quiet ionosphere.

This is an important result that shows the potential of the model for a number of applications. For instance, the model can be used for the computation of VTEC from GIM at epochs between the grid intervals instead of linear or quadratic interpolation. Although the potential of these applications is very high the practical use and feasibility of these applications have not been investigated in this thesis. The reason is that this kind of modeling is not very important for centimeter precision differential GNSS applications, the main focus of this thesis. Instead we focused on application of the recursive model for STEC modeling in small ionospheric patches. Because the model is based on a differential equation, it is ideally suited to derive a recursive formulation that can be used as time update equation in a Kalman filter. Two practical applications are discussed in this thesis. One is the modeling of VTEC, the other is to use the model for removing the geometry-dependent trend from time series of GPS-provided ionospheric delays in the modeling of traveling ionospheric disturbances (TID), which is the third part of this thesis.

Ignoring the geomagnetic field in the definition of the normal ionosphere increases the absolute error of the recursive model of VTEC. Future research will attempt to take the geomagnetic field into account. One may think of an empirical modeling of the discrepancy between computed VTEC values from the physic-based model and GIM as a function of local time and latitude.

8.5 Modeling Medium-Scale Traveling Ionospheric Disturbances

Accurate spatial ionospheric prediction is crucial for modern GNSS applications such as Network-RTK system that gives centimeter positioning accuracy for users in the field. In order to get centimeter accuracy after a short (minutes) initialization periods the ionospheric delay for the user's receiver needs to be predicted. This essentially boils down to very precise spatial interpolation between the IP points of the reference receivers at the double difference level. Having the cm-level accuracy in the ionospheric interpolation is essential for the carrier phase ambiguity resolution by the user. To achieve such a level of accuracy, the TID should be taken into account in the ionospheric prediction of the Network-RTK system. There are two major complications for the TID mitigation for Network-RTK that should be addressed. Firstly, mitigation must be taken care of at the double-difference level, which involves taking single differences between ionospheric delays for the same satellite between two different receivers, followed by differencing single differ-

ences for different satellites. This means that two different patches of the ionosphere are involved, each related to a different satellite, and each possibly associated with different TIDs. Secondly, for the Network-RTK a real-time strategy for TID detection and modeling is needed.

In the chapter 7, we proposed a real-time algorithm for detection and modeling of TIDs. The algorithm takes advantage of the high temporal resolution in data from a GPS network and uses the correlation function of time series of the data. A time series of the un-differenced ionospheric delay, provided from GNSS data, is decomposed in a geometry-dependent trend (regular variation) and the TID effects (mostly from medium-scale and large-scale TIDs). As we are mainly interested in medium-scale TID the geometry-dependent trend and the effect of large-scale TID need to be removed. In a medium scale GPS network, between-receiver single-differencing of the ionospheric delays is a simple way to remove the effect of large-scale TID, this is not true for the geometry-dependent trend, because of the horizontal TEC gradient and the Earth's curvature. It is therefore essential to detrend the single-difference data. The detrending can be done in different ways such as a high-pass filtering or a polynomial fit. In this thesis, as an application of the recursive model developed in the chapter 6, we fitted the recursive model to the time series of the single-difference ionospheric delays. Any deviations with respect to the recursive model is attributed to MS-TID effects. An advantage of this approach is that it can be used for detrending of the data in real-time.

The detrended time series in a reference network is the main input of the algorithm for MS-TID modeling. The algorithm consists of the six individual steps: initialization, detection, scraping, cross-correlation, parameter estimation, and ending. A MS-TID is supposed to be as a planar longitudinal traveling wave with spatially independent amplitude that propagates in the ionospheric patch. Characteristic parameters of MS-TID wave (e.g. period, phase velocity, propagation direction, amplitude) were considered to be time dependent and the Doppler shift caused by the satellite motion was taken into account in the estimation step.

The performance of the algorithm was tested in a medium scale permanent GPS network in time span of 5 hour for two satellites. The tracks of the detected TIDs showed that a TID wavelike fluctuations in the ionosphere not only is not a perfect planar wave but also its amplitude is spatially dependent. The new algorithm was able to model (in time and in space) TID effects on data that its performance for Network-RTK was comparable with the Kriging interpolation method. That is an important result because these two methods are developed using totally different principles, but give similar result. The Kriging is basically a linear spatial interpolation method that takes spatial correlations into account. Our algorithm has been developed based on the physics of ionosphere using the planar wave approximation for the TID. An advantage of our algorithm is that there is a possibility to improve the algorithm for the phase-offset modeling by avoiding the planar wave approximation, while is not possible Kriging.

A number of recommendations can be given for improvement of the new algorithm: (1) Estimating the TID phase velocity as a function of azimuth, and avoid the plane wave approximation in the estimation of the propagation parameters, (2) The amplitude of the TID can be considered spatial-dependent.

The quality of the estimated TID parameters still needs to be evaluated. The algorithm

for TID modeling is not invariant with respect to the reference point in the network. This implies that for different selection of the reference point the estimates of the propagation parameters can be differed. Further research should be done either to develop a strategy for selection of the best reference point in the network or to use all possible baselines in the network at the same time. The quality of the estimated parameters also depend on the geometry of the network. Further analysis and development are recommended for these area.

The real time algorithm for MS-TID modeling can also be employed in an active reference GPS network as a tool for TID monitoring.

Bibliography

Afraimovich, E. L., K. S. Palamartchoul, and N. P. Perevalova (1998). Gps radio interferometry of travelling ionospheric disturbances. *Journal of Atmospheric and Solar-Terrestrial Physics 60*, 1205–1223.

Afraimovich, E. L., V. V. Vodyannikov, S. V. Voyeikov, N. P. Perevalova, and A. F. Yakovets (2002). Dynamical characteristics of traveling wave packets of total electron content disturbances. *Radiophysics and Quantum Electronics 45*(10).

Allain, D. J. and C. N. Mitchell (2008). Ionospheric delay corrections for single-frequency GPS receivers over europe using tomographic mapping. *GPS Solution 13*, 141–141. doi:10.1007/s10291-008-0107-y.

Anderson, D. N. (1973). A theoretical study of the ionospheric F region equatorial anomaly - I theory. *Planet. Space Science 21*, 409–419.

Anderson, D. N. (1981). Modeling the ambient, low latitude F-region ionosphere: A review. *Journal of Atmospheric and Solar-Terrestrial physics 43*(8), 753–762.

Azpilicueta, F., P. Coisson, B. Nava, C. Brunini, and S. M. Radicella (2003). Optimized nequick ionospheric model for point positioning. In *Proceedings of International Symposium on GPS/GNSS*, November 15-18 2003, Tokyo, Japan.

Bailey, G. J., N. Balan, and Y. Z. Su (1997). The sheffield university plasmasphere ionosphere model: A review. *Journal of Atmospheric and Solar-Terrestrial physics 59*(13), 1541–1552.

Banks, P. M. (1969). The thermal structure of the ionosphere. *Proceedings of the IEEE 57*(3). (March,1969).

Bassiri, S. and G. A. Hajj (1993). Higher-order ionospheric effects on the Global Positioning System observables and means of modeling them. *Manuscripta Geodaetica 18*, 280–289.

Bendat, J. S. and A. G. Piersol (1971). *Engineering application of correlation and spectral analysis*. A Wiley-Interscience publication, John Wiley & Sons, INC. Second edition.

Bilitza, D. (1990). International reference ionosphere 1990, report 90-22. Technical report, National Space Science Data Center/World Data Center A for Rockets and Satellites, Code 930.2, Goddard Space Flight Center.

Bradley, P. A. (1990). Mapping the critical frequency of the F2-layer: Part 1 - requirements and developments to around 1980. *Adv. Space Res. 10*(8), 47–56.

Brent, P. R. (1973). *Igorithms for Minimization without derivatives*. by Prentice-Hall, Inc., Englewood Califs, N.J., in U.S.A.

Brunner, F. K. and M. Gu (1991). An improved model for the dual frequency ionospheric correction of gps observations. *Manuscripta Geodaetica 16*, 205–214.

Calais, E., J. S. Haase, and J. B. Minster (2003). Detection of ionospheric perturbations using a dense gps array in southern california. *Gheophysical Research Letters 30*(12). doi:10.1029/2003GL017708.

Candido, C. M. N., A. A. Pimenta, J. A. Bittencourt, and F. Becker-Guede (2008). Statistical analysis of the occurrence of medium-scale traveling ionospheric disturbances over brazilian low latitudes using oi 630.0 nm emission all-sky images. *Gheophysical Research Letters 35*. doi:10.1029/2008GL035043.

Cappellari, J. O., C. E. Velez, and A. J. Fuchs (1976). Mathematical theory of the goddard trajectory determination system, Report X-582-76-77. Technical report, Goddard Space Flight Center, Greenbelt, MD, U.S.A., April 1976, Section 7.6.2, pp 7-44, 7-52.

Chapman, S. (1931a). The absorption and dissociative or ionizing effect of monochromatic radiation in an atmosphere on a rotating earth. *Proc. Phys. Soc. 43*(26), 26–45.

Chapman, S. (1931b). The absorption and dissociative or ionizing effect of monochromatic radiation in an atmosphere on a rotating earth, part II. Grazing incidence. *Proc. Phys. Soc. 43*(26), 483–501.

Chapman, S. (1931c). The absorption and dissociative or ionizing effect of monochromatic radiation in an atmosphere on a rotating earth, part ii. grazing incidence. In *Proc. Phys. Soc.*, Vol. 43, No. 26, London.

Chong, E. and S. Zak (2001). *An Introduction to Optimization*. A Wiley-Interscience Publication, JOHN WILEY & SONS, INC.

Clark, R. M., K. C. Yeh, and C. H. Liu (1971). Interaction of internal gravity waves with the ionospheric F2-layer. *Journal of Atmospheric and Terrestrial physics 33*, 1567–1576.

CODE (2007). *Center for Orbit Determination in Europe, Astronomy Institute University of Berne (AIUB)*. Website: *http://www.aiub.unibe.ch/ionosphere.html*.

Coisson, P., S. M. Radicella, R. Leitinger, and B. Nava (2006). Topside electron density in iri and nequick: Features and limitations. *Adv. Space Res. 37*, 937–942.

Davies, K. (1965). *Ionospheric Radio Propagation*. Dover Publications Inc., Manufactured in the United States of America.

Davies, K. (1989). *Ionospheric Radio*. Peter Peregrinus Ltd., London, United Kingdom.

Dendy, R. O. (1993). *Plasma physics, An introductory Course*. The press syndicate of the university of Cambridge.

Dieminger, W., G. K. Hartmann, and R. Leitinger (1996). *The upper atmosphere, Data analysis and interpretation*. Springer Verlag, Berlin-Heidelberg.

Dudeney, J. R. (1983). The accuracy of simple methods for determining the height of the maximum electron concentration of the f2-layer from scaled ionospheric characteristics. *J. Atmos. Terr. Phys. 45*(8/9).

Feltens, J. (1998). Chapman profile approach for 3-d global TEC representation. In *Proceedings of the 1998 IGS Analysis Centers Workshop, ESOC*, February, Darmstadt, Germany, pp. 285–297.

Garrison, J. L., G. S.-C. Lee, J. S. Haase, and E. Calais (2007). A method for detecting ionospheric disturbances and estimating their propagation speed and direction using a large gps network. *Radio Science 42*. doi:10.1029/2007RS003657.

Giovanni, G. D. and S. R. Radicella (1990). An analytical model of the electron density profile in the ionosphere. *Adv. Space Res. 10*, 27–30.

Giraud, A. and M. Petit (1978). *Ionospheric techniques and phenomena*. Geophysics and astrophysics monographs, D. Reidel Publishing Company.

Gran, A. A. (1965). Solar ionizing flux as determined from sunrise measurements of electron content. Scientific report no. 240, The Pennsylvania state university, May 30, 1965.

Hajj, G. A., B. D. Wilson, C. Wang, X. Pi, and I. G. Rosen (2004). Data assimilation of ground gps total electron content into a physics-based ionospheric model by use of the kalman filter. *Radio Science 39*(RS1S05). doi:10.1029/2002RS002859.

Hall, M. P. M., L. W. Barclay, and M. T. Hewitt (1996). *Propagation of radio waves*. The Institute of Electrical Engineers, London.

Hanbaba, R. (1995). Statistical use of ionosonde data for IRI. *Adv. Space Res. 15*(2), 17–22.

Hartmann, G. K. and R. Leitinger (1984). Range errors due to ionospheric and tropospheric effects for signal frequencies above 100 MHz. *Bulletin Geodesique 58*, 109–136.

Hein, G. W. and B. Eissfeller (2007). Gnss system of systems, part 1. *Inside GNSS January-February*.

Hernandez-Pajares, M., J. M. Juan, and J. Sanz (2006a). Medium-scale traveling ionospheric disturbances affecting gps measurements: Spatial and temporal analysis. *Journal of Geophysical Research 111*, A07S11,doi:10.1029/2005JA011474.

Hernandez-Pajares, M., J. M. Juan, and J. Sanz (2006b). Real time mstids modeling and application to improve the precise gps and galileo navigation. In *Proceedings of ION GNSS 19th International Technical Meeting of the Satellite Division*, 26-29 September 2006, Fort Worth, TX, USA, pp. 1358–1368.

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

Hochegger, G., B. Nava, S. Radicella, and R. Leitinger (2000). A family of ionospheric models for different uses. *Phys. Chem. Earth (C) 25*(4), 307–310.

Hofmann-Wellenhof, B., H. Lichtenegger, and J. Collins (2001). *Global Positioning System: Theory and practice*. 5th revised edition: Springer-Verlag.

Hofmann-Wellenhof, B., H. Lichtenegger, and E. Wasle (2008). *Global Navigation Satellite Systems: GPS, GLONASS, Galileo and more*. SpringerWienNewYork.

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

ICD-GPS-200C (2004). *GPS Interface Control Document, ICD-GPS*. Navstar GPS Space Segment and Navigation User Interfaces, U.S. Air Force.

Ivanov-Kholodny, G. S. and A. V. Mikhailov (1986). *The Prediction of Ionosphere Conditions*. Institute of applied geophysics, geocohydromet, Moscow, U.S.S.R.

Jones, W. B. and R. M. Gallet (1962a). Methods for applying numerical maps of ionospheric characteristics. *Journal of Research NBS 66D (Radio Propagation)* (6), 649–662. November-December.

Jones, W. B. and R. M. Gallet (1962b). The representation of diurnal and geographic variations of ionospheric data by numerical methods. *Telecommunication J. 29*(5), 129–149. (May 1962).

Jong, C. D. D. (1998). A unified approach to real-time integrity monitoring of single- and dual-frequency gps and glonass observations. *Acta Geod. Geoph. Hung. 33*, 247–257.

Jonge, P. D. and C. C. J. M. Tiberius (1996). The LAMBDA method for integer ambiguity estimation: implementation aspects. Technical report, Geodetic Computing Centre, LGR Series, 12, 45 p.

Kalikhman, A. D. (1980). Medium scale travelling ionospheric disturbances and thermospheric winds in the F-layer. *Journal of Atmospheric and Terrestrial physics 42*, 697–703.

Kleijer, F. (2004). *Troposphere modeling and filtering for precise GPS leveling*. Ph. D. thesis, Delft University of Technology, Publication on Geodesy, 56, Netherlands Geodetic Commission, Delft.

Kunitsyn, V. E. and E. D. Tereshchenko (2003). *Ionospheric Tomography*. Springer Verlag, Berlin-Heidelberg.

Lee, J. K., F. Kamalabadi, and J. J. Makela (2008). Three-dimensional tomography of ionospheric variability using a dense gps receiver array. *Radio Science 43*(RS3001). doi:10.1029/2007RS003716.

Leick, A. (2003). *GPS Satellite Surveying (3rd ed.)*. John Wiley and Sons, New York.

Leitinger, R., L. Ciraolo, L. Kersley, S. S. Kouris, and P. Spalla (2005). Relations between electron content and peak density: Regular and extreme behavior. *Annals of Geophysics 47*(2/3). Supplement to Vol. 47.

Leitinger, R., M. L. Zhang, and S. M. Radicella (2005). An improved bottomside for the ionospheric electron density model nequick. *Annals of Geophysics 48*(3).

Lejeune, S. and R. Warnant (2008). A novel method for the quantitative assessment of the ionosphere effect on high accuracy gnss applications, which require ambiguity resolution. *Journal of Atmospheric and Solar-Terrestrial Physics 70*, 889–900.

Matsushita, S. and W. Campbell (1967). *Physics of geomagnetic phenomena*. Academic Press INC. (London) LTD. International Geophysics series, volume I.

Memarzadeh, Y. and H. van der Marel (2006a). A comparison between nequick-derived and measured slant ionospheric delay from GPS data. In *Proceedings of The European Navigation Conference 2006*, 8-10 May, Manchester, The UK (ENC-GNSS 2006).

Memarzadeh, Y. and H. van der Marel (2006b). Spatial dependency of the effective ionization level parameter for NeQuick. In *Proceedings of 3rd ESA workshop on satellite navigation user equipment technologies, NAVITEC 2006*, 11-13 December 2006, ESA/ESTEC, Noordwijk, The Netherlands.

Mercier, C. (1986). Observations of atmospheric gravity waves by radiointerferometry. *Journal of Atmospheric and Terrestrial Physics 48*(7), 605–624.

Misra, P. and P. Enge (2006). *Global positioning system: Signals, Measurements, and Performances*. Ganga-Jamuna Press, In the united states of America. Second edition.

Nagy, A. F., J. D. Winningham, and P. M. Banks (1973). The effect of conjugate photoelectron impact ionization on the pre-dawn ionosphere. *Journal of Atmospheric and Terrestrial Physics 35*, 2289–229139.

Nava, B., P. Coisson, G. M. Amarante, F. Azpilicueta, and S. M. Radicella (2005). A model assisted ionospheric electron density reconstruction method based on vertical TEC data ingestion. *Annals of Geophysics 48*(2).

Nava, B., P. Coisson, and S. M. Radicella (2008). A new version of the nequick ionosphere electron density model. *Journal of Atmospheric and Solar-Terrestrial Physics 70*, 1856–1862.

Nava, B., S. M. Radicella, R. Leitinger, and P. Coisson (2007). Use of total electron content data to analyze ionosphere electron density gradients. *Advances in Space Research 39*, 1292–1297.

NOAA (2005). *National Oceanic and Atmospheric Administration, Space Environment Center*. Website: *www.sec.noaa.gov*.

Odijk, D. (2002). *Fast precise GPS positioning in the presence of ionospheric delays*. Ph. D. thesis, Delft University of Technology, Publication on Geodesy, 52, Netherlands Geodetic Commission, Delft.

Parkinson, B. W. and J. J. Spilker (1996). *Global positioning system: Theory and applications*. American institute of aeronautics and astronautics.

Perez, R. O. (2005). *Contributions on the improvement, assessment and application of the Global Ionospheric VTEC Maps computed with GPS data*. Ph. D. thesis, Universitat Politechnica de Catalunya (UPC) Spain, Research group of Astronomy and Geomatics.

Pi, X., M. Mendillo, and M. W. Fox (1993). Diurnal double maxima patterns in the f region ionosphere: Substorm-related aspects. *Journal of Geophysical Research 98*(A8), 677–692. 13.

Pi, X., C. Wang, G. A. Hajj, G. Rosen, B. D. Wilson, , and G. J. Bailey (2003). Estimation of e×b drift using a global assimilative ionospheric model: An observation system simulation experiment. *Journal of Geophysical Research 108*(A2). doi:10.1029/2001JA009235.

Prieto-Cerdeira, R., R. Orus, and B. Arbesser-Rastburg (2006). Assessment of the ionospheric correction algorithm for Galileo single frequency receivers. In *Proceedings of 3rd ESA workshop on satellite navigation user equipment technologies, NAVITEC 2006*, 11-13 December 2006, ESA/ESTEC, Noordwijk, The Netherlands.

Prolss, G. W. (2004). *Physics of the Earth's space environment, An introduction*. Springer-Verlag Berlin Heidelberg.

Radicella, S. M. and R. Leitinger (2001). The evolution of the DGR approach to model electron density profiles. *Adv. Space. Res. 27*(1), 35–40.

Radicella, S. M., R. Leitinger, B. Nava, and P. Coisson (2003). A flexible 3d ionospheric model for satellite navigation applications. In *Proceedings of International Symposium on GPS/GNSS*, Tokyo, Japan, November 15-18.

Radicella, S. M. and M. L. Zhang (1995). The improved DGR analytical model of electron density height profile and total electron content in the ionosphere. *Annali di geofisica XXXVIII*(1). March.

Ratcliffe, J. A. (1972). *An introduction to the ionosphere and magnetosphere*. Cambridge University Press, Great Britain.

Rawer, K. (1983). Replacement of the present sub-peak plasma density profile by a unique expression. *Adv. Space. Res. 2*(10), 183–190.

Rieger, M. and R. Leitinger (2002). Assessment of tid activity from GPS phase data collected in a dense network of GPS receivers. *Acta. Geod. Hung. 37(2-3)*, 327–341.

Rishbeth, H. (1972). Thermospheric winds and the f-region: A review. *Journal of Atmospheric and Terrestrial Physics 34*, 1–47.

Rush, C. (1986). Ionospheric radio propagation models and predictions- a mini-review. In *IEEE Transactions on Antennas and Propagation*, Volume AP-34, September.

Schluter, S., Y. Beniquel, C. Bourga, B. Arbesser, N. Jakowski, F. Amarillo, D. Klahn, and T. Noack (2004). Ionosphere related contribution of the atmospheric processing and assessment facility to gstb-v1. In *Proceedings of The European Navigation Conference 2004*, 16-19 May, Rotterdam, The Netherlands.

Schmid, P. E., R. B. Bent, S. K. Liewellyn, G. Nesterczuk, and S. Rangaswamy (1973). NASA-GSFC Ionospheric corrections to satellite tracking data, Report X-591-73-281. Technical report, Goddard Space Flight Center, December.

Schunk, R. W. and A. F. Nagy (2000). *Ionospheres, Physics, Plasma physics and Chemistry*. Cambridge, Atmospheric and Space science series.

Schunk, R. W., L. Scherliess, J. J. Sojka, D. C. Thompson, D. N. Anderson, M. Codrescu, C. Minter, T. J. Fuller-Rowell, R. A. Heelis, M. Hairston, and B. M. Howe (2004). Global assimilation of ionospheric measurements (GAIM). *Radio Science 39*(RS1S02). doi:10.1029/2002RS002794.

Seeber, G. (2003). *Satellite Geodesy: Foundations, methods and applications*. Walter de Gruyter, Berlin, New York.

SIS-ICD (2006). *Galileo Open Service, Signal In Space Interface Control Document, SIS-ICD-2006*. European Space Agency.

Tascione, T. F. (1994). *Introduction to the space environment*. Krieger publishing company, Malabar, Florida. second edition.

Teunissen, P. J. G. (1990). Nonlinear least squares. *Manuscripta geodaetica* (15), 137–150.

Teunissen, P. J. G. (1995). The least-squares ambiguity decorrelation adjustment: a method for fast gps integer ambiguity estimation. *Journal of Geodesy 70*, 65–82.

Teunissen, P. J. G. (2002). Theory of carrier phase ambiguity resolution. In *Proc. of 2002 International Symposium on GPS/GNSS, 9th GNSS Workshop*, November 6-8, 2002, Wuhan, China.

Teunissen, P. J. G. (2003). A new class of gnss ambiguity estimators. *Artificial Satellites 37(4)*, 111–119.

Teunissen, P. J. G. (2004). Integer least-squares. In *Proc. V Hotine-Marussi Symposium on Mathematical Geodesy*, Matera, Italy, 17-21 June, 2002, Reihe: International Association of Geodesy Symposium, Band 127 (2004).

Teunissen, P. J. G. (2007). Influence of ambiguity precision on the success rate of gnss integer ambiguity bootstrapping. *Journal of Geodesy 81*, 351–358.

Teunissen, P. J. G. and A. Kleusberg (1998). *GPS for Geodesy*. Springer, Berlin. Chapter 7: Quality control and GPS, by: P.J.G. Teunissen.

Teunissen, P. J. G., D. G. Simons, and C. C. J. M. Tiberius (2005). *Probability and observation theory*. Faculty of Aerospace Engineering, Delft University of Technology: Delft University. Lecture notes AE2-E01.

Titheridge, J. E. (2000). Modeling the peak of the ionospheric E-layer. *Journal of Atmospheric and Solar-Terrestrial Physics 62*, 93–114.

Tsugawa, T., A. Saito, and Y. Otsuka (2004). A statistical study of large-scale traveling ionospheric disturbances using the gps network in japan. *Journal of Geophysical Research 109*. doi:10.1029/2003JA010302.

van der Marel, H. (1993). Modelling of GPS ionospheric delays for geodetic applications. In *URSI commissie G meeting*, March 12, Eindhoven, The Netherlands.

van der Marel, H. (1998). Virtual GPS reference stations in the netherlands. In *Proceedings of ION GNSS 11th International Technical Meeting of the Satellite Division*, September 15-18 1998, Nashville, USA, pp. 49–58.

Velthoven, V. (1990). *Medium scale irregularities in the ionospheric electron content*. Ph. D. thesis, Technische Universititiet Eindhoven, Eindhoven.

Verhagen, A. A. (2005). *The GNSS integer ambiguities: estimation and validations*. Ph. D. thesis, Delft University of Technology, Publication on Geodesy, 58, Netherlands Geodetic Commission, Delft.

Walt, M. (1994). *Introduction to Geomagnetically Trapped Radiation*. Cambridge, Atmospheric and Space science series.

Wienia, R. J. (2008). Use of global ionospheric maps for precise point positioning. MSc scriptie, Delft University of Technology, Faculty of Aerospace Engineering, Section Mathematical Geodesy and Positioning (MGP).

Index