

Die approbierte Originalversion dieser Dissertation ist an der Hauptbibliothek der Technischen Universität Wien aufgestellt (<http://www.ub.tuwien.ac.at>).

The approved original version of this thesis is available at the main library of the Vienna University of Technology (<http://www.ub.tuwien.ac.at/englweb/>).



TECHNISCHE
UNIVERSITÄT
WIEN

VIENNA
UNIVERSITY OF
TECHNOLOGY

DISSERTATION

VLBI as a tool to probe the ionosphere

carried out in order to obtain the academic degree “doctor of the technical sciences” under the supervision of

o.Univ.-Prof.Dr.-Ing. Harald Schuh

E128

Institute of Geodesy and Geophysics
Advanced Geodesy

presented at the Vienna University of Technology
Faculty of Mathematics and Geoinformation

by

Dipl. Ing. Thomas Hobiger

E9725159

1020 Wien, Engerthstrasse 161/603

Vienna, October 19th, 2005



DISSERTATION

Untersuchung der Ionosphäre mittels VLBI

ausgeführt zum Zwecke der Erlangung des akademischen Grades eines Doktors der technischen Wissenschaften unter der Leitung von

o.Univ.-Prof.Dr.-Ing. Harald Schuh
E128
Institut für Geodäsie and Geophysik
Höhere Geodäsie

eingereicht an der Technischen Universität Wien
Fakultät für Mathematik und Geoinformation

von
Dipl. Ing. Thomas Hobiger
E9725159
1020 Wien, Engerthstrasse 161/603

Wien, am 19. Oktober, 2005

SUPERVISORS

o.Univ.-Prof.Dr.-Ing. Harald Schuh
E128
Institute of Geodesy and Geophysics
Advanced Geodesy

Dr. Tetsuro Kondo
Kashima Space Research Center
National Institute of Information
and Communications Technology, Japan

聞くは一時の恥、聞かぬは末代の恥
"If you ask you might be blamed for one time,
but if you don't ask you will be blamed forever"
(Japanese proverb)

Abstract

Precise measurements with dual frequency instruments have a long tradition in geodesy. Among several techniques Very Long Baseline Interferometry (VLBI) started its routine observations already in the late 80ies and has carried them out until today. The dual frequency receiving systems, operating at X- and S-band, were designed to correct ionospheric time delays within geodetic analysis. The potential for ionospheric research of this correction was disregarded until the community around the Global Position System (GPS) showed that parameters of the ionosphere can be deduced from dual frequency satellite measurements. With this thesis a method has been developed, which enables estimation of similar parameters from dual frequency VLBI measurements without any external information. Due to the fact that VLBI is a differential technique, the calculated ionospheric corrections depend on the differences of the propagation media over the stations. Additionally, an instrumental delay offset per station causes a bias of the ionospheric measurements. Within this thesis a method is presented, which is capable of estimating ionospheric parameters, i.e. values of vertical total electron content, from VLBI data. The obtained results are cross-validated against GPS, satellite altimetry data and theoretical models of the ionosphere. As VLBI observations cover more than two complete solar cycles, longer than all other space geodetic techniques using radio signals, the relation to space weather indices on long time-scales can be shown. Generally it can be stated that the overall agreement between VLBI and GPS is within the formal error of each technique and that both systems detect the same periods of ionospheric variations. But only VLBI is able to reveal long period signals like the solar cycle, since it covers a sufficiently long time-span. Apart from explanations for small biases among the techniques also deficiencies of theoretical models are discussed. Instrumental biases, a by-product of ionospheric parameter estimation, demonstrate how receiving systems evolved with the time, as instrumental changes are absorbed in this parameter. The usage of (fringe) phase information from VLBI measurements is a new and challenging field of research, which can be utilized for the detection of short period variations (scintillations). A method for the extraction of such disturbances is discussed and by an example it is shown that such a short period variation can be detected very precisely. Therefore VLBI can be used to detect both, long-term trends and short period variations of the ionosphere and thus it can contribute to ionospheric research as a new independent technique.

Zusammenfassung

Präzise Messungen auf zwei Frequenzen haben eine lange Tradition in der Geodäsie. Unter den verschiedenen Verfahren hat die Radiointerferometrie auf langen Basislinien (VLBI) bereits in den späten achtziger Jahren ihre routinemäßigen Beobachtungen begonnen und führt diese immer noch in beinahe unveränderter Art aus. Zwei-Frequenzempfänger, die Signale im X- und S-Band aufzeichnen, wurden konstruiert, um die ionosphärischen Laufzeitverzögerungen im geodätischen Auswerteprozess korrigieren zu können. Jedoch wurde der eigentlichen physikalischen Bedeutung dieser Korrektur lange Zeit keine Aufmerksamkeit gewidmet. Erst nachdem die Möglichkeit aufgezeigt worden war, Ionosphärenparameter aus GPS-Messungen zu bestimmen, setzte das Interesse ein, physikalische Zustandsgrößen der Ionosphäre mittels VLBI zu bestimmen. Diese Arbeit stellt ein Verfahren zur Bestimmung solcher Parameter vor, ohne Information von anderen Techniken zu benötigen. Da jedoch die VLBI ein differentielles Verfahren ist, liefert die gemessene Ionosphärenkorrektur nur eine Aussage über die Unterschiede des Ionosphärenzustandes zwischen den Stationen. Zusätzlich verfälschen instrumentelle Einflüsse an den Stationen die Ionosphärenkorrekturen. Innerhalb dieser Arbeit wird gezeigt wie Ionosphärenparameter, im speziellen vertikale Gesamtelektronendichtewerte, aus VLBI-Messungen gewonnen werden können. Die daraus erhaltenen Zeitserien werden mit GPS-Modellen, Satellitenaltimetrie-Messungen und theoretischen Modellen der Ionosphäre verglichen. Da VLBI als einziges geodätisches Weltraumverfahren, das im Radiofrequenzbereich beobachtet, den Zeitraum der letzten beiden Sonnenzyklen abdeckt, können auch die Zusammenhänge mit Parametern des Weltraumwetters untersucht werden. Die ermittelten Werte von VLBI und GPS stimmen innerhalb ihrer formalen Fehler überein und die Zeitserien aus beiden Verfahren beinhalten dieselben Perioden. Einzig VLBI ist in der Lage, langperiodische Signale wie den elfjährigen Sonnenzyklus exakt aufzudecken. Im weiteren Verlauf der Arbeit werden kleine systematische Unterschiede auf meßspezifische Fehlerquellen und Modellierungsansätze zurückgeführt und es wird gezeigt, daß sich theoretische Modelle unter Berücksichtigung der VLBI-Ergebnisse verbessern lassen. Die mitgeschätzten instrumentellen Einflüsse regieren auf kleinste Veränderungen an den Stationen und zeigen deutlich die Evolution der Empfangssysteme. Die Verwendung der Phaseninformation (fringe phases) als neues Forschungsgebiet innerhalb der geodätischen VLBI, wird ebenfalls behandelt. Phaseninformationen innerhalb eines Beobachtungszeitraums ermöglichen einen Nachweis von kurzperiodischen Variationen der Ionosphäre, sogenannten Szintillationen. In einem eigenen Kapitel dieser Arbeit wird gezeigt, wie solche ionosphärischen Störungen aus den Daten gewonnen werden und wie gut die Ergebnisse mit Resultaten anderer Verfahren übereinstimmen. VLBI kann daher eingesetzt werden, um sowohl langperiodische also auch kurzperiodische Variationen der Ionosphäre aufzuzeigen und bietet sich als neues unabhängiges Verfahren für die Untersuchung der Ionosphäre an.

Contents

1	Introduction	1
1.1	Outline	1
2	The ionosphere	3
2.1	Introduction	3
2.2	Temporal and spatial variations of the ionosphere	4
2.2.1	Temporal variations	5
2.2.2	Spatial variations	6
2.3	The Chapman layer	10
2.3.1	Electron production	11
2.3.2	Recombination	11
2.3.3	Vertical transport	12
2.4	Models of the ionosphere electron density	13
2.4.1	Chiu ionospheric model	13
2.4.2	Bent ionospheric model	14
2.4.3	International Reference Ionosphere	14
2.4.4	Parameterized Ionospheric Model	14
2.5	Radio waves in a magnetic plasma	14
2.5.1	Wave propagation in a slightly inhomogeneous medium	18
2.5.2	Transionospheric wave propagation	19
3	Very Long Baseline Interferometry	21
3.1	Introduction	21
3.2	Basic principles of VLBI	22
3.3	The VLBI system	23
3.3.1	Receiving systems	23
3.3.2	The correlator	24
3.3.3	Precise determination of delay and delay rate	29
3.3.4	Measurement precision	32
3.4	Applications	33
3.4.1	Astronomical applications	34
3.4.2	Geodetic and astrometric applications	36
4	VLBI and the ionosphere	39
4.1	Impact of the ionosphere on space geodetic techniques	39
4.1.1	Index of refractivity for phase delay of space geodetic techniques	40
4.1.2	Index of refractivity for group delay to space geodetic techniques	44
4.2	Treatment of ionospheric delays in VLBI processing	46
4.2.1	Ionospheric contribution to delays	46
4.2.2	Ionosphere free linear combinations	48
4.3	Instrumental biases	48
4.4	Treatment of interplanetary and interstellar plasma	49
4.5	The thin-shell approximation and its mapping function	51
5	Determination of ionospheric parameters from group delay measurements	53
5.1	Methods and algorithms	53
5.1.1	The station-dependent model	55
5.1.2	Piece-wise linear function	56
5.1.3	Instrumental offsets and their treatment	57
5.1.4	Proper height of the thin-shell layer	58
5.1.5	Adjustment and non-negative constraints	60
5.1.6	Stochastic model, weighting of the data	62

5.1.7	Formal error of the results	63
5.2	Results for globally distributed stations	64
5.2.1	Results for stations located in different ionospheric regions	65
5.2.2	Daily mean values of globally distributed stations	67
5.2.3	Formal error of the estimated VTEC	69
5.3	Gradient-like parameter c_i	69
5.4	Frequency characteristics of the results	71
5.4.1	Treatment of irregularly sampled data - the CLEAN algorithm	72
5.4.2	Wavelet analysis	76
5.5	Comparison to ionospheric models derived from GPS measurements	79
5.5.1	Finding the optimum weight parameter k from CONT02 measurements	79
5.5.2	Differences between GPS and VLBI results	81
5.6	Comparison to satellite altimetry data	84
5.7	Comparison to a theoretical model - IRI2001	85
5.8	Correlations with solar and geomagnetic indices	88
5.8.1	Correlations with sunspot numbers	89
5.8.2	Correlations with solar flux	90
5.8.3	Correlations with Kp and Dst	91
5.9	Station-dependent instrumental offsets	91
6	Determination of ionospheric parameters from intra-scan phases	94
6.1	Functional and stochastic model	94
6.2	An example for the detection of short-period TEC variations	95
6.2.1	VLBI data	95
6.2.2	GPS data	98
6.3	Possible fields of application and requisites	100
7	Summary and outlook	101
7.1	Summary	101
7.2	Outlook	101
A	Solving the quadratic programming problem	103
B	Estimation of Fourier components of unequally sampled data	105
C	BWS plots for baseline HARTRAO - HOBART26, Jan. 7th, 2004	110
	Bibliography	112
	Acknowledgments	117
	Curriculum Vitae	118

1 Introduction

Based on investigations of geomagnetism in the 19th century, the scientific interest in an environment, called the ionosphere, arose about 80 years ago with the increasing use of radio waves in communication. First studies in this new field of research were made during the 1930s, and World War II significantly pushed the observation technique as the development of the radar was intensified by all countries. The military technology of this era also brought rocketry as a tool for high altitude sounding and provided the basics for the first satellite launches in the late 1950s, which within only a few years brought a great expansion of space activity. Direct and indirect observation methods were developed and applied to measure time- and space-dependent variations of different quantities which characterize the ionosphere. With the inauguration of the first satellites of the Global Positioning System (GPS) in the 1980s it became possible to monitor ionospheric parameters using a dense and well distributed station network in order to cover most parts of the Earth. In interaction with satellite altimetry missions and occultation measurements even gaps over the oceans were closed and global coverage could be achieved.

But what is the ionosphere? According to Hargreaves (1992, [29]) the ionosphere refers to the ionized part of the atmosphere which contains significant numbers of free electrons and positive ions that exert a great influence on the medium's electrical properties. This means that all radio waves traveling through the ionosphere are effected by a change in their propagation characteristics.

Approximately at the same time when the interest in the ionosphere started to grow Jansky discovered the radio emission from the Milky way, which is now seen as the birth of the new science of radio astronomy. When he published his results in his paper in December, 1932 he could not foresee the big impact of his discovery on astronomy and related science fields. His investigations enabled scientists to observe objects which could not be seen with optical telescopes and provided a basis for new astrophysical models and theories. Soon after first successful single antenna experiments it was clear that two or more phase connected antennas, utilized like an interferometer, can be used to observe objects with the same angular resolution which would be obtained by a single dish instrument with a diameter equal to the distance between the antennas. The first radio antenna arrays connected by cables were called "connected-element interferometers" and used to make images of distant radio sources. Later, as antenna separation distances were too large for stable phase connection, high frequency oscillators, so called "clocks", with sufficient time-stable characteristics were installed at each antenna to synchronize the data in a later processing stage. In the 1970s the development of atomic clocks, which are essentially frequency standards, experienced a great progress in manufacturing very stable and accurate clocks. From now on it became possible to separate antennas up to a distance of the Earth diameter, which was called Very Long Baseline Interferometry (VLBI). Approximately at the same time the geodetic community got interested in this technique as terrestrial reference frames could be realized and Earth orientation parameters could be observed by this astronomical tool, too. Major achievements in the 1980s revealed plate tectonics between the continents and provided high accurate parameters which relate the Earth's position and rotation to an inertial reference frame, realized by distinct radio sources, so called quasars. A huge contribution to the total error budget of VLBI observation came from unaccounted time delays caused by the ionosphere. Therefore lots of geodetic VLBI sites got equipped with dual frequency receivers to get rid of this restriction. But so far, the ionospheric correction that can be obtained from routine VLBI observations has not been used to deduce characteristic parameters of the ionosphere itself. Except in one paper (Kondo, 1993, [44]) no group has focused how to gain ionospheric information from dual-frequency VLBI measurements. VLBI is only sensitive to the differences of the ionospheric influences at each station pair within a network, is not a continuous observing technique and does not have such a dense distribution of stations like GPS. Nevertheless, VLBI provides capabilities to probe the ionosphere in an absolute sense, which will be shown in this thesis.

1.1 Outline

Chapter 2 provides an introduction to the ionosphere and describes spatial and temporal characteristics of this medium. A model for the electron height distribution is introduced and several theoretical

models of the ionosphere are presented. In the last part of chapter 2 an in-depth discussion about the propagation characteristics of electromagnetic waves in a magnetic plasma is given. The 3rd chapter deals with VLBI, starting with the receiving system and technical requisites for a successful determination of the observables. Thereafter, the correlator and its mathematical background are presented and it is shown how the geodetic observables (delay and delay rate) can be gained. The chapter is completed with an overview of astronomical, astrometric, and geodetic applications. In chapter 4 the impact of the ionosphere on VLBI measurements is investigated. Numerical analyses investigate whether the first order approximation of the ionospheric propagation characteristic is precise enough to obtain mm-accuracy from geodetic measurements at S- and X-band. Furthermore it is described how ionospheric impacts can be removed from the observables and it is revealed that instrumental imperfectness causes additional biases to the measurements. In the last subsection a single layer mapping function is presented, which is capable of transforming slant measurements of total electron content to a vertical measure. The findings from this chapter are used in chapter 5 to develop a method for the determination of ionospheric parameters from group delay measurements. Apart from the functional and stochastic model also physical properties of the mapping function are discussed. A special adjustment method is presented, which takes the physical properties of the estimates into account. Thereafter, results for globally distributed stations are shown and their comparison to GPS and satellite altimetry data is presented. The spectral characteristics of the obtained time-series are analyzed by special Fourier methods and by Wavelet analysis. The VLBI results are compared against theoretical models and the correlation to solar and geomagnetic events is checked. The chapter is concluded with a discussion about instrumental biases and their origin. The 6th chapter shows how short period variations of the ionosphere can be detected by VLBI fringe phase measurements and verifies these results by GPS measurements. The proposed algorithm considers also higher order delay variations and is capable to obtain geodetic observables from parameter period data. The thesis is concluded with a summary and an outlook, both arranged in chapter 7.

2 The ionosphere

2.1 Introduction

The ionized part of the atmosphere, the ionosphere, contains significant numbers of free electrons and positive ions. There are also some negative ions at the lower altitudes. The medium as a whole is electrically neutral, there being equal numbers of positive and negative charges within a given volume. Although the charged particles may be only a minority amongst the neutral ones they exert a great influence on the medium's electrical properties, and herein lies their importance. Already in the 19th century first suggestions of electrified layers within the higher levels of the terrestrial atmosphere were made, but the scientific interest started to grow when Marconi made his experiments by transmitting radio signals from Cornwell in England to Newfoundland in Canada in 1901. Within a cooperation with Kennelly and with suggestions from Heaviside, who worked independently of the other two, it was concluded that, because of the Earth's curvature, the waves must have been reflected by an ionized layer. R. Watson-Watt proposed to call this region ionosphere, which came into common use in about 1932.

Studies of the ionosphere have been carried out since that time and most of the characteristics, though not all, are now fairly well understood and can be explained by physical and chemical processes of the upper atmosphere. A typical vertical structure is shown in figure 1. The identification of the regions

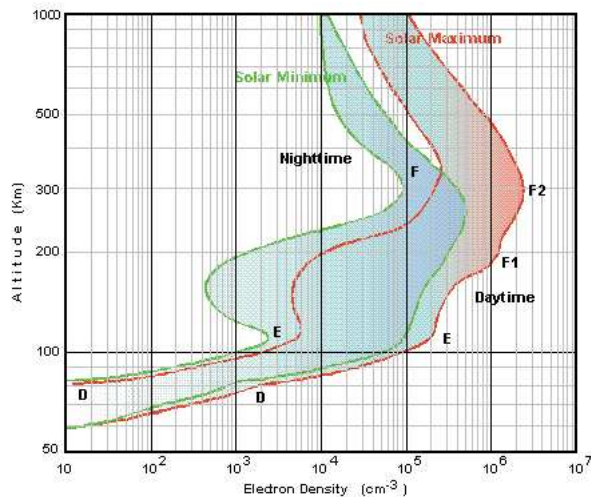


Figure 1: Typical vertical profiles of electron densities in the mid-latitude ionosphere (NASA, 2005, [51]).

was mainly dominated by their signatures as seen in so-called ionograms, which usually emphasize inflections in the profile, but may not be separated by distinct minima. The green lines in figure 1 represent usual profiles as observed during sunspot minima, the red ones show the corresponding behavior during sunspot maxima. The main regions are designated D, E, F1 and F2, with the following daytime characteristics (table 1): The D and F1 region vanish at night, and the E region becomes

region	extent	electron density
D	60-95 km	$10^8 - 10^{10} m^{-3}$
E	55-150 km	several $10^{11} m^{-3}$
F1	160-180 km	several $10^{11} - 10^{12} m^{-3}$
F2	max. around 300 km	up to several $10^{12} m^{-3}$

Table 1: Daytime characteristics of the main ionospheric regions.

much weaker. The F2 region persists though, at a reduced intensity. The behavior of the terrestrial ionosphere indicates that it can be divided into two parts, corresponding to the two principal regimes of magnetospheric circulation (Hargreaves, 1992, [29]).

In this section, first, basic principles and processes will be introduced (following Hargreaves (1992, [29]) and Hakegard (1995, [28])). Later, these foundations will be used to reveal the influence of the ionosphere on signals traveling through this medium (as described by Budden (1988, [11]) and Leitinger (2005, [49])).

2.2 Temporal and spatial variations of the ionosphere

The Earth's ionosphere is strongly related to the solar activity and the behavior of the geomagnetic field. For future use it will be very convenient to introduce two coordinate systems, which take the characteristics of the medium into account.

- Sun-fixed reference system: such a system is defined by three rectangular axes with a geocentric origin, the Z-axis pointing along the terrestrial rotation axis and the X-axis being in the plane defined by the meridian of the Sun.
- Geomagnetic reference system: this system is defined so that its Z-axis is parallel to the magnetic dipole axis. The Y-axis of this system is perpendicular to the geographic poles such that if $\vec{D}\vec{P}$ is the dipole position and $\vec{S}\vec{P}$ is the south pole $\vec{Y} = \vec{D}\vec{P} \times \vec{S}\vec{P}$. Finally the X-axis completes a right-handed orthogonal set. The relation between the geographic coordinates (φ_g, λ_g) and geomagnetic coordinates (φ_m, λ_m) is given in equations (2.1) and (2.2), using the geographic coordinates of the geomagnetic north pole $\varphi_0 = 87.2^\circ N$ and $\lambda_0 = 114.4^\circ W$ (IGRF, 2005, [36]).

$$\sin \varphi_m = \sin \varphi_g \sin \varphi_0 + \cos \varphi_g \cos \varphi_0 \cos (\lambda_g - \lambda_0) \quad (2.1)$$

$$\sin \lambda_m = \frac{\cos \varphi_g \sin (\lambda_g - \lambda_0)}{\cos \varphi_m} \quad (2.2)$$

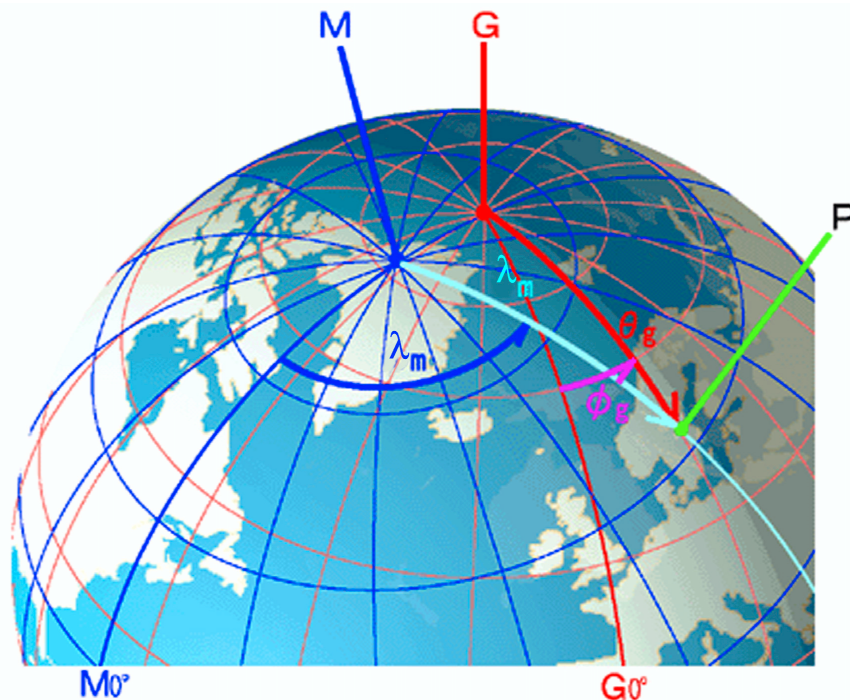


Figure 2: Relation between the geographic and geomagnetic system, where G is the geographic pole, M : geomagnetic pole, P : an arbitrary point, θ_g : geographic co-latitude ($= 90 - \varphi_g$), θ_m : geomagnetic co-latitude ($= 90 - \varphi_m$), and geographic and geomagnetic longitude are expressed by λ_g and λ_m (by courtesy of World Data Center for Geomagnetism, Kyoto (2005, [36])).

Many measurements have been made using different techniques from locations around the world, in order to understand the complex relationship that exists between solar conditions, the geomagnetic field, geographic location, time of day, season, etc. The most important time and spatial variations are discussed in the following subsections.

2.2.1 Temporal variations

Seasonal variations The period of these variations is about one year as the Sun moves from the north to the south with respect to the plane of the equator and reaches a maximum / minimum declination at about $\pm 23^\circ$ (i.e. the so called solistices). The location of the maximum electron distribution changes when the Sun moves from the northern to the southern hemisphere and vice versa.

Storms and disturbances in the ionosphere Sun flares are a good indicator of the solar activity as they have direct consequences for the ionosphere because they are the sources of sporadic particle and electromagnetic emissions, which lead to perturbations of the electron distribution in the Earth's atmosphere. A solar flare is the sudden brightening of a small area of the photosphere that may last from a few minutes to several hours. Flares tend to develop near the long lasting dark areas named sunspots and are classified according to the area covered as class 1, 2 or 3 increasing with size. Additionally the notation of + or - was introduced to indicate sizes larger or smaller than average in class. When a solar flare is observed it is often accompanied by an increase in ultraviolet and X-ray radiation, which raises the ionization in the Earth's ionosphere. This effect is called sudden ionospheric disturbance (SID) and is most noticeable in the lowest region of the ionosphere (D region). This phenomenon can be detected only in the sunlit hemisphere, and it can last from twenty to one hundred minutes. Furthermore it is supposed that at the time when the flare occurs, many charged particles leave the Sun and travel outward at a considerable speed, called Coronal Mass Ejection (CME), producing a disturbance named ionosphere storm, which takes a quite different form to that of a SID. It can easily be noticed in the D region and in a higher F region, where maximum electron concentrations are. Apparently there is a 2 to 4 days gap between the solar flare and the corresponding ionospheric storm that causes it. The effects of the ionospheric storm are greater in the equinoxes and in summer than in winter and are greater in higher latitudes. Therefore the equatorial regions are less affected than the polar regions. Sometimes after the appearance of large flares, a huge increase in the cosmic ray intensity is detected and the flux of the energetic protons produces additional ionization in the D region of the polar ionosphere. Subsequent studies have shown that these effects produce an absorption in the whole polar cap region. Due to this, they are called polar cap absorption events. Acoustic-gravity waves are generated in the neutral atmosphere, which can perturb the ionized component through collision, too. The nature of the ionosphere's response depends mainly on the altitude. As these perturbations show time differences between distant observation sites they are called traveling ionospheric disturbances (TIDs). The wavelengths of these phenomena vary from about 50 to 500 kilometers, whereas the periods can last from a few minutes to several hours. TIDs are more frequently found in mid-latitude regions during the solar maximum, in winter and near the local noon.

Ionospheric scintillations The term scintillation is used to describe the irregular variations of the amplitude or of the phase of a radio signal received after passing through, or being reflected by the ionosphere. The main regions that are affected by scintillations are the equatorial region, the auroral region and the polar cap. They appear to be strongest from afternoon to midnight, weaker in the pre-noon zone, and weakest in the early morning (figure 3).

Solar cycle Another important variation is produced by the solar cycle, which has an average duration of about 11 years. The sunspots (observed in white light) are seen as dark areas, tending to occur in groups and can be found in solar latitudes from 5° to 30° . Usually the Wolf sunspot number R is taken to quantify the solar activity.

$$R = k(f + 10g) \quad (2.3)$$

Where f is the total number of observed sunspots, g is the number of disturbed regions and k is a constant value for each observatory. Figure 4 shows the annual sunspot numbers (derived from daily values) starting from the 17th century. The typical cycle is not symmetric, the time from minimum

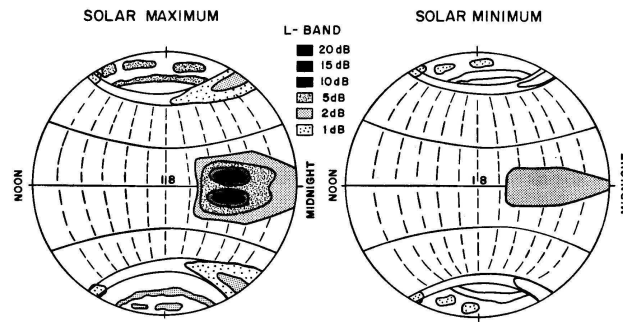


Figure 3: Worst case fading because of ionospheric scintillation during solar maximum and solar minimum (Hakegard, 1995, [28]).

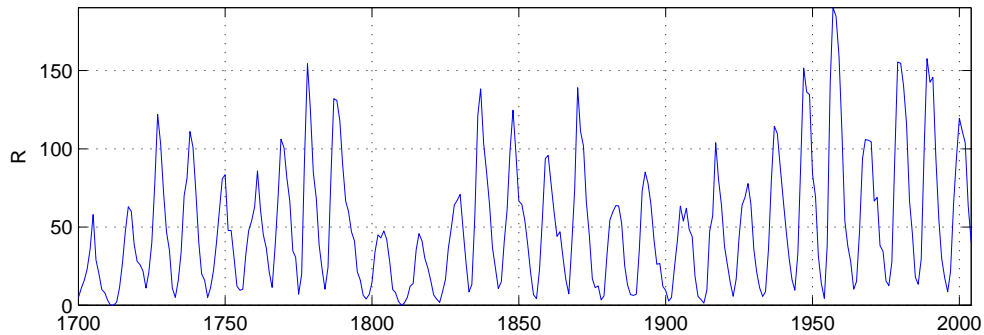


Figure 4: Plot of the annual sunspot numbers 1700-present (NGDC, 2005, [53]).

to maximum is about 4.3 years and the time from maximum to minimum is 6.6 years on average. The cycles are not equally strong, which cannot be seen on long term modulations of the 11 year cycle but in periods of 57 and 95 years. Additionally the 11-year cycle is not stable but varies between 9 and 14 years. The most recent sunspot maximum happened in 2000/2001.

2.2.2 Spatial variations

The free electron distribution changes at different heights and at different latitudes and longitudes defined in a Sun-fixed system. The latitudinal variations make it possible to divide the ionosphere into three quite different zones, whose boundaries are not constant, but vary according to the solar activity and geomagnetic conditions: the equatorial, the mid latitude and the polar zone. The path from the illuminated region to the dark one produces longitudinal variations in the Sun-fixed system.

Vertical structure So called ionospheric sounding measurements can be taken to identify plasma frequencies of the ionosphere as a function of height, which makes it possible to identify typical vertical structure properties. As the density of gas grows from the top to the bottom of the atmosphere, when the ionizing radiation penetrates from outside through the atmosphere, more and more electrons per unit volume are produced. The radiation is absorbed in the process, however, and below a certain height the rate at which the radiation intensity decreases is greater than the rate of gas density increases. Therefore the rate of production of electrons decreases as lower regions are reached. The first studies of the ionosphere were made by radio sounding from the ground. But the ideas of the nature of the processes that produce the ionosphere were incomplete and they could not explain the complex and diverse behavior of the ionosphere. The various ionospheric regions are formed by the effect of various parts of the solar spectrum. They consist of different ion compositions and their formation mechanisms are also different. This section will describe the main characteristics of the different ionospheric layers. As already mentioned in section 2.1, the ionosphere is divided into the region called D, E and F layer (see figure 1) according to the level at which their electron density peaks occur, as recommended by the Committee of the Institute of Radio Engineers.

The D layer The lowest parts of the ionosphere belong to this region going from about 60 to 95 kilometers with an electron density of about 10^6 electrons/ m^3 . The D region is the most complex part of the ionosphere from the chemical point of view. There are several different sources of ionization contributing significantly to ion production. The most important are: Lyman- α that ionizes NO , ultraviolet radiation that ionizes O_2 and N_2 , hard X rays that depend strongly on the solar activity and are not significant in the D region at sunspot minima, and galactic cosmic rays that affect mainly the lower parts in the D region, especially at high and mid latitudes. The relative contributions of these different sources vary with the time of the day (i.e. the longitude in the Sun-fixed coordinate system), latitude and solar activity. This region shows an abrupt change in the electron concentration near sunrise and sunset and remains almost constant during the day. At mid latitude it can be easily observed as it is strongly controlled by the Sun and vanishes in the nighttime. In the D region there is radio absorption, which can be used by multi-frequency measurements to obtain information about the height distribution of this region. It varies with solar zenith angles χ direct proportional to $(\cos \chi)^n$, where n is between 0.7 and 1. During the winter months however, the absorption exceeds the amount that would be expected by a factor two or three. This phenomenon is called winter anomaly of ionospheric radio absorption. Additionally sudden ionosphere disturbances, which were described before, produce an increase in the electron content of the D layer.

The E layer It extends from 95 to 170 kilometers and the electron density is about 10^{11} electrons/ m^3 . The critical frequency (i.e. the frequency which is no longer reflected by an ionospheric layer, but penetrates it) of the E layer varies with time of day, season, epoch in the solar cycle and with the position of the Earth. The variations are regular and are mainly controlled by the Sun and may be described by a Chapman law (see section 2.3), because the electron concentration is governed by fast chemical reactions, and consequently depend on the solar zenith angle rather than on dynamic transport processes. The E layer does not vanish at night, but a weakly ionized layer remains with an electron density of approximately 10^9 electrons/ m^3 . The most important anomaly of this region is called sporadic-E. This phenomenon is seen on ionograms as an echo at constant height, which extends to a higher frequency than usual for the E layer. Sporadic-E tends to be particularly severe at low latitudes, partly due to the instabilities in the equatorial anomalies. At high latitudes, it is attributed to ionization by incoming energetic particles.

The F layer It lies at a height of about 150 kilometers and is divided into two layers with different behavior, called F1 layer and F2 layer.

The F1 layer can be represented by Chapman's theory (section 2.3). Its critical frequency varies as a function of the solar zenith angle as $(\cos \chi)^{1/4}$, indicating so-called α -Chapman behavior. Its presence is more relevant in summer and at sunspot minimum and it is absent at night. The altitude of the F1 layer is around 200 kilometers and the electron density is about $10^{11} - 10^{12}$ electrons/ m^3 .

The F2 layer has the greatest concentration of free electrons, which is about 10^{12} electrons/ m^3 , and unfortunately it is the most variable, anomalous and difficult to predict. It is situated at about 300 kilometers. If the behavior of this region should be explained using Chapman theory many anomalies will be detected, mainly due to the diffusion effect, which produces the F2 peak and is not considered in the model. These anomalies may be summarized as follows:

- The diurnal variation can be asymmetrical with respect to noon. There are rapid changes at sunrise or no changes at sunset. In summer the daily peak occurs before or after local noon and in winter it occurs near noon.
- The daily pattern of variation may not repeat from day to day.
- Noon values of the F2 peak and electron content are usually greater in winter than in summer, although the Chapman theory predicts the opposite. This is called the seasonal anomaly. The average electron content difference between the hemispheres is about 20% larger in December than in June, which is called annual anomaly. The electron content is exceptionally high at the equinoxes, giving the semi-annual anomaly.

- The mid latitude F2 region does not vanish at night, but remains at a substantial level up to the next sunrise.

Several causes are proposed to explain these phenomena, including changes in ion constituents concentration, changes in reaction constants with temperature, ionization by impinging particles and movements of the plasma caused by winds in the neutral atmosphere. The height variations of the electron density for the F2 layer take a variety of shapes, which are often roughly semi-parabolic, with a half-thickness of about 100 kilometers. This shape is particularly common during the night. Near the equator the curve is nearly straight right up to the end. The critical frequency of the F2 layer follows the solar cycle strongly, being highest at sunspot maxima. In the polar regions the variations with the solar cycle are different at different places, hours and seasons. In addition intense irregularities may occur in the equatorial F2 layer, especially after sunset when plasma instabilities develop. So the normal ionospheric condition can be affected by different processes of distinct nature. Some effects are transient (such as TIDs and irregularities), while others can last for several days (ionospheric storms). Figure 5 summarizes the distribution of the ionospheric layers on the day and night side of the Earth.

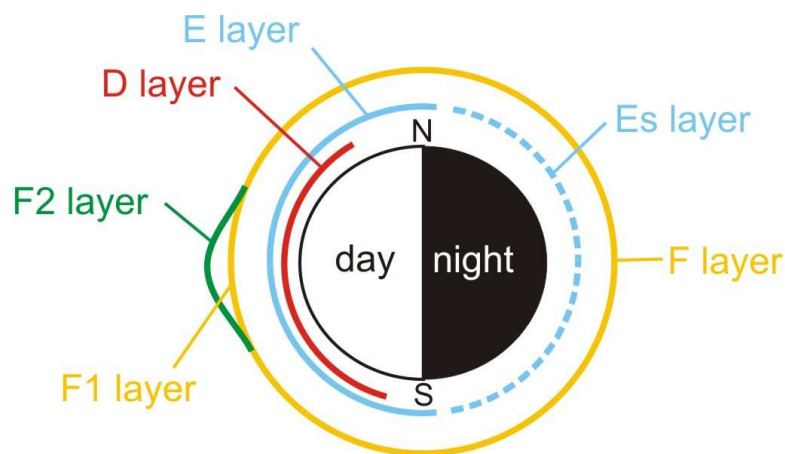


Figure 5: Ionospheric layers, schematic illustration after Wild (1994, [82]).

Latitudinal variations Here the magnetic field will be described, as it is one of the main agents that influence the latitudinal distribution of the electron density in the ionosphere. The magnetosphere is the part of the terrestrial environment where the geomagnetic field has the largest influence. Generally, the energy density of the magnetic field exceeds that of the plasma, which can be stated by

$$\frac{B^2}{2\mu_0} > nkT \quad (2.4)$$

where B is the magnetic flux density, μ_0 is the magnetic permeability of a vacuum, k is the Boltzmann constant, n is the particle density contained in the plasma and T is the plasma temperature. The ratio of energy density between particles and magnetic field, usually written β , is an important parameter. For $\beta \ll 1$ the magnetic field controls the motion of the plasma, for $\beta = 1$ the plasma can change the shape of the magnetic field lines. There is no sharp boundary that defines the magnetosphere, and more semantics than physical definitions separate it from the ionosphere. The outer boundary is determined by the interaction with the solar wind. Since the magnetosphere responds rapidly to changes in the various solar emissions, form and structure are determined largely by emissions from the Sun. The solar corona is not in hydrostatic equilibrium but it expands continuously, with matter leaving the Sun and streaming out into space. This flow is named solar wind. The Sun ejects a lot of matter per second, mainly protons (H^+), but only about 5% of total emission are α particles (He^{++}). The solar wind passes through the bow shock that slows it from being supersonic to subsonic. This effect evokes an increase in the plasma temperatures and the magnetic field, and its density enhances through compression. This shock gives origin to magnetosheath plasma that is immediately adjacent to

the magnetopause (figure 6). Part of the energy produced by this shock penetrates the magnetopause to enter the boundary layer. On the sunward side the geomagnetic field is compressed and behind the Earth it stretches to enormous distances. In the anti-sunward direction the magnetosphere is extended into a long tail, usually known as magnetotail. It has been proved that on the night side of the Earth the magnetospheric field extends beyond about 10 Earth radii running in the Sun-Earth direction. As the solar wind flows past the magnetosphere, it exerts a tangential stress on the magnetopause boundary, driving large-scale plasma convection in the magnetosphere. The tangential stress arises primarily from magnetic field line reconnection at the day side magnetopause. The resulting reconnection-driven convection can be visualized by following a southward solar wind magnetic field line that reconnects with a northward terrestrial field line at the day side magnetopause. The two new open field lines (connecting the Earth to the solar wind) are carried over the poles, where they again reconnect in the tail of the magnetosphere, forming a newly closed terrestrial field line and a solar wind field line. The newly closed field line moves towards the Earth, then around the Earth and out to the day side magnetopause, where the process begins anew. Upon magnetotail reconnection, earthward-bound closed field lines compress the plasma forming the plasma sheet. The Earth magnetic field plays an

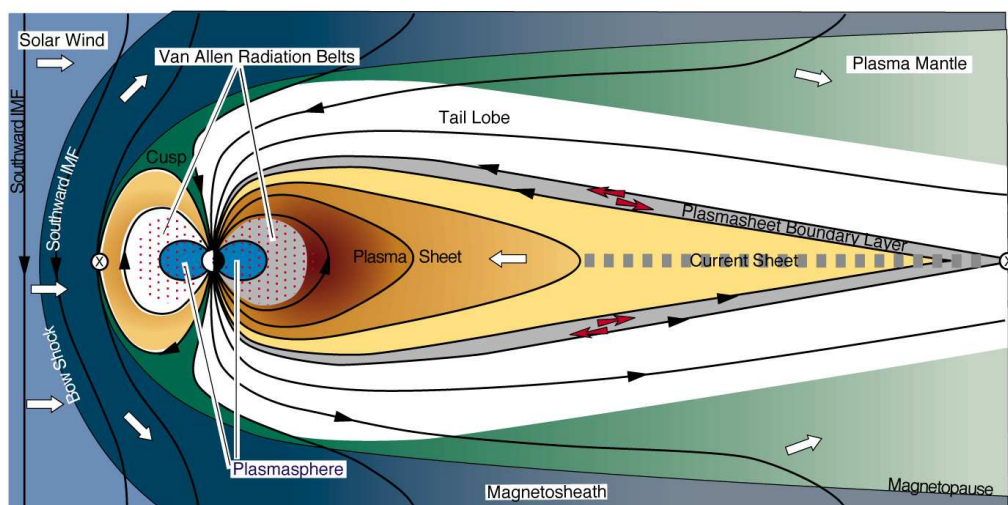


Figure 6: Geomagnetic field (Reiff, 1999, [58]).

important role in the distribution of free electrons. Higher than 4 Earth radii and specially for high latitudes, the magnetic field lines get distorted by the solar wind (figure 6) but within 2 or 3 Earth radii the field can be approximated by a simple dipole field. Although not very accurate (about 30%), the dipolar form is useful for making approximate calculations about the inner magnetosphere. The geographic coordinates and the transformation to geomagnetic coordinates are already given at the beginning of this subsection. Usually the geomagnetic field near the ground varies from about 25000 nT at the equator up to 75000 nT at the poles. The behavior of the ionosphere indicates that it can be divided into three regions: equatorial, mid latitude and high latitude (figure 7). The boundaries between the regions are not constant, but vary according to local time, geomagnetic condition and solar activity.

Equatorial region In this region the so called total electron content (TEC, see section 2.5.2) and the peak electron density have the highest values. As mentioned previously, the scintillation effect has its greatest amplitude at these latitudes. Elongated areas shown in the equator region (figure 7) are known as equatorial anomaly. Various processes significantly disturb these areas, which display a strong diurnal dependence.

Mid latitude region This geographical part of the ionosphere is the least variable one. There exist several ionospheric models that estimate the mean ionosphere in this region with a high degree of accuracy. There are, however, variations in this region as the investigations of day-to-day variability

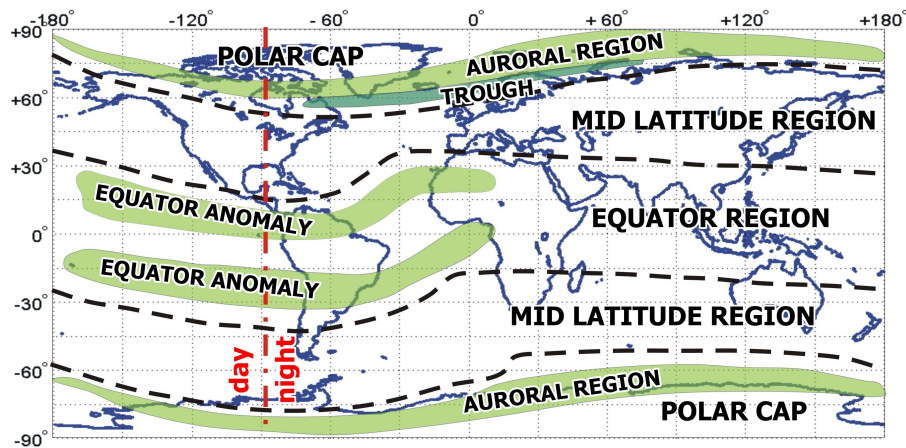


Figure 7: The most important latitudinal regions of the ionosphere.

(difference between average conditions and the current conditions at a fixed local time) have revealed that even in the least variable region, the day-to-day residual mean square (r.m.s.) variations in TEC from the average are about 20-30 %.

High latitude region The main source of ionization is the photoionization, but there is another important driver, which is not entirely absent at mid latitudes and is very important at high latitudes: the ionization by energetic particles. Geomagnetic field lines can guide energetic electrons and protons from magnetosphere or magnetosheath down to the Earth's atmosphere. Precipitating particles lose their energy via collisions with the neutral particles and ionize them at approximately the same altitude range than solar UV radiation when creating the ionosphere. Precipitating particles also lose their energy before collisions with the neutral particles through the particle-wave interaction, which finally generates intense electromagnetic waves named "auroral kilometric radiation (AKR)", whose frequency is 100-500 kHz. In addition, some of the atmospheric constituents are excited to higher energy levels: this can lead to emission of auroral light. Most of this activity occurs within auroral oval (the major exceptions being the Sun-aligned arcs and low-latitude aurorae). The aurora is the most obvious characteristic of the high latitude ionosphere. These features can be visualized in an image of the auroral oval from space. It gives as much illumination on the ground as the full moon does. The location of the aurora at fixed time is not circular but oval. The maximum is near 67° at midnight, increasing to about 77° at noon. Another feature they have in common is that they show sub-storm behavior. They tend to occur in bursts, each lasting about 30-60 minutes, which are separated by quiet intervals of several hours.

2.3 The Chapman layer

When calculating the electron density in an ionized layer, first the electron production has to be analyzed, then the rate at which electrons are removed has to be taken into account. The ionosphere is formed by ionization of atmospheric gases such as N_2 , O_2 , and O . At mid and low latitudes the energy comes from the solar radiation, i.e. the ultraviolet and X rays. At high latitudes energetic particles play an important role, too. The ions and electrons tend to recombine and to react with other gases, which produces other ions. Consequently, the electron density N depends on the relative speed of the production and loss processes and can be described by the continuity equation (2.5).

$$\frac{\partial N}{\partial t} = q - L - \nabla \cdot (N\vec{v}) \quad (2.5)$$

Whereas q represents the production rate, L is the loss rate due to recombination, $\nabla \cdot (N\vec{v})$ expresses the loss of electrons by movement and \vec{v} is the mean drift velocity. At different altitude intervals, the contributing terms of equation (2.5) have different magnitudes. For instance, in E and F1 regions it is feasible to neglect the loss of electrons by movement (transport processes) as compared

with other processes while the typical recombination time constant is so short that under normal conditions $\partial N_e/\partial t = 0$, and the equilibrium concentrations are determined just by the photochemical circumstances. On the other hand, in the upper part of the ionosphere, transport is the main process, while recombination and ionization are insignificant contributions. In the F2 layer all these processes must be considered simultaneously. Furthermore, effects that are related to the non-stationary states become also important.

2.3.1 Electron production

The rate of ion-electron pair production q can be represented as a product of four contributing terms (Hargreaves, 1991, [29])

$$q = \eta\sigma nI \quad (2.6)$$

I is the intensity of ionization radiation at some level of the atmosphere, n is the density of atoms or molecules capable of being photo-ionized, σ is the absorption cross-section and η is the ionization efficiency. Equation (2.6) expresses the so called Chapman production function. To analyze this function it is assumed that the atmosphere is composed of a single species whose height distribution is exponential with constant scale height. There is no variation in the horizontal plane and the absorption of solar radiation is proportional to the concentration of gas particles. Another assumption is that the absorption coefficient is constant (i.e. monochromatic radiation). So the Chapman production function can be written as

$$q = q_{m0} \exp(1 - z - \sec \chi \exp(-z)) \quad (2.7)$$

where z is the reduced height for the neutral gas, $z = (h - h_{m0})/H$, H is the scale height, χ is the solar zenith angle, h_{m0} is the height of maximum production rate (for $\chi = 0$) and q_{m0} is the production at h_{m0} and can be derived by $q_{m0} = \eta\sigma n_{m0}$, using the density of ionizable atoms at level of maximum ion production n_{m0} . If equation (2.7) is expressed as

$$\frac{q}{q_{m0}} = \exp(1) \exp(-z) \exp(-\sec \chi \exp(-z)) \quad (2.8)$$

then the first term is constant, the second expresses the height variation of ionizable atoms and the last term expresses the intensity of the ionizing radiation. The Chapman function plays an important role in theoretical models as it is the base of ionospheric formation and of radiation absorption in an exponential atmosphere. The real ionospheric behavior is much more complicated, but the Chapman theory gives a good reference to start with. The profiles represented in figure 1 are in fact diffused lines, since a real atmosphere consists of more than one ionizable gas and the ionizing radiation is composite so the resulting production function $q(z)$ is the sum of the functions generated by the separate ionizing processes. It can be shown that even in that case the peak of the production rate q_m is proportional to $\cos \chi$, as for a single gas. The next subsection will deal with processes involved in the loss of free electrons. Generally electrons can disappear from a given volume for two reasons, either being removed by recombination or being dragged off by transport processes.

2.3.2 Recombination

There are two kinds of recombination: recombination between electrons and positive ions and the recombination by attachment. In the first case it is assumed that there are no negative ions and that electrons recombine with positive ions only:



X^+ is an ion and e^- is an electron. The rate of loss can be defined then by

$$L = \alpha [X^+] N \quad (2.10)$$

using the recombination coefficient α and the ion density $[X^+]$. If it is assumed that the ion density is equal to the electron density, L can be rewritten as

$$L = \alpha N^2. \quad (2.11)$$

At equilibrium (equation (2.5)) production rate and loss rate are of the same size, which gives

$$q = L = \alpha N^2. \quad (2.12)$$

Taking now the production rate q from the Chapman production function (2.7) yields

$$N = N_{m0} \exp \left[\frac{1}{2} (1 - z - \sec \chi \exp(-z)) \right] \quad N_{m0} = \sqrt{\frac{q_{m0}}{\alpha}}. \quad (2.13)$$

The region where the electron density follows this behavior is called α -Chapman layer. In the second case, when the electron loss is produced by attachment between electrons and neutral particles and negative ions are obtained. The attachment type of this reaction can be written as



taking the neutral molecule M and the electron e . In such a case the concentration of neutral molecules or atoms is assumed constant because they can be found in large number, so removing a few of these species does not have an important effect on the total remaining. The rate of electron loss can be stated by

$$L = \underbrace{b[M]}_{\beta} N = \beta N \quad (2.15)$$

using the attachment coefficient β . Using the equilibrium condition

$$q = \beta N \quad (2.16)$$

and relating it to the Chapman production function (2.7) results, in an analog way as in the previous case, in the electron density

$$N = N_{m0} \exp(1 - z - \sec \chi \exp(-z)) \quad N_{m0} = \frac{q_{m0}}{\beta}. \quad (2.17)$$

The region where the electron density follows this behavior is called β -Chapman layer. However, the concentration of neutral molecules varies with height, so the β coefficient varies with z (Hargreaves, 1992, [29]). In the F layer the electron loss occurs in two steps. First, a charge exchange between the positive ion and a molecule ($A^+ + XY \rightarrow A + XY^+$) takes place, then the recombination of an electron with the charged molecule by a process that results in dissociating it into the atoms X and Y (these atoms may be left excited, $(XY^+ + e^- \rightarrow X' + Y')$) happens. The rate of the first process is $\beta[A^+]$ and of the second one is $\alpha[XY^+]$. At low altitudes β is large, the first process goes quickly and the rate is governed by an α -Chapman process. At higher altitudes (more than about 160 kilometers) β is small and the rate is governed by the β -Chapman process. (Ratcliffe, 1972, [56])

2.3.3 Vertical transport

The last term of the continuity equation (2.5) represents the transport processes inside the ionosphere. It takes into account the change of electron and ion density at a given location due to the volume movement of the plasma. If production and loss processes are removed, and only vertical movement is taken, the continuity equation can be rewritten as

$$\frac{\partial N}{\partial t} = -\nabla \cdot (N\vec{v}) = -\frac{\partial (Nw)}{\partial h} \quad (2.18)$$

whereas w is the vertical velocity of the plasma. Assuming that the drift is caused by diffusion, then the velocity is proportional to the pressure gradient (more details in Hargreaves (1992, [29])) and the vertical velocity can be denoted by

$$w = -\frac{D}{N} \frac{\partial N}{\partial h} \quad (2.19)$$

using the diffusion coefficient $D = kT/(m\nu)$ containing the collision frequency ν and the Boltzmann constant k and where the temperature T and the particle mass m are introduced. The expression for

D was derived by equating the driving force due to the pressure gradient to the drag force due to collisions as a minority gas. In the given case the minority gas is the plasma composed of ions and electrons, and the majority gas is the neutral air. The ions are heavier and tend to drift away from the electrons, which evokes electric forces from the field proportional to eE on the the charged particles. Therefore the drift of the plasma for each species (subscript e for electrons, subscript i for ions) can be denoted (Hargreaves, 1992, [29]).

$$\begin{aligned} -\frac{dp_e}{dh} &= N_e m_e g + eEN + N_e m_e \nu_e w_e = -kT_e \frac{dN_e}{dh} \\ -\frac{dp_i}{dh} &= N_i m_i g + eEN + N_i m_i \nu_i w_i = -kT_i \frac{dN_i}{dh} \end{aligned} \quad (2.20)$$

whereas g is the acceleration due to gravitational forces. Adding both components and assuming $m_e \ll m_i$, $m_e \nu_e \ll m_i \nu_i$, $N_e = N_i = N$ and $w_e = w_i = w$ gives

$$Nm_i g + Nm_i \nu_i w = -k(T_i + T_e) \frac{dN}{dh}. \quad (2.21)$$

Hence for the plasma

$$Nw = D_p \left[\frac{dN}{dh} + \frac{N}{H_p} \right] \quad (2.22)$$

is obtained whereas $D_p = k(T_e + T_i)/(m_i \nu_i)$ is the plasma diffusion coefficient and $H_p = k(T_e + T_i)/(m_i g)$ is the plasma scale height. Therefore the continuity equation can be written as

$$\frac{dN}{dt} = \frac{\partial}{\partial h} \left[D_p \left(\frac{dN}{dh} + \frac{N}{H_p} \right) \right]. \quad (2.23)$$

At equilibrium $\frac{dN}{dt} = 0$ and using $T_e = T_i$ in equation (2.23), the electron density can be represented as a Chapman layer but with about twice the scale height (Hargreaves, 1992, [29]). If the plasma is not in equilibrium, the electron density distribution depends on time, at a rate depending on the value of the plasma diffusion coefficient. It is inversely proportional to the collision frequency and therefore increases with the altitude. If H is the scale height of the neutral gas, D can be written as

$$D = D_0 \exp \left(\frac{h - h_0}{H} \right). \quad (2.24)$$

Diffusion becomes more important than photochemistry at higher altitudes, which has to be taken into account when developing theoretical models for this region.

2.4 Models of the ionosphere electron density

There are numerous models that have been developed incorporating several techniques of ionospheric measurements. Four of them will be described here briefly, the first being the Chiu model, as it is one of the oldest phenomenological models that uses combinations of Chapman functions to model the electron density profiles. Thereafter the Bent model, which is often quoted in literature and which appears to be the father of such models, will be presented. Later the IRI model, created from an international cooperation between ionospheric scientists under the umbrella of the Committee on Space Research (COSPAR) and the International Union of Radio Science (URSI) and the PIM model, developed by the U.S. Air Force Research Laboratory, will be described.

2.4.1 Chiu ionospheric model

This global phenomenological model describes the large scale variations of ionospheric electron density with local time, latitude, and solar sunspot number (Ching and Chiu, 1973, [12]). It is based on ionosonde data from 50 stations spanning the period 1957 to 1970. The model profile is obtained as the sum of three Modified Chapman functions for E, F1, and F2 layers. The model was improved by Chiu (1975, [13]) and served as the starting point for the FAIM model. The model is fairly simple, using less than 50 coefficients, which limits its application for equatorial and higher latitudes. It is, however, fast and easily manipulable and a good choice for first-order estimates. An extension for the polar cap ionosphere is being constructed.

2.4.2 Bent ionospheric model

The model describes the ionospheric electron density as a function of latitude, longitude, time, season, and solar radio flux. The topside is represented by a parabola and three exponential profile segments, and the bottom side by a bi-parabola. The model is based on about 50,000 Alouette topside ionograms (1962-1966), 6,000 Ariel 3 in situ measurements (1967- 1968), and 400,000 bottom side ionograms (1962-1969). For the F2-peak the CCIR (International Radio Consultative Committee) maps are used. The model has been widely used for ionospheric refraction corrections in satellite tracking. It does not include the lower layers (D, E, F1) and uses a simple quadratic relationship between CCIR's M(3000)F2 factor and the height of the F2-peak. A comparison between the Bent model and the IRI model and their application for satellite orbit determination was discussed by Bilitza et al. (1988, [7]). IRI showed better results because of the more detailed representation of the bottom side density structure.

2.4.3 International Reference Ionosphere

The International Reference Ionosphere (IRI) is an international project sponsored by the Committee on Space Research (COSPAR) and the International Union of Radio Science (URSI). These organizations formed a Working Group in the late sixties to produce an empirical standard model of the ionosphere based on all available data sources. Several steadily improved editions of the model have been released. For given location, time and date, IRI describes the electron density, electron temperature, ion temperature, and ion composition in the altitude range from about 50 km to about 2000 km; and also the electron content. It provides monthly averages in the non-auroral ionosphere for magnetically quiet conditions. The major data sources are the worldwide network of ionosondes, the powerful incoherent scatter radars (Jicamarca, Arecibo, Millstone Hill, Malvern, St. Santin), the ISIS and Alouette topside sounders, and in situ instruments on several satellites and rockets. IRI is updated every year during special IRI Workshops (e.g., during COSPAR general assembly). More information can be found in the workshop reports. Several extensions are planned, including models for the ion drift, description of the auroral and polar ionosphere, and consideration of magnetic storm effects (Bilitza, 2001, [6]).

2.4.4 Parameterized Ionospheric Model

The Parameterized Ionospheric Model (PIM) is a fast global ionospheric and plasmaspheric model based on a combination of the parameterized output of several regional theoretical ionosphere models and an empirical plasmaspheric model. It represents the climatological portion of the Parameterized Real-time Ionospheric Specification Model (PRISM), currently the operational model for specifying the current state of the ionospheric weather. The development of PRISM and PIM was funded by the Battlespace Environment Division of the Air Force Research Laboratory (AFRL). PIM consists of portable FORTRAN source code and a large database of coefficients for an orthogonal function expansion. For user specified geophysical conditions and spatial coordinates, PIM produces electron density profiles (EDPs) between 90 and 25000 km altitude, corresponding critical frequencies and heights for the ionospheric E and F2 regions, and Total Electron Content (TEC). The ionospheric portion of PIM is a parameterization of the results of several regional theoretical ionospheric models. This allows PIM to be computationally fast while retaining the physics of the theoretical ionospheric models. The parameterization compresses the output from the theoretical ionospheric models while preserving important characteristics such as density peaks and scale heights. The large base of data used by PIM contains coefficients from the parameterization. This constitutes PIM as described in Daniell et al. (1995, [17]).

2.5 Radio waves in a magnetic plasma

The ionosphere, i.e. a magnetic plasma, is an anisotropic and birefringent medium (being shown later), which makes the usage of tensor notation for conductivity $\hat{\sigma}$ (as the electric field strength vectors and current density vectors don't point into the same direction) necessary. The dielectric coefficient has

to be treated with a tensor $\hat{\epsilon}$ in the same way. Dealing only with the “electric” components three frequencies have to be taken into account,

- the plasma frequency for electrons ω_P (Demtröder, 1999, [18])

$$\omega_P^2 = \frac{e^2 N_e}{m_e \epsilon_0}, \quad (2.25)$$

- the electron gyro (or synchrotron) frequency ω_g

$$\omega_g = \frac{e}{m_e} \vec{B}_{earth}, \quad (2.26)$$

- and the collision frequency of the electrons ν ,

whereas e and m_e are respective charge and mass of the electron, N_e is the free electron density, \vec{B}_{earth} is the Earth’s magnetic field vector and ϵ_0 is the vacuum permeability. Generally the propagation of electromagnetic waves can be explained by Maxwell’s equations in the form

$$\nabla \times \vec{H} = \frac{\partial \vec{D}}{\partial t} = \epsilon_0 \frac{\partial \vec{E}}{\partial t} + \vec{i} \quad (2.27)$$

$$\nabla \times \vec{E} = -\frac{\partial \vec{B}}{\partial t} \quad (2.28)$$

$$\nabla \cdot \vec{D} = \rho_e \quad (2.29)$$

$$\nabla \cdot \vec{B} = 0 \quad (2.30)$$

where the following notation was used

- \vec{H} : magnetic field strength (of the wave),
- \vec{B} : flux density (of the wave),
- \vec{D} : electric displacement,
- \vec{E} : electric field,
- \vec{i} : current density,
- ρ_e : free charge density.

The motion of electrons can be described by

$$m_e \frac{d\vec{v}}{dt} = e(\vec{E} + \vec{v} \times \vec{B}) \quad (2.31)$$

$$\vec{i} = e\vec{v}N_e \quad (2.32)$$

Thereafter it is useful to introduce the frequency ω , which takes time-dependent variations in the form of $\exp(-j\omega t)$ with $j = \sqrt{-1}$ into account. With the abbreviations

$$X = \frac{\omega_P^2}{\omega^2} \quad Y = \frac{\omega_g}{\omega} \quad Z = \frac{\nu}{\omega} \quad (2.33)$$

or

$$\tilde{X} = \frac{X}{1 + jZ} \quad \tilde{Y} = \frac{Y}{1 + jZ} \quad (2.34)$$

the tensor of conductivity $\hat{\sigma}$ (in a Cartesian coordinate system, z-axis in direction of \vec{B}_{earth}) can be found (Budden, 1988, [11]) by

$$\hat{\sigma} = j\omega\epsilon_0 \begin{pmatrix} \tilde{X} & j\tilde{X}\tilde{Y} & 0 \\ \frac{1 - \tilde{Y}^2}{j\tilde{X}\tilde{Y}} & \frac{1 - \tilde{Y}^2}{\tilde{X}} & 0 \\ -\frac{j\tilde{X}\tilde{Y}}{1 - \tilde{Y}^2} & \frac{\tilde{X}}{1 - \tilde{Y}^2} & \tilde{X} \end{pmatrix}. \quad (2.35)$$

Moreover the relations $\vec{B} = \mu_0 \mu \vec{H}$ with $\mu = 1$ (μ_0 : permeability) and $\rho_e = 0$ (the plasma can be treated neutral if length of scale is bigger than the Debye-length¹ (Budden, 1988, [11])) are valid. Together with the Ohm's law $\vec{i} = \hat{\sigma} \vec{E}$, one can write

$$\frac{\partial \vec{D}}{\partial t} = \varepsilon_0 \frac{\partial \vec{E}}{\partial t} + \vec{i} = \varepsilon_0 \frac{\partial \vec{E}}{\partial t} + \hat{\sigma} \vec{E}. \quad (2.36)$$

As \vec{E} is already dependent on time the expression $\frac{\partial}{\partial t}$ can be replaced by $-j\omega$, which simplifies equation (2.36) to

$$\vec{D} = \varepsilon_0 \left(1 - \frac{1}{j\omega\varepsilon_0} \hat{\sigma} \right) \vec{E} = \varepsilon \hat{\varepsilon} \vec{E} \quad (2.37)$$

and the dielectric tensor can be denoted in the following way (using the unity tensor \hat{U}).

$$\hat{\varepsilon} = \left(\hat{U} - \frac{1}{j\omega\varepsilon_0} \hat{\sigma} \right) = \begin{pmatrix} 1 - \frac{\tilde{X}}{1 - \tilde{Y}^2} & \frac{j\tilde{X}\tilde{Y}}{1 - \tilde{Y}^2} & 0 \\ -\frac{j\tilde{X}\tilde{Y}}{1 - \tilde{Y}^2} & 1 - \frac{\tilde{X}}{1 - \tilde{Y}^2} & 0 \\ 0 & 0 & 1 - \tilde{X} \end{pmatrix} \quad (2.38)$$

Therefore Maxwell's equations can be rewritten

$$\nabla \times \vec{H} = -j\omega\varepsilon_0 \hat{\varepsilon} \vec{E} \quad \nabla \times \vec{E} = j\omega\mu_0 \vec{H} \quad \nabla \cdot (\varepsilon_0 \hat{\varepsilon} \vec{E}) = 0 \quad \nabla \cdot (\mu_0 \vec{H}) = 0 \quad (2.39)$$

and the case of planar waves in the form $\vec{F} = \vec{F}_0 \exp(j\vec{k}\vec{r} - j\omega t)$ is investigated further. The position vector is represented by \vec{r} , \vec{k} points into propagation direction and the absolute value of its real part is called wave number and related to the wavelength λ in the following way.

$$|\mathcal{R}(\vec{k})| = \frac{2\pi}{\lambda} \quad (2.40)$$

The nabla operator ∇ can be easily replaced now by the expression $j\vec{k}$ as \vec{F}_0 is constant in space and time for planar waves. Taking the first and second Maxwell equation from (2.39) gives

$$j\vec{k} \times \vec{H} = -j\omega\varepsilon_0 \hat{\varepsilon} \vec{E} \quad j\vec{k} \times \vec{E} = j\omega\mu_0 \vec{H}. \quad (2.41)$$

Division by j of both equation and multiplying the second one by $\vec{k} \times$ gives $\vec{k} \times (\vec{k} \times \vec{E}) = \omega\mu_0 (\vec{k} \times \vec{H})$. Now the first equation of (2.41) can be substituted on the right side and

$$\vec{k} \times (\vec{k} \times \vec{E}) = -\omega^2 \varepsilon_0 \mu_0 \hat{\varepsilon} \vec{E} \quad (2.42)$$

is obtained. Expression (2.42) represents an equation system for the three components of \vec{E} . Solutions can only exist (as this system is homogeneous) if the corresponding determinant equals to zero. For practical reasons the index vector $\vec{n} = (c/\omega)\vec{k}$ ($c = 1/\sqrt{\varepsilon_0\mu_0}$ speed of light in vacuum) is introduced. Simplifying equation (2.42) to

$$\vec{n} \times (\vec{n} \times \vec{E}) = -\hat{\varepsilon} \vec{E} \quad (2.43)$$

and applying the cross product twice gives $(\vec{n} \cdot \vec{E})\vec{n} - (\vec{n} \cdot \vec{n})\vec{E} = -\hat{\varepsilon} \vec{E}$. Rearranging the left side to $(n^2 - \vec{n}\vec{n}\cdot)\vec{E} = -\hat{\varepsilon} \vec{E}$ states the problem now very clearly. The expression $(\vec{n}\vec{n}\cdot)$ is an operator and can be treated as a tensor (in cartesian coordinates).

$$\hat{Q} = (\vec{n}\vec{n}\cdot) = \begin{pmatrix} n_x n_x & n_x n_y & n_x n_z \\ n_y n_x & n_y n_y & n_y n_z \\ n_z n_x & n_z n_y & n_z n_z \end{pmatrix} \quad (2.44)$$

It can be assumed now, without restriction of generality, that the z-axis of the cartesian coordinate system coincides with the Earth's magnetic field vector \vec{B}_{earth} and that the y-axis is in the plane,

¹The Debye length is given by $l_D = (\varepsilon_0 k T / N e^2)^{\frac{1}{2}}$, (Budden, 1988, [11])

spanned by the vectors \vec{k} and \vec{B}_{earth} . The components of the index vector can be written now explicitly in the form

$$n_x = 0, \quad n_y = n \sin \Theta, \quad n_z = n \cos \Theta \quad (2.45)$$

using the angle Θ between \vec{B}_{earth} and \vec{k} . The absolute value of n is defined by $n = \sqrt{\vec{n} \cdot \vec{n}}$, but one has to be aware that is a still complex scalar value. Equation (2.44) can be rewritten now, using expressions (2.45).

$$\hat{Q} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & n^2 \sin^2 \Theta & n^2 \sin \Theta \cos \Theta \\ 0 & n^2 \sin \Theta \cos \Theta & n^2 \cos^2 \Theta \end{pmatrix} \quad (2.46)$$

which enables reformulation of (2.43) by the tensor relation $(n^2 \hat{U} - \hat{Q} - \hat{\epsilon}) \vec{E} = 0$, which defines the polarization relations $(E_x : E_y : E_z)$ for the electric field, using the Cartesian coordinate system defined above. If the determinant of the bracket expression equals zero a solution of equation (2.46) will exist.

$$\left| n^2 \hat{U} - \hat{Q} - \hat{\epsilon} \right| = \begin{vmatrix} n^2 - 1 + \frac{\tilde{X}}{1 - \tilde{Y}^2} & \frac{j\tilde{X}\tilde{Y}}{1 - \tilde{Y}^2} & 0 \\ -\frac{j\tilde{X}\tilde{Y}}{1 - \tilde{Y}^2} & n^2 \cos^2 \Theta - 1 + \frac{\tilde{X}}{1 - \tilde{Y}^2} & -n^2 \sin \Theta \cos \Theta \\ 0 & -n^2 \sin \Theta \cos \Theta & n^2 \sin^2 \Theta - 1 + \tilde{X} \end{vmatrix} = 0 \quad (2.47)$$

multiplying the third line by $\cos \Theta / \sin \Theta$ and adding it to the second line gives

$$\begin{vmatrix} n^2 - 1 + \frac{\tilde{X}}{1 - \tilde{Y}^2} & \frac{j\tilde{X}\tilde{Y}}{1 - \tilde{Y}^2} & 0 \\ -\frac{j\tilde{X}\tilde{Y}}{1 - \tilde{Y}^2} & -1 + \frac{\tilde{X}}{1 - \tilde{Y}^2} & -(1 - \tilde{X}) \cot \Theta \\ 0 & -n^2 \sin \Theta \cos \Theta & n^2 \sin^2 \Theta - 1 + \tilde{X} \end{vmatrix} = 0. \quad (2.48)$$

Because equation (2.48) has two solutions for n^2 , also $(n^2 \hat{U} - \hat{Q} - \hat{\epsilon}) \vec{E} = 0$ has two solutions. This means that the medium is birefringent for electromagnetic waves. Finding now the two solutions for n^2 leads to the dispersion formula, also called Appleton-Hartree equation.

$$n^2 = 1 - \frac{\tilde{X}(1 - \tilde{X})}{1 - \tilde{X} - \tilde{Y}_T^2/2 \pm \sqrt{\tilde{Y}_T^4/4 + \tilde{Y}_L^2(1 - \tilde{X})^2}} \quad (2.49)$$

Two new expressions were introduced in equation (2.49) by splitting \tilde{Y} into two components. The longitudinal component $\tilde{Y}_L = \tilde{Y} \cos \Theta$ and the transversal component $\tilde{Y}_T = \tilde{Y} \sin \Theta$ account for the angle between the propagation direction and the Earth's magnetic field. The index vector can be split into real and imaginary parts

$$\vec{n} = \mathcal{R}(\vec{n}) + j\mathcal{I}(\vec{n}) = \vec{\mu} + j\vec{\chi} \quad (2.50)$$

whereas $|\vec{\mu}| = \mu$ represents the index of refractivity and $|\vec{\chi}| = \chi$ can be seen as a (constant) damping factor, decreasing the energy of the propagated waves. Phase velocity, i.e. the propagation speed of surfaces with constant phase, can be denoted now.

$$v = \frac{c}{\mu} = \frac{\omega}{k} \quad (2.51)$$

Equation (2.49) reveals the following properties of a magnetic plasma (the ionosphere can be treated as such a medium).

- Dispersive: the index of refraction depends on the used frequency and it can be shown that the group velocity differs from phase velocity.
- Absorptive: the index of refraction is a complex number and the imaginary part χ , called extinction coefficient, describes the energy absorption. This process is dissipative as wave energy is converted into heat through collision processes.

- Birefringent: the index of refraction has two distinct values, which suggests the possibility of two ray paths; each one characterized by different phase and group velocities.
- Anisotropic: each of the two indices of refraction is a separate function of the orientation of the normal to the surface of constant wave phase with respect to the background (uniform) magnetic field.

One of the biggest challenges for any system operating in the radio frequency band is the determination of the propagation velocity of its signal. If such waves propagate through vacuum, the traveled distance would just be the product of speed of light in vacuum with the propagation time between sender and receiver. When signals travel through magnetic plasma like the ionosphere phase propagation speed however, is modified according to equation (2.51). A signal (i.e. a modulated carrier wave) can be treated like a superimposition of a group of different wavelength waves centered at the frequency of the carrier. In a dispersive medium the phases propagate with a different velocity, depending on the frequency. The group velocity v_g is then given by

$$v_g = \frac{c}{\mu_g} \quad (2.52)$$

where μ_g is the refraction index for the group. It can be related to the refraction index in the following way.

$$\mu_g = c \frac{\partial k}{\partial \omega} = \frac{\partial}{\partial \omega} (\omega \mu) = \mu + \omega \frac{\partial \mu}{\partial \omega} \quad (2.53)$$

Finally μ and μ_g can be related to each other starting with

$$\mu \frac{\partial \mu}{\partial \omega} = \frac{1}{2} \frac{\partial(\mu^2)}{\partial \omega} = \frac{1}{2} \frac{\partial(\mu^2)}{\partial(\omega^2)} \frac{\partial(\omega^2)}{\partial \omega} = \omega \frac{\partial(\mu^2)}{\partial(\omega^2)} \quad (2.54)$$

which leads to

$$\mu \mu_g = \mu^2 + \omega \mu \frac{\partial \mu}{\partial \omega} = \mu^2 + \omega^2 \frac{\partial(\mu^2)}{\partial(\omega^2)} = \frac{\partial(\omega^2 \mu^2)}{\partial(\omega^2)}. \quad (2.55)$$

Equation (2.53) can be evaluated together with the Appleton-Hartree expression (2.49) and the relation (2.50) in order to obtain the refractive index for group propagation within a magnetic plasma. This step will be carried out later in section 4.1.2.

2.5.1 Wave propagation in a slightly inhomogeneous medium

Waves can only propagate if the real part of $n^2 > 0$ and therefore $\mu^2 > 0$. Contributions due to collisions Z can be neglected in an isotropic medium (i.e. a plasma without an outer magnetic field). Taking the most simple form of the dispersion relation (2.49) gives (when all fundamental constants are absorbed in the parameter A)

$$\mu^2 = 1 - X = 1 - \frac{\omega_P^2}{\omega^2} = 1 - \frac{f_P^2}{f^2} = 1 - A \frac{N_e}{f^2} \quad (2.56)$$

which means that waves can only propagate if $f > f_P$. Sending out radio waves at a given frequency from a ground station will help to determine the plasma frequency and therefore the electron density. It is assumed that the electron density N_e grows with height and the plasma frequency grows proportionally to $\sqrt{N_e}$. Waves can only propagate up to a certain region where $f_p = f$ is given. It can be shown that at this point total reflection of the wave will happen. The condition for reflection $X = 1$ is valid for the so called ordinary component of the wave but theoretical considerations (only neglecting Z) have shown that a second component, called extra-ordinary, will be reflected under the condition $(1 - X^2)^2 = Y^2$ in the same way. The point of reflection can be treated as being surrounded by a very inhomogeneous medium (n^2 tends to zero very fast), but the main part of the propagation medium can be taken as a slightly inhomogeneous medium only. The approximation holds if the index of refraction μ and the damping constant χ change only slightly (maximum of about 10%) within one

wavelength. This simplifies further calculations as propagation can be approximated by geometrical optics. So any characteristics of the radio wave can be described, using the Eikonal Ansatz.

$$\vec{F}(\vec{r}) = \vec{G}_0 \frac{S_0}{S} \exp \left\{ j \left[\left(\int_{\vec{r}_0}^{\vec{r}} \mu ds \right) - \omega t \right] \right\} \quad \text{using the approximation} \quad S = \int_{\vec{r}_0}^{\vec{r}} ds \quad (2.57)$$

In above equation S symbolizes the length of the ray path, reaching from the source at \vec{r}_0 to any given place \vec{r} and S_0 is a reference value which determines geometric thinning. The ray path in a slightly inhomogeneous medium can be determined by taking the approximation from geometric optics and its law of refraction, but anisotropy of the medium has to be accounted for by treating the ray and the normal to the wavefront separately. The law of refraction (Snellius' law) is valid for the normal to the wavefront and can be written by $\mu \sin \alpha = \text{const}$. Continuous refraction is obtained, but reflection cannot be explained by this approach. A simple reason for this property can be found as the reflected parts annihilate themselves due to interference. Partial reflections happen if the conditions for geometrical optics are violated, i.e. locations with big gradients of the refraction index. If the change of μ on a distance $\Delta s \ll \lambda$ is strong (but nearly constant) the electromagnetic wave will be affected by this discontinuity, but this case can be treated as a piecewise inhomogeneous medium with (mathematical) joint faces, separating the different regions. At these joint faces fraction and (partial) reflection happen and amplitude changes can be dealt with the Fresnel equations (continuity of the tangential component within the joint face).

If the transmitted frequency is smaller than the maximum useful frequency (MUF), radio waves sent out at slant directions will be refracted continuously until they reach ground again. This phenomenon is called "slant reflection", but has physically nothing to do with reflection. This property can be used to transmit radio waves over huge (intercontinental) distances without direct inter-visibility between transmitter and receiver.

2.5.2 Transionospheric wave propagation

If the transmitted frequency is much bigger than the maximum plasma frequency of the ionosphere (i.e. $f \gg f_{P,max}$) radio waves will propagate through the ionosphere. With the evolution of artificial satellites and the observation of extra terrestrial radio sources, effects caused by penetration of the ionosphere had to be accounted for. For transmitted frequencies $f > 100$ MHz the quasi-longitudinal approximation can be taken and damping of the wave can be neglected in almost all cases. This means that only

$$\mu^2 = 1 - \frac{X}{1 \pm Y \cos \Theta} \quad (2.58)$$

represents the refraction index, depending on X , Y and the angle between the direction of the Earth's magnetic field and the propagation direction. Taking the square root of equation (2.58) and applying a Taylor approximation up to first order gives

$$\mu \approx 1 - \frac{1}{2} \frac{f_P^2}{f^2 \pm f f_g} = 1 - \frac{A}{2} \frac{N_e}{f^2 \pm f f_g}. \quad (2.59)$$

Only if transionospheric propagation influences of higher order have an impact, the Taylor expansion of the square root of equation (2.58) has to be extended to higher terms. For the estimation of the first order propagation effects on the propagation of phase and modulated signals on frequencies $f > 100$ MHz caused due to the Earth's magnetic field will be neglected. As the highest synchrotron frequency of the electrons is about 1.7 MHz the ratio f_g/f can be neglected, which gives the "isotropic high-frequency approximation" $\mu \doteq 1 - (1/2)X$. The assumption of the neglect of higher order terms also has the advantage that the ray path between the transmitter and the receiver can be treated as a straight line. Therefore the phase φ of a signal at frequency f can be determined analytically.

$$\varphi = \frac{2\pi f}{c} \int_S^R \left(1 - \frac{A N_e}{2 f^2} \right) ds_0 - 2\pi f t + \psi = \frac{2\pi f}{c} \left[\rho - \frac{A}{2 f^2} \int_S^R N_e ds_0 \right] - 2\pi f t + \psi \quad (2.60)$$

where S and R are the phase centers of the transmitting and receiving antennas, ds_0 is the path element of the ray \overline{SR} , ψ is an arbitrary phase constant and ρ is the geometric distance \overline{SR} . A value for A in S.I units can be found easily as $A = F_p^2/N_e = e^2/(2\pi m_e \epsilon_0) = 80.6$. From equation (2.60) it can be seen that the ionospheric influence is proportional to the integrated electron density. The integral will be given its own abbreviation and unit because of its importance. The slant total electron content STEC is defined by

$$STEC = \int_S^R N_e ds_0 \quad (2.61)$$

and denoted in a so called total electron content unit (TECU), where 1 TECU is defined by 10^{16} electrons $/m^2$. This means that 1 TECU is equal to column with a cross section of $1 m^2$ along the ray path in which 10^{16} free electrons can be found. The influence of the ionosphere on complex signals (e.g. modulated signals) can be described by analyzing the sum to several phases combinations. Another important characteristic of the ionosphere causes a phase shift between the left and right circularly polarized components. This phenomenon is called Faraday effect and can be described, by taking equation (2.59).

$$\begin{aligned} \varphi_l - \varphi_r &= \frac{2\pi A}{2cf^2} \left[\int_S^R \left(-\frac{N_e}{1 + Y \cos \Theta} + \frac{N_e}{1 - Y \cos \Theta} \right) ds_0 \right] \\ &= \frac{2\pi A}{2cf^2} \int_S^R \frac{2Y \cos \Theta}{1 - Y^2 \cos^2 \Theta} N_e ds_0 \approx \frac{2\pi A}{cf^2} \int_S^R Y \cos \Theta N_e ds_0 \end{aligned} \quad (2.62)$$

Now Y can be re-substituted and

$$\varphi_l - \varphi_r = \frac{2\pi A}{cf^2} \frac{e}{2\pi m_e f} \int_S^R B \cos \Theta N_e ds_0 = \frac{Ae}{m_e cf^2} \int_S^R B \cos \Theta N_e ds_0 \quad (2.63)$$

is obtained.

3 Very Long Baseline Interferometry

3.1 Introduction

Until a few decades ago man’s knowledge of the universe outside the Earth came almost entirely from optical-astronomy observations. Beginning millenniums ago with purely visual techniques, astronomy made rapid advances after the invention of the optical telescope in the early seventeenth century and the application of photographic methods in the last century. All observations were in the visible part of the electromagnetic spectrum in a band about one octave wide. During the last seven decades astronomical observations at radio wavelengths have created a new branch of astronomy called radio astronomy. The older astronomy in the visible spectrum is now often called optical astronomy to distinguish it from the newer branch. The positions of optical and radio astronomy in the electromagnetic spectrum coincide with the two principal transparent bands of the Earth’s atmosphere and ionosphere (figure 8). These transparent bands are commonly referred to as the optical and radio windows. The optical window extends from about 0.4 to 0.8 micron (1 octave) while the much broader radio window extends from about 1 cm to 10 m (about 10 octaves). The values of 1 cm and 10 m are nominal and arbitrary. Because of some relatively transparent bands in the millimeter region and occasional ionospheric “holes” at decimeter wavelengths, more extreme limits of the radio window may be placed at 1 mm and 150 m (i.e. about 17 octaves). The short-wavelength limit is a function of the atmospheric composition, cloud cover, etc., while the long-wavelength limit depends on the electron density in the ionosphere (see section 2.5). The methods of radio astronomer often appear to be quite

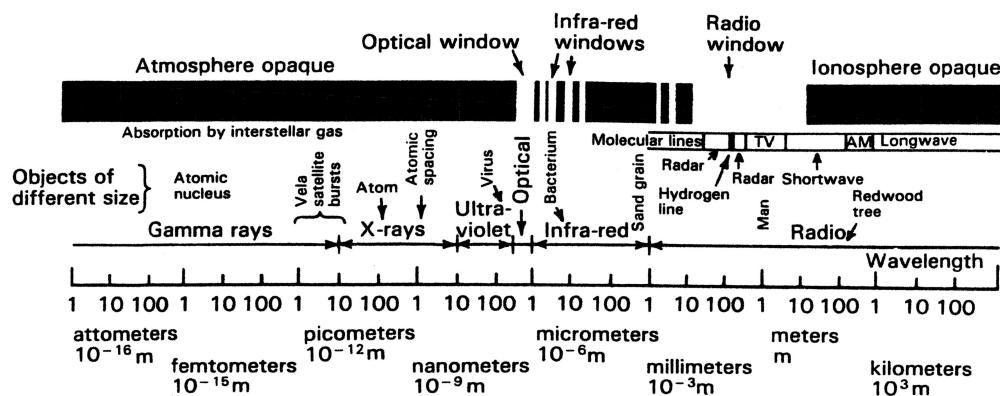


Figure 8: Electromagnetic spectrum showing the relative transparency of the Earth’s atmosphere and ionosphere (Kraus, 1986, [47]).

different from those of the optical astronomer, but there is the same aim in all of astronomy, from the radio to the X-ray domain. Nature presents us with a distribution of brightness in the sky, and it is the task of the astronomer to deduce, from this brightness distribution of electromagnetic radiation, what the radiating sources are, and what physical processes are acting. The distinguishing feature of a radio telescope is that the radiation energy gathered by the instrument is not measured immediately, a process known as detection in radio terminology. Instead, the radiation is amplified and manipulated coherently, preserving its wave-like character, before it is finally detected. The instrumental goals of the radio astronomer - obtaining a larger collection area, greater angular resolution and more sensitive detectors - are otherwise the same as they are for all the other astronomical disciplines (after Kraus (1986, [47])). But also the interests of geophysicists and geodesists are directed towards radio astronomy. They have found out how to use this expensive technique to monitor Earth rotation and plate tectonics, long before global navigation satellite systems were launched. Radio telescopes are taken to link the rotating Earth to the celestial sky and to deduce Earth orientation parameters from the inertial system defined by radio sources. Other geophysical parameters, station coordinates and atmospheric condition numbers can be determined using this technique.

3.2 Basic principles of VLBI

Two antennas separated by any distance, pointing at the same radio source and collecting the signal in the same frequency bands can be seen as the simplest possible configuration for a system called Very Long Baseline Interferometry (VLBI). These sources emit radio waves which are propagating through the universe like an expanding sphere. When these wavefronts reach the Earth's surface, they arrive as a plane wave perpendicular to the antenna's pointing direction. The emitted signal consists of a continuous sequence of waves. Figure 9 takes only the arrival of a single wave front into account. First the reference antenna is reached and after a certain time delay the wave front strikes the second (remote) antenna. This time-of-arrival difference is called time delay and is the main measurement type for geodetic and geophysical applications. The following sections follow the text books by Takahashi et al. (2000, [73]), Kraus (1986, [47]), Whitney (1974, [81]) and Thompson et al. (2001, [75]). The wave fronts of radio signals are essentially planar in nature as the distance between

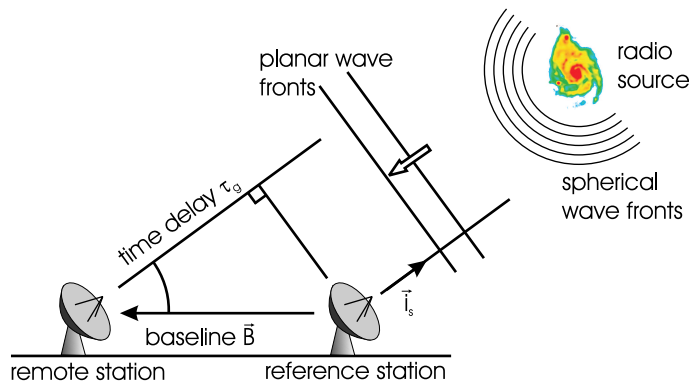


Figure 9: The principle of time delay.

the source and the Earth is usually large enough. This means that the time delay simply becomes one of the sides of a right triangle. The length of this side is found as the product of time delay and the speed of light. The angular direction to the source is usually known and the vector between the reference points of both antennas is called baseline \vec{B} . The time delay τ_g can be described using the pointing vector in radio source direction \vec{i}_s and the speed of light c .

$$\tau_g = -\frac{\vec{B} \cdot \vec{i}_s}{c} \quad (3.1)$$

Here the negative sign defines the delay in time-of-arrival at the remote station as a positive value. A baseline error $\Delta\vec{B}$ and a source direction error $\Delta\vec{i}_s$ cause a time delay difference $\Delta\tau_g$, which can be written as

$$\Delta\tau_g = -\frac{\Delta\vec{B} \cdot \vec{i}_s}{c} - \frac{\vec{B} \cdot \Delta\vec{i}_s}{c}. \quad (3.2)$$

Assuming that source directions are known at a certain accuracy level (i.e. $\Delta\vec{i}_s = 0$) it is possible to adjust $\Delta\vec{B}$ by observing several sources in different sky directions. Therefore VLBI can be taken to obtain precise baseline lengths over huge distances. If baseline length errors are sufficiently small (i.e. $\Delta\vec{B} = 0$) it will be possible to estimate source parameters. When the time delay is measured to an accuracy of about 50 ps, a conversion can be made to obtain the angular direction to the source with an accuracy of $0.015/|\vec{B}|$ (rad). Consequently the longer the baseline distance the greater the accuracy of the measured angular value, which corresponds to the angular resolution of the VLBI system. For a baseline length of 8000 km (roughly the distance between the U.S.A. and Japan), resolution becomes approximately 0.4 mas. With a single radio telescope angular resolution is generally defined as the inverse value of the antenna dish diameter. If two antennas are spaced several kilometers apart, however, the distance between them will become equivalent to the diameter of a single antenna, and resolution is increased accordingly. The development of VLBI systems has radically increased the ability of astronomers to increase the angular resolution by antenna spacing. Current VLBI systems feature large-scale antennas located in some cases on several continents. Space-based systems,

consisting of an orbiting telescope working in conjunction with an Earth-based station obtain even greater antenna spacing and consequently higher resolution. Thus, while VLBI gives us a very accurate instrument to measure the Earth's surface, Earth rotation, and other characteristics, it is also a powerful tool as a high resolution telescope with up to global scale diameter, which allows to study radio objects in the cosmos.

3.3 The VLBI system

The first radio antennas were connected by cables, but as antenna separation became too far for cable connection, high frequency oscillators (“clocks”) with stability characteristics permitting installation at each antenna site were taken for synchronization of the dislocated sites. The key to overcome certain limitations due to stability of the clocks came with the introduction of the atomic clock, which is essentially a frequency standard and can be taken to provide a stable signal in order to synchronize interferometer operations between the reference station and each remote station even of distances of several thousands of kilometers. Figure 10 shows the VLBI signal flow at each station. In the old days

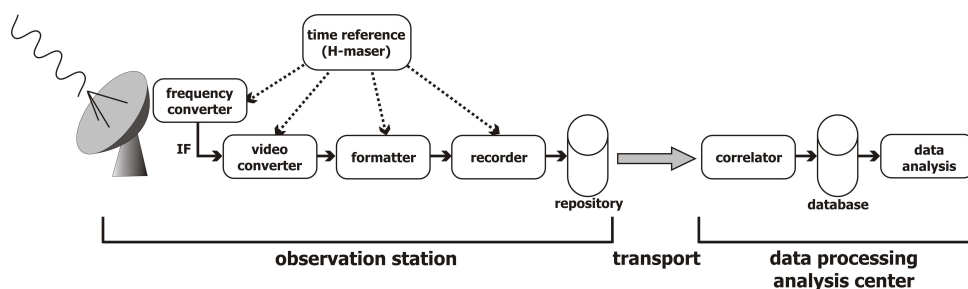


Figure 10: VLBI signal flow.

data were stored on magnetic tapes and shipped to correlators and processed further. Nowadays hard discs are used to record the data and are either transported by surface mail or sent over the Internet to the correlation center. The sensitivity of the system mainly depends on the diameters of the two antennas (D_1 , D_2) and the system noise temperature of the receivers (T_1 , T_2) and can be described by

$$\text{sensitivity} \propto \frac{D_1 D_2}{\sqrt{T_1 T_2}}. \quad (3.3)$$

Due to the fact that modern low-noise amplifiers (LNA) normally have noise temperature characteristics of several 10s of Kelvins, it is sufficient that the product of the antenna diameters is about 100 m^2 or more to be suitable for geodetic applications. Additionally the system should have the capability to accommodate observing frequencies over a wide bandwidth to compute group delay parameters with a sufficient accuracy.

3.3.1 Receiving systems

Generally radio signals (RF frequencies) caught by the antenna and amplified by the LNA are mixed with a local oscillator (LO) signal which is phase-locked to a frequency standard, and then are down-converted to an intermediate frequency (IF) signal, which lowers transmission losses efficiently as the signal passes through the coaxial cables. By using the local oscillator signal the coherency of phase information on the RF signal can be preserved during the down-conversion process, which becomes important in the data recovery procedure. New optical fiber cables enable transmission of signal frequencies above 1 GHz with low transmission losses.

In the next stage the IF signal is separated in several channels and down-converted to baseband frequencies (also called video frequencies, as conversion is done by video converters) ranging from 0 to several MHz. The image rejection mixer of the video converter filters the final signal into upper sideband (USB) and lower sideband (LSB) signals. As the local oscillator signals of the video converter are also phase-locked to the frequency standard signal coherency is also preserved within this stage. The video signal in each channel is then converted from an analog signal to a digital one by the A/D

converter, whereas the sampling frequency should be twice the video channel bandwidth. Usually the video channel frequency is 2 MHz, which gives a sampling frequency of 4 MHz. Like most system signals the sampling timing signal is also generated by the frequency standard. Digitizing is normally done in 1-, 2- or 4-bit quantization, whereas for most of the geodetic applications 1-bit sampling is used. Later the digitized signal is tagged with timing marks, and is sent to the formatter section of the data recorder as a time-series digital signal. The formatter prepares the signal for recording and sends it to the recorder (i.e. tape drive or hard disc). As mentioned earlier, precise independent frequency standards (“clocks”) at each station are crucial for the realization of a VLBI system. However, stability of this frequency standards plays an important role in maintaining the coherency of the observed signal through all conversion processes in the receiving system. To fulfill the high standards necessary for VLBI operations, hydrogen (H) masers are used. Generally stability characteristics are a function of the time scale, e.g. over a 100 second interval (i.e. the average scan length within a geodetic experiment) the stability of a hydrogen maser is better than 10^{-13} , which is referred to as short-term stability. Long-term stability on the other hand (more than 1 hour up to one day) becomes more important in data processing where interval approximations are used to compute the clock parameters. Stability of the H maser on this time scale is even better than $3 \cdot 10^{-15}$ (this means that the clock would take about 10 million years to change by one second). Additionally, in order to achieve higher precision in the group delay measurements, the signals of each channel are used for bandwidth synthesis (see section 3.3.3, page 31). Therefore each signal must contain phase information in order to account for differences that may evoke occur the different processing steps individually. Using this phase information (phase calibration (PCAL) signal) the additional phase between the channels can be corrected. The phase calibration signal consists of a series of tone signal having a frequency of 1 MHz and is injected at the front end of the low noise amplifier (LNA) directly at the antenna.

Data from the recorder have to be transported to a central location where correlation takes place. As long as data were store on magnetic tapes, these media had to be sent by surface mail to the correlator, where they were put into playback units to provide the correlation units with the data stream. Nowadays data are stored on hard disks, which can be either shipped like the old fashioned tapes or the disk just works as buffer and the data stream is sent by broadband networks directly to the correlator where it is stored again on hard disks and can be taken as input for the correlator units. Recently experiments aiming at (near) real time VLBI were carried out and data were sent with no or minimal buffering using hard disks, but directly transmitting the formatted data over fiber networks to the correlation center (e.g. (Koyama et al., 2005, [46])).

Correlation involves basically integration and multiplication operations on the digital signals, which are sent from the two separate stations, and is done either by a hard-wired correlator or by software programs (software correlator see (Kondo et al., 2004, [45])), capable of performing the same tasks. One correlation unit is required to process the data from a single baseline, and one playback unit is required for each station contributing to the whole experiment. The following subsection will deal with the basic steps carried out within correlation and discuss the method how the geodetic observable (i.e. the time delay τ_g) can be obtained.

3.3.2 The correlator

The correlator is used to integrate the correlation function every several seconds by making continuous corrections to the signal data stream in order to compensate for variations in delay and Doppler shift due to the rotation of the Earth.

There are two types of correlators: the XF type correlator first performs the multiplication operation, then uses a Fourier transform to convert the signal’s time domain to the frequency domain. The FX type correlator first performs the Fourier transform of each data stream and then multiplies this output. The XF correlator was developed to achieve higher processing speed and is used for geodetic applications. Figure 11 shows a block diagram of an XF type correlator. Since in VLBI measurements the rotation of the Earth causes a continuous change in delay a suitable method of delay tracking must be implemented. Additionally, the Earth’s rotation causes the phase difference (fringe phase) between the signals received by the two stations to change continuously (fringe rotation). The effects on the signal frequency can be considered as a Doppler shift due to Earth’s rotational motion. Fringe

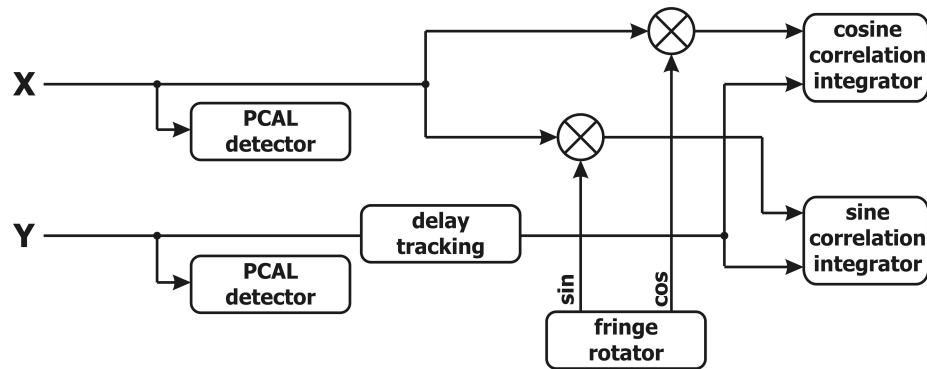


Figure 11: XF correlator block diagram.

rotation has to be compensated by a process called fringe stopping, which will be described later.

Delay tracking The correlator accounts for changes in the delay by modifying the read-address of the register of one of the two time-series. During playback of the data recorded from each station the signal is sent through the synchronous signal reproduction section to remove any large delay offsets. Then the buffer section (i.e. a shift register) makes the fine adjustment and tracks the delay. In real-time VLBI, the signals are transmitted directly from each station to the correlator, and geometric and transmission delays must be absorbed by the buffer section, whose capacity depends on the sampling frequency (rate) and on the a-priori delay. The delay is tracked by the correlator using steps in bit units, which causes a difference between a-priori delay and the discrete delay realized in the correlator. This difference must not exceed one bit unit within an integration period, which normally ranges from one to a few seconds. Additionally cumulative phase errors caused by fringe stopping must be held within a few 10s of degrees in order to maintain the correlation amplitude. A constant offset is permitted between a-priori delays but the a-priori delay rate must be known sufficiently precise. The delay can be expressed by

$$\tau = (\dot{\tau} + \Delta\dot{\tau})t + \frac{1}{2}(\ddot{\tau} + \Delta\ddot{\tau})t^2 + \dots \quad (3.4)$$

The contribution of $\Delta\dot{\tau}$, $\Delta\ddot{\tau}$, \dots must not exceed one bit over a unit integration time (parameter period) T_{pp} . This means, for a given sampling period T_{sample} the precision factor for a-priori calculations can be found by the following condition (omitting all terms past the 2nd derivative).

$$\Delta\dot{\tau}T_{pp} + \frac{1}{2}\Delta\ddot{\tau}T_{pp}^2 < T_{sample} \quad (3.5)$$

The magnitudes of τ , $\dot{\tau}$ and $\ddot{\tau}$ can be predicted using the geometrical delay model τ_g , deduced from station positions and source coordinates (Takahashi et al., 2000, [73]). A maximum delay of 21 ms will occur if one station observes the source in zenith direction and the other one is positioned such that the source is directly toward the horizon. The worst case scenario for the delay rate $\dot{\tau}_g$ gives a maximum value of $3.1 \mu\text{s/s}$, which means that delay changes by 3.1 microsecond when using a parameter period length of 1 second. Compared to $0.25\mu\text{s}$ bit length, according to 4 MHz sampling, delay rate changes are much too large to be neglected. Geometric delay acceleration $\ddot{\tau}_g$ can reach a maximum value of 110 ps/s^2 , which grows after 10 seconds integration period to only 6 ns, which is sufficiently small to be neglected for delay tracking purposes. It will become a topic if sampling rates reach 1 GHz or higher as this causes bit lengths in the nanosecond range. Anyway the 2nd time-derivative of the geometric time delay $\ddot{\tau}_g$ cannot be neglected in another process carried out during correlation and called fringe stopping.

Fringe stopping When the delay rate is a positive value, then viewed from the source, the remote station (Y) is steadily moving away from the reference station (X). Therefore the Doppler shift causes the signal received at station Y to be at lower frequency than the same signal received at station X. This means that the two signals cannot be correlated unless the frequency of the signal, observed

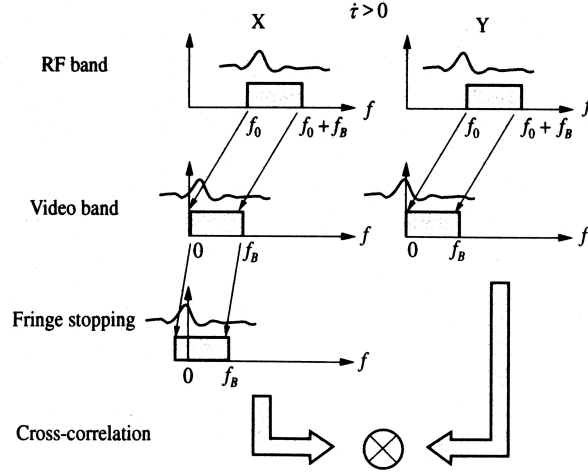


Figure 12: Qualitative description of fringe stopping (the spectrum is shown as a waveform in frequency direction, but in reality the spectrum should be flat), from Takahashi et al. (2000, [73]).

at station X is adjusted down (see figure 12). Frequency conversion is done by mixing the local oscillator signal f_{LO} with the radio signal f_{RF} to create the intermediate frequency of $f_{RF} + f_{LO}$ or $f_{RF} - f_{LO}$ in conventional receivers. The same principle is used in correlators, where the LO frequency can be considered as equivalent to the Doppler shift. This process is called fringe stopping and is accomplished by multiplying cosine and sine functions with respect to the time-series data of one of the stations (figure 11). Hard wired correlators usually use three level approximations of the trigonometric functions, stored in ROM chips, to save precious computation time. Software correlators can perform fringe stopping with any arbitrary resolution. The a-priori value of fringe phase is calculated by

$$\phi = 2\pi f\tau \quad (3.6)$$

which makes clear that, using an observation frequency of 8 GHz, the delay acceleration $\ddot{\tau}_g$ cannot be neglected now. Even for a parameter period length of 1 second, a fringe rotation of 160° would happen. In order to avoid degrading correlation amplitude, cumulative error in fringe phase rotation must not exceed a few 10s of degrees over one parameter period. The rate of change of the fringe phase (fringe rotation frequency) is given by

$$f_r = f\dot{\tau}. \quad (3.7)$$

Taking the maximum value of $3.1 \mu\text{s/s}$ and an observation frequency of 8 GHz gives 25 kHz for Earth based observations. Fringe stopping operations in the correlator produce the cosine and sine components, which are discussed now. The fringe stopping frequency f_r will be applied to the time-series $x(t)$ of station X. Multiplication gives the cosine (real) and sine (imaginary) parts as functions, depending on time t with the initial phase ϕ_0 .

$$x_r(t) = x(t) \cos(2\pi f_r t + \phi_0) \quad (3.8)$$

$$x_i(t) = x(t) \sin(2\pi f_r t + \phi_0) \quad (3.9)$$

Applying the Fourier transform on equations (3.8) and (3.9) gives

$$X_r(f) = \frac{1}{2} \left[X(f - f_r) e^{j\phi_0} + X(f + f_r) e^{-j\phi_0} \right] \quad (3.10)$$

$$X_i(f) = \frac{1}{2} \left[X(f - f_r) e^{j\phi_0} - X(f + f_r) e^{-j\phi_0} \right] \quad (3.11)$$

Introducing the time-series $y(t)$ of station Y, its Fourier transformation $Y(f)$ and its conjugate complex $Y^*(f)$ makes it possible to derive the correlation function output from the correlator

$$R_r(\tau) = \overline{x_r(t) \cdot y(t - \tau)} \quad (3.12)$$

$$R_i(\tau) = \overline{x_i(t) \cdot y(t - \tau)} \quad (3.13)$$

which can be obtained also from the inverse Fourier transform of the cross-spectrum

$$\begin{aligned} R_r(\tau) &= \int [X_r(f) \cdot Y^*(f)] e^{j2\pi f\tau} df \\ &= \frac{1}{2} [U(\tau) + L(\tau)] \end{aligned} \quad (3.14)$$

$$\begin{aligned} R_i(\tau) &= \int [X_i(f) \cdot Y^*(f)] e^{j2\pi f\tau} df \\ &= -\frac{1}{2}j [U(\tau) - L(\tau)] \end{aligned} \quad (3.15)$$

where

$$U(\tau) = \int [X(f - f_r) e^{j\phi_0} \cdot Y^*(f)] e^{j2\pi f\tau} df \quad (3.16)$$

$$L(\tau) = \int [X(f + f_r) e^{-j\phi_0} \cdot Y^*(f)] e^{j2\pi f\tau} df \quad (3.17)$$

was introduced. Equation (3.16) gives the cross-correlation between the signal from station Y and that one of station X increased in frequency by f_r . Equation (3.17) gives the same relationship, but the signal from station X is decreased by frequency f_r . This means that expression (3.16) is equivalent to the processing of a negative delay rate and $L(\tau)$ is the same when a positive delay rate is used. Solving equations (3.14) and (3.15) for $U(\tau)$ and $L(\tau)$ gives

$$U(\tau) = R_r(\tau) + jR_i(\tau) \quad (3.18)$$

$$L(\tau) = R_r(\tau) - jR_i(\tau) \quad (3.19)$$

Finally the cross-correlation function after fringe stopping can be summarized by

$$R(\tau) = R_r(\tau) \pm R_i(\tau) \quad (3.20)$$

where the positive sign is valid for $\dot{\tau} < 0$ and the negative one has to be applied for $\dot{\tau} > 0$. Both, positive and negative fringe stopping operations are possible by considering the corresponding signs of the sine and cosine components even though fringe rotation may remain a positive value continuously. Therefore, the above equation is applicable for fringe stopping with positive and negative delay rates. Although bidirectional fringe stopping could be realized and it would make following processes simpler, it is usually not implemented in the correlator.

Bit shifts and 90-degree phase jumps Fringe stopping is usually performed at frequencies calculated at the baseband f_0 or at the center of the video band $f_0 + f_B/2$, using the video band range from 0 to f_B . The baseband is defined as the frequency of the RF signal which is finally converted to the zero (video) frequency. As delay tracking can only be done in discrete values due to its implementation in bit units, a continuously varying error will develop between the correlator provided delay and the a-priori delay. Figure 13(a) compares delay tracking using a-priori delay and the discretized delay produced by the correlator and the resulting errors. Expressed in bit units, this error represents the fraction part of the bit, and is consequently called fractional bit T_{fbit} , which can be described by

$$T_{fbit} = nint\left(\frac{\tau}{T_{bit}}\right) - \frac{\tau}{T_{bit}} \quad (3.21)$$

where τ represents the a-priori delay and one bit length is defined by $T_{bit} = 1/(2f_B)$. The function $nint()$ returns the closest whole integer. The first term of equation (3.21) represents discrete delay tracking as provided by the correlator and restricts T_{fbit} to values ± 0.5 bits. Such errors are completely compensated in the frequencies at which fringe stopping takes place. But phase changes at other frequencies than the fringe stopping frequency will arise according to

$$\phi = 2\pi f T_{fbit} \quad (3.22)$$

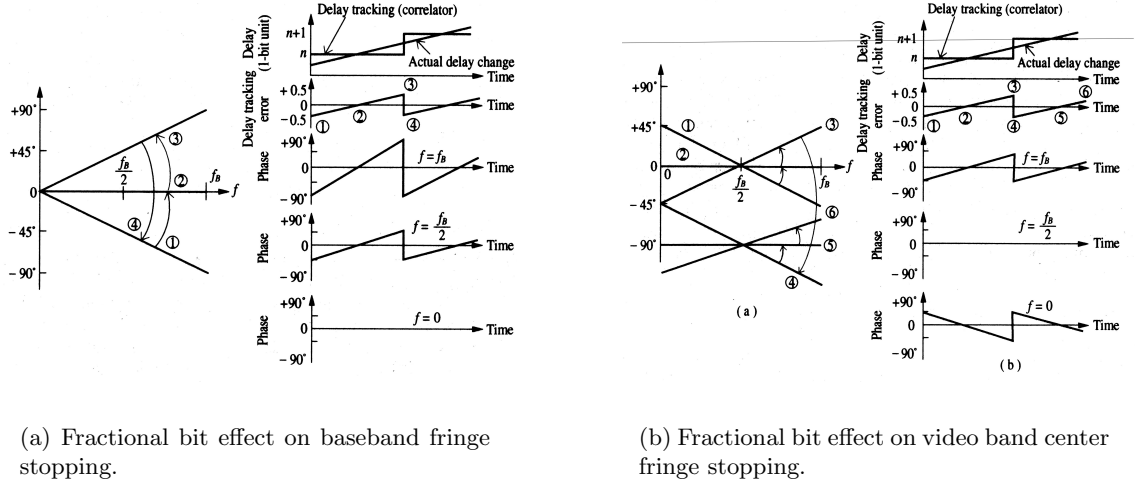


Figure 13: Implication of fringe stopping on fractional bits (Takahashi et al., 2000, [73]).

and will consequently cause shifts between -90° and $+90^\circ$. Figure 13(a) illustrates this effect, the left part illustrates the effect in dependency of the taken frequency. The transition between points 3 and 4 (figure 13(a)) reveals a phase change, which is caused by setting the zero point of the video frequency to the baseband. This will result in a degradation of the correlation amplitude, which can be predicted as follows. Taking a sampling period (1 bit length) of T_{bit} , a video bandwidth of f_B and a cross-spectrum of $X(f) = 1$ the correlation amplitude can be gained by

$$P_A = \int_{-T_{bit}/2}^{T_{bit}/2} \int_0^{f_B} X(f) df dt = f_b T_{bit} = \frac{1}{2}. \quad (3.23)$$

If fringe stopping at the baseband is taken into account phase will change according to $\exp(j2\pi ft)$ and the integrated correlation amplitude will be

$$\begin{aligned} P_B &= \int_{-T_{bit}/2}^{T_{bit}/2} \int_0^{f_B} X(f) e^{j2\pi ft} df dt \\ &= \frac{1}{2} \left(1 - \frac{\pi^2}{2^2 \cdot 3 \cdot 3!} + \frac{\pi^4}{2^4 \cdot 5 \cdot 5!} + \frac{\pi^6}{2^6 \cdot 7 \cdot 7!} + \dots \right) \end{aligned} \quad (3.24)$$

Now degradation of the correlation amplitude can be predicted by

$$\begin{aligned} \frac{P_B}{P_A} &= 1 - \frac{\pi^2}{2^2 \cdot 3 \cdot 3!} + \frac{\pi^4}{2^4 \cdot 5 \cdot 5!} + \frac{\pi^6}{2^6 \cdot 7 \cdot 7!} + \dots \\ &\approx 1 - \frac{\pi^2}{72} \end{aligned} \quad (3.25)$$

which means that a loss of about $\pi^2/72 (\approx 13.7\%)$ in the correlation amplitude has to be taken into account, when fringe stopping at the baseband is carried out.

To avoid such a limitation, fringe stopping can also be performed at the RF frequency corresponding to the center frequency of the video band. Figure 13(b) shows the effects caused by this approach and the phase can be calculated by

$$\phi = 2\pi \left(f - \frac{f_B}{2} \right) T_{bit}. \quad (3.26)$$

Now phase changes become zero at the video band center frequency and changes occur only in a band of $\pm 45^\circ$. But at the bit shift point, the phase changes around the baseband axis and at the 90° phase

shift point the sequence will be repeated again. This means that with each bit shift, fringe phase will change in the same direction, which makes correlation impossible if it is not accounted for. To overcome this weakness a 90° jump is forced by the correlator each time a bit shift occurs. Indeed, this means an additional capability, which has to be fulfilled by the correlator, but the loss of correlation amplitude can be reduced remarkably, which will be shown now. Approximation of the correlation amplitude can be done like above, introducing a phase change expressed by $\exp(j2\pi(f - f_B/2)t)$ gives

$$\begin{aligned} P_C &= \int_{-T_{bit}/2}^{T_{bit}/2} \int_0^{f_B} X(f) e^{j2\pi(f - \frac{f_B}{2})t} df dt = \int_{-T_{bit}/2}^{T_{bit}/2} \int_{-f_B/2}^{f_B/2} X(f' + \frac{f_B}{2}) e^{j2\pi f' t} df' dt \\ &= \frac{1}{2} \left(1 - \frac{\pi^2}{4^2 \cdot 3 \cdot 3!} + \frac{\pi^4}{4^4 \cdot 5 \cdot 5!} + \frac{\pi^6}{4^6 \cdot 7 \cdot 7!} + \dots \right) \end{aligned} \quad (3.27)$$

Therefore

$$\begin{aligned} \frac{P_C}{P_A} &= 1 - \frac{\pi^2}{4^2 \cdot 3 \cdot 3!} + \frac{\pi^4}{4^4 \cdot 5 \cdot 5!} + \frac{\pi^6}{4^6 \cdot 7 \cdot 7!} + \dots \\ &\approx 1 - \frac{\pi^2}{288} \end{aligned} \quad (3.28)$$

a loss of correlation amplitude of $\pi^2/288 (\approx 3.4\%)$ is expected, which is much smaller than using fringe stopping at baseband frequency.

3.3.3 Precise determination of delay and delay rate

The correlator integrates the data using a-priori values of delay and its time derivative, and splits the data into time segments of a few seconds length. In a later step carried out by a computer, the integrated data are being corrected for differences between a-priori values and actual delay and delay rates, whereas corrections to the a-priori values are added to maximize the correlation coefficient. As a result corrected values for delay and delay rate are obtained. For a successful correlation a certain degree of precision of the a-priori values is needed. This means that these values are close enough to real (observed) values, so that correlation is not lost during the parameter period T_{pp} chosen for integration. Precision of the delay rate is the most important factor for running the correlation process successfully over T_{pp} , as it has a direct impact on fringe phase rotation. Correlation will fail if phase changes exceed 180 degrees during one parameter period. Furthermore, time delay must be kept within the limits of the search window, otherwise no correlation maximum can be found. The search for delay and delay rate residuals are preceded by normalization of the correlator output and extraction of the phase calibration (PCAL) information, which will become important for bandwidth synthesis processing. Taking the cosine (real) $a_r(n, k, d)$ and sine (imaginary) $a_i(n, k, d)$ output from the correlation integrator count and the integrator bit number counts $A_r(n)$ and $A_i(n, k)$ makes it possible to obtain the normalized complex cross-correlation function by

$$\mathcal{R}(R(n, k, d)) = \frac{2a_r(n, k, d) - A_r(n, k)}{A_r(n, k)} \quad (3.29)$$

$$\mathcal{I}(R(n, k, d)) = -\frac{2a_i(n, k, d) - A_i(n, k)}{A_i(n, k)} \quad (3.30)$$

where n , k and d represent the corresponding frequency channel, parameter period and correlator lag (in bit units). The correlation function $R(n, k, d)$ has to be compensated for losses due to 1-bit sampling, fringe stopping and fractional bit occurrences, in order to obtain the true correlation function $r(n, k, d)$

$$r(n, k, d) = \sin(C \cdot R(n, k, d)) \approx C \cdot R(n, k, d) \quad (3.31)$$

using the coefficient

$$C = \frac{\pi}{2} \cdot \frac{3}{4} \cdot \frac{\pi}{4 \cos\left(\frac{\pi}{8}\right)} \cdot \frac{1}{1 - \frac{\pi^2}{288}} \quad (3.32)$$

which might change due to correlator architecture and parameter period length (integration time). Because the correlation function is in all cases much smaller than one the Taylor expansion in equation (3.31) can be neglected after order one. If a Fourier transform is performed on the correlation function (3.31), the video cross-spectrum is obtained. If J is the number of lags of each correlation function, then J point zero data are used to extend the length of the dataset, which causes an interpolation in the frequency domain. Then the whole $2 \cdot J$ point correlation function is Fourier transformed. Therefore the cross-spectrum consists of the values from the division of the video band into a number J of equal points. Then $j = 0 \dots J$ can be taken as an index within the video band and the video cross-spectrum $S_v(j, k, n)$ can be obtained. It will serve as a basis for the coarse search function in order to determine delay and delay rate.

Coarse delay search function The coarse search function deals with the correction of the residual delay error ($\Delta\tau$) in the video bands only. It is mainly used to assess the correlation results and to check if the correlator parameters are set properly. If correlation amplitude shows no clear peak or the maximum appears to be on the edge of the lag window, a-priori values will have to be modified and the correlation process has to be done again. The correction in the video bands, for channel n can be expressed mathematically by

$$F(n, \Delta\tau, \Delta\dot{\tau}) = \frac{1}{K} \sum_{k=1}^K \left\{ \frac{1}{J-1} \sum_{j=1}^{J-1} S_v(j, k, n) e^{-j2\pi f_j^v \Delta\tau} \right\} e^{-j2\pi f_0^n \Delta\dot{\tau} \Delta t k} \quad (3.33)$$

where f_j^v is the frequency corresponding to the index j in the video band, f_0^n is the RF frequency of channel n , $\Delta\tau$ and $\Delta\dot{\tau}$ are the residual errors of delay and delay rate, which have to be searched, and Δt is the time interval of k (i.e. the T_{pp} as defined above). In other words, equation (3.33) corrects phase rotation in the video band caused by $\Delta\tau$ and corrects fringe phase rotation in time direction due to $\Delta\dot{\tau}$ with respect to the video cross spectrum. Summation is carried out between $j = 1$ to $j = J - 1$ to get rid of the DC offset (bias) and to compensate for degradation at both edges of the video band. The coarse delay resolution function itself is defined by integrating the absolute values of

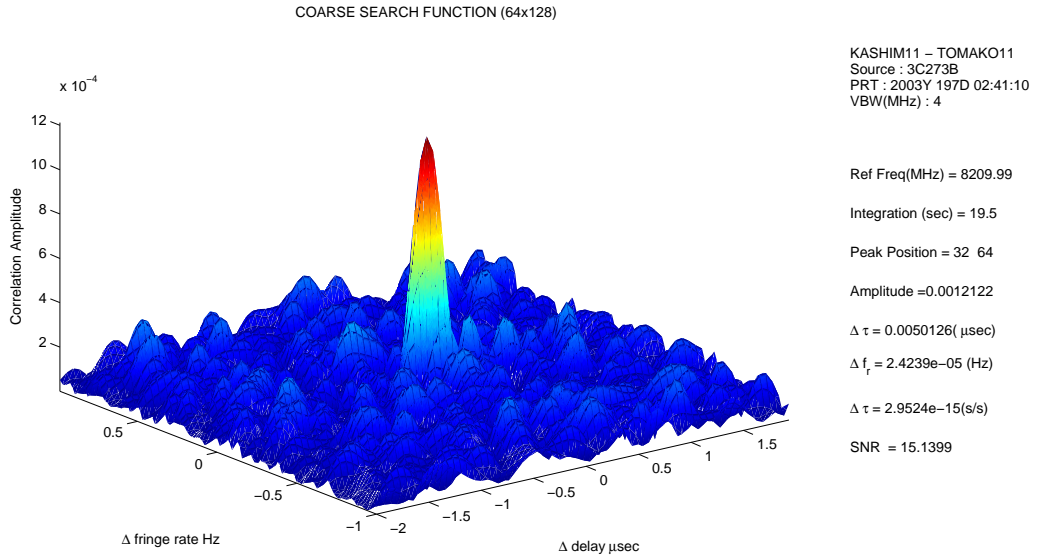


Figure 14: Single channel coarse delay resolution function, Hobiger and Kondo (2005, [32]).

$F(n, \Delta\tau, \Delta\dot{\tau})$ over all frequency channels.

$$F(\Delta\tau, \Delta\dot{\tau}) = \frac{1}{N} \sum_{n=1}^N |F(n, \Delta\tau, \Delta\dot{\tau})| \quad (3.34)$$

$F(\Delta\tau, \Delta\dot{\tau})$ has to be maximized by varying the values of $\Delta\tau$ and $\Delta\dot{\tau}$, which would take a lot of computation time if given mathematical relations are not considered. Since equation (3.33) is a

two-dimensional Fourier transform with respect to $\Delta\tau$ and $f_0^n \Delta\hat{\tau}$, algorithms, using the fast Fourier transform (FFT), can be applied to calculate the coarse delay resolution function. In this approach only the frequency components at the video band edges zero ($j = 0$) and f_B ($j = J$) and the lower side band (LSB) values have to be set to zero, then a FFT can be carried out on the cross video spectrum as obtained from the correlator. There are several methods to find the maximum value of equation (3.34). The method of steepest descent (or Newton's method) will lead to the maximum, but a simpler approach called "mountain climbing" is usually taken, as it can be implemented easily (see Takahashi et. al, 2000, [73]). Figure 14 shows as an example the coarse delay function from a single channel (Hobiger and Kondo, 2005, [32]) from a geodetic experiment.

Bandwidth synthesis for precise delay search In the fine search process residual errors of the delay are solved by making corrections over all frequency channels by adjusting phase rotation. The procedure of combining data from different frequency channels is referred to as bandwidth synthesis. Inter-channel phase corrections (i.e. phase differences in all channels) can be accomplished by phase calibration signal phase, which is extracted for each station within correlation processes (see figure 11). The residual errors $\Delta\tau$ and $\Delta\hat{\tau}_s$ from coarse search are used in a first step function

$$D_s(n, k) = \frac{1}{J-1} \sum_{j=1}^{J-1} S_v(j, k, n) e^{-j2\pi(f_j^v \Delta\tau_s + f_0^n \Delta\hat{\tau}_s \Delta t k)} \quad (3.35)$$

to correct the correlation function corrected with results from coarse search processes. This can be used to set up the fine delay resolution function

$$D(\Delta\tau, \Delta\hat{\tau}) = \frac{1}{N} \sum_{n=1}^N \left\{ \frac{1}{K} \sum_{k=1}^K D_s(n, k) e^{-j(2\pi f_0^n \Delta\hat{\tau} \Delta t k + \Delta\phi_n)} \right\} e^{-j2\pi f_0^n \Delta\tau} \quad (3.36)$$

where the mean values of the PCAL phase from stations X and Y in channel n are used.

$$\Delta\phi_n = \phi_{nx} - \phi_{ny} \quad (3.37)$$

The term outside the brackets in equation (3.36) represents the inter-channel correction of the delay residuals. Considering this correction is called bandwidth synthesis, which increases delay precision roughly proportional to the inverse of the band, which is usually much better than single band delay precision. When different frequency channels are combined, smaller side lobe levels in the bandwidth synthesis are better. When a small number of channels is involved, and with certain restrictions (i.e. frequency steps, frequency range), channels can be combined by computing bandwidth synthesis for all possible frequency channels. Thereafter the channel combination with the minimum side lobe maxima can be chosen, which is also very effective for avoiding interference signals in the reception band. Like for the coarse delay resolution function (equation (3.33)) a two-dimensional Fourier transform can be used instead of double summations in equation (3.36). Therefore a small modification must be made, by extracting the lowest frequency (f_0^1) of the channels involved from the summation.

$$\begin{aligned} D(\Delta\tau, \Delta\hat{\tau}) &= \frac{1}{N} \sum_{n=1}^N \left\{ \frac{1}{K} \sum_{k=1}^K D_s(n, k) e^{-j(2\pi f_0^n \Delta\hat{\tau} \Delta t k + \Delta\phi_n)} \right\} e^{-j2\pi f_0^n \Delta\tau} \\ &= e^{-j2\pi f_0^1 \Delta\tau} \left[\frac{1}{N} \sum_{n=1}^N \left\{ \frac{1}{K} \sum_{k=1}^K D_s(n, k) e^{-j(2\pi f_0^n \Delta\hat{\tau} \Delta t k + \Delta\phi_n)} \right\} e^{-j2\pi (f_0^n - f_0^1) \Delta\tau} \right] \end{aligned} \quad (3.38)$$

The phase term (outside the squared brackets) has to be considered in a later step. The values $\Delta\tau$ and $\Delta\hat{\tau}$, maximizing $D(\Delta\tau, \Delta\hat{\tau})$ are found again by a mountain climbing method. Representing $\Delta\tau$ and $\Delta\hat{\tau}$ (corrections from the fine search procedure) by $\Delta\tau_m$ and $\Delta\hat{\tau}_m$ makes it easier to denote the final residuals (to the a-priori values) by

$$\Delta\tau_{obs} = \Delta\tau_m + l \cdot t_{am} \quad (l = 0, \pm 1, \pm 2, \pm 3 \dots) \quad (3.39)$$

$$\Delta\dot{\tau}_{obs} = \Delta\dot{\tau}_s + \Delta\dot{\tau}_m + \dot{\Psi}_X - \dot{\Psi}_Y \quad (3.40)$$

where t_{am} describes the inverse of the greatest common measure of the frequency spacing in the reception channels, and can be treated like an ambiguity unit. During bandwidth synthesis, t_{am} represents the pattern repetition period in $\Delta\tau$ domain, due to limitation of discrete frequencies. These remaining ambiguities have to be solved for in the final step, when doing baseline analyses. $\dot{\Psi}_X$ and $\dot{\Psi}_Y$ the time derivatives of the phase calibration signal phases and represent a drift rate of the PCAL signal during one scan. Finally the observed correlation amplitude is given by

$$\rho_{obs} = |D(\Delta\tau_m, \Delta\dot{\tau}_m)| \quad (3.41)$$

and the corresponding fringe phase $\arg(D(\Delta\tau_m, \Delta\dot{\tau}_m))$.

Ambiguities in bandwidth synthesis From the prior steps a given frequency f_n for channel n with phase $\phi_n + 2m_n\pi$, where $m_n = l = 0, \pm 1, \pm 2, \pm 3 \dots$ is obtained. Without loss of generality $\phi_0 = 0$ can be set, which makes it possible to express any frequency f_n depending on the frequency of the first channel and on the frequency spacing.

$$\begin{aligned} f_n &= f_{n-1} + \Delta f_{n-1} \\ &= f_1 + \Delta f_1 + \Delta f_2 + \dots + \Delta f_{n-1} \\ &= f_1 + \sum_{k=1}^{n-1} \Delta f_k \end{aligned} \quad (3.42)$$

where the frequency difference between channel k and channel $k+1$ is denoted by Δf_k . The delay of each channel is computed from the reference frequency and phase with respect to integer multiples of 2π . When channel 1 is used as the reference for frequency and phase, then the delay τ_n for channel n is given by

$$\tau_n = \frac{2m_n\pi}{2\pi(f_n - f_1)} = \frac{m_n}{\sum_{k=1}^{n-1} \Delta f_k}. \quad (3.43)$$

As the delay should be the same in all channels the condition $\tau_2 = \tau_3 = \dots = \tau_N$ (using N channels) has to be fulfilled. This leads to

$$\frac{m_2}{\Delta f_1} = \frac{m_3}{\Delta f_1 + \Delta f_2} = \dots = \frac{m_N}{\Delta f_1 + \Delta f_2 + \dots + \Delta f_{N-1}} \quad (3.44)$$

and introduction of $\Delta f_n = a_n \Delta f_{max}$ using the integer value a_n modifies equation (3.44) to

$$\frac{m_2}{a_1 \Delta f_{max}} = \frac{m_3}{(a_1 + a_2) \Delta f_{max}} = \dots = \frac{m_N}{\Delta f_{max} \sum_{k=1}^{N-1} a_k}. \quad (3.45)$$

As a_k can only take integer values, the assumption of $m_n = \sum_{k=1}^{n-1} a_k$ means that all terms of expression (3.45) become $1/\Delta f_{max}$. Therefore, the ambiguity value becomes a minimum, when Δf_{max} is the greatest common divisor of the frequency spacings. E.g. a value of $\Delta f_{max} = 10$ MHz is equal to an ambiguity of 100 ns, a value which can be rather easily adjusted.

3.3.4 Measurement precision

The signal-to-noise ratio can be defined as

$$SNR = \rho_0 \sqrt{2BT} \quad (3.46)$$

where ρ_0 is the correlation amplitude, B the bandwidth and T represents total integration length. This means $\sqrt{2BT}$ will increase SNR when scan lengths get longer, which can be explained by the

suppression of the (random) noise components with longer integration times. It can be shown that the standard deviation of the fringe phase σ_ϕ is related to SNR by

$$\sigma_\phi = \frac{1}{SNR}. \quad (3.47)$$

Also the formal error of delay and delay rate after bandwidth synthesis can be deduced from SNR . First the variance of N frequency channels has to be defined by

$$\begin{aligned} \sigma_f^2 &= \frac{1}{N} \sum_{n=1}^N (f_n - \bar{f})^2 \\ &= \frac{1}{N} \sum_{n=1}^N f_n^2 - \left(\frac{1}{N} \sum_{n=1}^N f_n \right)^2 \end{aligned} \quad (3.48)$$

then the standard deviation of the adjusted delay is given by

$$\sigma_\tau = \sqrt{\frac{1}{4\pi^2 \sigma_f^2 SNR^2}}. \quad (3.49)$$

This means that a frequency setup with a spread channel distribution will make delay observables more precise. But one has to be aware that separation of the frequency channels will also reduce ambiguity spacing, which makes bandwidth synthesis more complicate (as discussed in equation (3.45)). In a similar way the standard deviation for delay rate $\sigma_{\dot{\tau}}$ can be denoted by

$$\sigma_{\dot{\tau}} = \sqrt{\frac{1}{4\pi^2 f^2 \sigma_t^2 SNR^2}} \quad (3.50)$$

where

$$\sigma_t^2 = \frac{1}{K} \sum_{n=1}^K (t_k - \bar{t})^2 \quad (3.51)$$

is the time variance using K parameter periods. If scan length gets longer, the time variance will grow, too, and the formal error of the delay rate will decrease. One way to obtain more precise measurements is to increase the SNR , which can be done by extending the bandwidth or recording longer scans (see equation (3.46)). The correlation amplitude itself depends on the antenna noise temperatures (T_{nx} and T_{ny}) and on the equivalent noise temperature of the source signal (T_{ax} and T_{ay}) as received by antennas X and Y and can be expressed as follows.

$$\rho_0 = \sqrt{\frac{T_{ax} T_{ay}}{(T_{ax} + T_{nx})(T_{ay} + T_{ny})}} \quad (3.52)$$

So, if the antenna noise temperatures can be reduced by the usage of improved electronic equipment the signal-to-noise ratio will increase and the precision of the observable will get better.

3.4 Applications

There are many fields of research which profit from the observations made by VLBI, directly or indirectly. On the one hand there is the radio astronomy community, which is interested in any characteristics of the observed objects, on the other hand there are geophysicists and geodesists who investigate Earth orientation and Earth environmental processes by using VLBI. In between is the astrometric community, which takes care of the inertial reference system, realized by radio sources and provides reliable and precise coordinates of these objects. From correlation processing the following types of observables can be obtained:

- phase ϕ , i.e. the fringe phase ϕ for each parameter period,

- correlation amplitude ρ ,
- delay τ (after bandwidth synthesis) and
- delay rate $\dot{\tau}$ (after bandwidth synthesis).

To some extent, geodetic and astronomical VLBI differ in what are considered to be observables and what are nuisance parameters (Walker, 2000, [79]). An obvious case is the delay. For both the delays are determined with a fringe fit to phase slopes. For geodesy most subsequent processing is based on the delays and the amplitudes and phases are not used beyond the fringe fit. For astronomy the delays are only used to flatten the fitted phase slopes. Subsequent processing is based on the calibrated amplitudes and phases. For geodesy and astrometry the scientific targets of the observations are the geometric parameters. Source structure, to the extent that it affects the data, is a nuisance parameter. For astronomy, on the other hand, source structure is usually the scientific target and any errors in the a priori geometry are nuisance parameters. One thing that both groups agree on is that phase offsets introduced by the troposphere and ionosphere are nuisance parameters. There are differences in the way models are treated for geodesy and astronomy. For geodesy, the correlator model only needs to be good enough to reduce the phase slopes in time (fringe rate) and frequency (delay) to allow a reasonable amount of averaging before the fringe fit and fairly simple models can be used. After the fringe fit the residuals are added to the model removed by the correlator to construct the “totals” that are used in all further processing. The most important aspect of the model is that it is accountable, so that accurate totals can be constructed. Also, to simplify model independence, the baselines are often treated independently in both the correlator and the fringe fitting. The time tags on the data are usually the wavefront arrival times at one of the stations. For astronomy, the model is rarely modified after correlation. For simple imaging experiments, the model quality is also not very important as errors are removed by the fringe fitting. But for phase referencing a high quality model is required. Phase referencing is based on the assumption that the phase errors are the same on calibrator and target, and this is only true with a good model. It is much simpler to analyze such observations if a good model is used on the correlator.

3.4.1 Astronomical applications

Two radio telescopes pointing at the same radio source can be utilized to gain a “map” of this object. Usually the baseline vector \vec{B} is measured in wavelengths at the center of the frequency of the observing band, and the baseline has the components (u, v, w) in a right-handed coordinate system, where u and v are measured in a plane normal to the direction of the phase-reference position (see figure 15). The spacing component v is measured towards north as defined by the plane through the origin, the source, and the pole, and u is measured towards east. The component w is measured in the direction of the phase reference position of the source. The radio object has an intensity variation $I(x, y)$, where x and y are directions on the sky with respect to u and v . The intensity distribution of the source

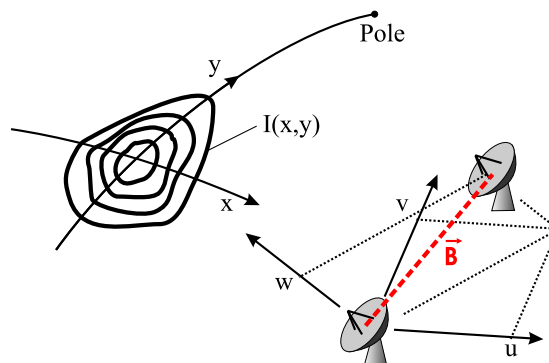


Figure 15: Geometric relationship between the source with an intensity distribution $I(x, y)$ and an antenna pair.

$I(x, y)$ can be related to the measured visibility $V(u, v, w)$ (i.e. a flux density in $\text{W m}^{-2} \text{Hz}^{-1}$) by

$$V(u, v, w) = e^{-j2\pi w} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} I(x, y) e^{-j2\pi(ux+vy)} dx dy. \quad (3.53)$$

It is customary to absorb the factor in front of the integral to the left hand side of equation (3.53), which gives (without changing notation of V)

$$V(u, v) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} I(x, y) e^{-j2\pi(ux+vy)} dx dy. \quad (3.54)$$

Since equation (3.54) is a Fourier transform, it may be formally inverted

$$I(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} V(u, v) e^{j2\pi(ux+vy)} du dv \quad (3.55)$$

which makes it possible to conclude on the source structure from measuring the visibility $V(u, v)$ by radio telescopes. The visibility is a complex quantity, where its amplitude is related to the correlation amplitude and its phase is the correlated fringe phase, both available for each parameter period within one scan. Hence the relation between visibility amplitude and normalized correlation amplitude (ignoring digital processing factors) is given by

$$|V(u, v)| = 2k \cdot \rho(u, v) \sqrt{\frac{T_{s1} T_{s2}}{A_1 A_2}} \cdot 10^{26} \quad [\text{Jy}] \quad (3.56)$$

where A_1 and A_2 are the antenna surfaces, T_{s1} and T_{s2} are the system temperatures and k is the Boltzmann constant. Here the flux density is given in Jansky ($1 \text{ Jy} = 10^{26} \text{ W m}^{-2} \text{Hz}^{-1}$) for convenience. As the Earth rotates the baseline components u and v will change according to observing geometry and observation length. If one source is tracked sufficiently long and several stations participate in the same experiment, the $u - v$ plane will contain many samples to perform the inverse transformation (equation 3.55) and to obtain an image of the radio source. In other words, in practice, the spatial coherence function $V(u, v)$ is not known everywhere, but is sampled at particular places on the $u - v$ plane. The sampling can be described by a sampling function $S(u, v)$, which is only one where data have been taken. Then, instead of equation (3.55) one may calculate

$$I^D(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} V(u, v) S(u, v) e^{j2\pi(ux+vy)} du dv \quad (3.57)$$

where $I^D(x, y)$ is the so-called dirty image (Clark, 1995, [14]). Its relation to the desired intensity distribution $I(x, y)$ is seen by using the Fourier convolution theorem

$$I^D(x, y) = I(x, y) * B(x, y) \quad (3.58)$$

where the asterisk denotes convolution, and

$$B(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} S(u, v) e^{j2\pi(ux+vy)} du dv \quad (3.59)$$

is called the synthesized beam, point spread function, or dirty beam corresponding to the sampling function $S(x, y)$. The final goal is now to find the most reasonable (physical and mathematical) model of the source which fulfills the convolution equation (3.59).

3.4.2 Geodetic and astrometric applications

The geodetic and geophysical applications from VLBI are based on the use of an inertial reference frame formed by a given set of extremely compact extragalactic radio sources. VLBI measures very accurately the angles between the Earth-fixed baseline vectors and the space-fixed radio sources. Thus, even the smallest changes in the baseline lengths and in the angles between the reference systems can be detected. The main geodynamic phenomena such as polar motion, UT1 variations, nutation and precession, Earth tides, ocean tidal response, and tectonic plate motions can be monitored with high accuracy. The group delay τ obtained from bandwidth synthesis processing acts as the main observable for all computations as it determines the baseline-source geometry, which relates the location of the radio telescopes on a rotating Earth to the infinitely distant radio sources. These (nearly) point-like objects without proper motion can be treated as fixed beacons in the heavens, which make it possible to monitor small discrepancies from the computed motions of the receiving stations. By using least-squares parameter estimation, a Kalman filter or other estimation methods the most likely values of the unknown quantities can be solved for. In a prior step systematic effects on the raw observations have to be removed in order to improve the final accuracy of the results. Figure 16 shows a flow diagram of geodetic data analysis. The left side of figure 16 deals with the actual observations, which

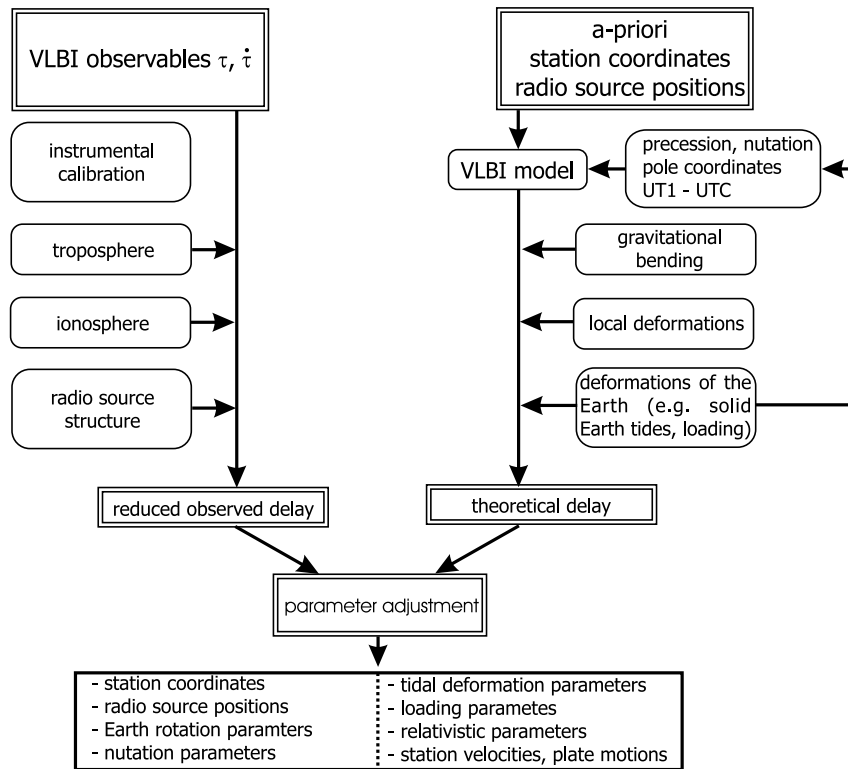


Figure 16: Flow diagram of geodetic VLBI data analysis.

are corrected for instrumental and environmental influences. The right side of the diagram shows the computation of the so-called theoreticals having a set of a-priori values, which are taken as initial values for the unknown parameters of the VLBI model. In parameter estimation processes (e.g. a least-squares adjustment) the differences “observed minus computed” are formed and corrections to the initial values of the unknowns are obtained. Group delay ambiguities (see section 3.3.3, page 32) have to be fixed to a common level, after a first adjustment was carried out, using only delay rate observables. Additionally, one has to consider delay closure as an additional constraint on ambiguity fixing. Effects of the atmosphere on VLBI observations are some of the big contributions to the total error budget, because widely separated stations pointing at the same source experience different meteorological conditions. The neutral atmosphere, essentially the troposphere, exerts an influence on radio signals, which adds up to an extra zenith path of 1.8 to 2.5 m and increases with zenith distance. The contribution of the dry part is rather stable and can be covered by a suitable mapping

function, but the wet component, although it has a smaller contribution, changes rapidly and needs special treatment. It either has to be estimated within parameter adjustment using special mapping functions (e.g. Niell, 1996, [52] or Boehm and Schuh, 2004, [9]) or it has to be corrected by external measurements like water vapor radiometers (WVR), which measure the microwave thermal emission from water vapor at 22 GHz in the line-of-sight. Ionospheric influences on the observed delay can be compensated by measuring at two distinct frequencies (see section 4.2.2). This correction has to be applied together with the other ones, before the adjustment can be carried out. Although the radio sources are nearly compact (point-like) they show a structure of a few mas, which sets up a limit on the accuracy of the radio reference frame. Therefore permanent monitoring of the radio sources by means of radio astronomy (mapping) enables geodetic analysts to consider effects of source structure on the reference frame. Figure 9 describes the geometric situation which relates the delay τ_g to the baseline vector \vec{B} and to the unit vector \vec{i}_s in the direction of the radio source.

$$\tau_g = -\frac{\vec{B} \cdot \vec{i}_s}{c} = -\frac{1}{c} [B_x \cos \delta \cos h(t) + B_y \cos \delta \sin h(t) + B_z \sin \delta] \quad (3.60)$$

with the geocentric baseline components B_x, B_y, B_z , the radio source position α, δ and the Greenwich hour angle $h(t) = GST - \alpha$ (GST - Greenwich sidereal time). Starting from the basic form in a geocentric system the rather complex relativistic formulation in the solar system barycenter (SSB) has to be carried out, where the relativistic formulation includes both the effect of special relativity (SRT) and of general relativity (GRT). The effects of special relativity are due to the fact that Lorentz transformations ($\sim v/c^2$) connect the coordinate frames which are moving relative to each other. The effect of gravity on the propagation of electromagnetic waves is not less important and covered by general relativity. The space-time structure will be deformed due to the presence of huge masses. Bodies as the Sun and the Earth itself have to be taken into account when ray paths come close to them, but even for Jupiter a noticeable effect must be considered. The baseline vector \vec{B} (in the terrestrial coordinates system) and the radio source positions in the quasi-inertial celestial reference frame in space can be related to each other by

$$\tau_g = -\frac{1}{c} \vec{B} \cdot \mathbf{W} \cdot \mathbf{S} \cdot \mathbf{N} \cdot \mathbf{P} \cdot \vec{i}_s \quad (3.61)$$

where

- \mathbf{W} is the rotation matrix for polar motion (wobble),
- \mathbf{S} is the diurnal spin matrix,
- \mathbf{N} is the nutation matrix and
- \mathbf{P} is the precession matrix.

The orientation of the Earth with respect to the celestial system (precession and nutation), the motion of the Earth's axis with respect to the crust (polar motion) and the phase angle of the Earth's rotation ($dUT1 = UT1 - UTC$) have to be modeled at the same accuracy levels as all the other components. Earth tides and ocean tides have to be taken into account when a-priori models of the Earth rotation parameters are calculated. Any target parameters can be obtained from equation (3.61), if the network geometry and the number of observations is sufficiently high so that the unknowns can be determined reliably within an adjustment process. Thus, only measurements carried out in large networks, i.e. with baselines longer than 6000 km, are normally used for the determination of the Earth's orientation. Another important factor is the geometry of the VLBI network. Long east-west baselines are mainly sensitive to UT1 whereas networks with long north-south components allow to measure precisely both components of the pole position. Periodic and aperiodic deformations of the Earth's crust have to be taken into account as well. Solid Earth tides show diurnal and semidiurnal oscillations, which cause vertical deformations in a range of several 10s of cm and horizontal displacements of about 30% of the vertical effect. More difficult to model is the loading by the water masses of ocean tides and currents (ocean loading), which amounts to as much as a decimeter on some coastal or island sites.

Additionally loading effects due to air pressure variations (atmospheric loading) also reach the level of significance in VLBI modeling and should be considered in routine analysis. Finally deformation of the antenna structure has to be considered, too. Then the theoretical delays can be compared with the reduced observed delays by any parameter estimation technique (e.g. a least-squares fit). Finally, the estimated parameters can be separated into two groups.

- Auxiliary parameters which have to be computed but are not “usual” geodetic products, e.g. clock parameters and atmospheric parameters
- Geodetic/astrometric parameters, which are of interest for the user community, e.g. baseline components or station coordinates with respect to a reference station, Earth orientation parameters (UT1-UTC, pole coordinates, nutation coefficients) or radio source positions.

4 VLBI and the ionosphere

4.1 Impact of the ionosphere on space geodetic techniques

As discussed in chapter 2 the ionosphere influences or completely prohibits propagation of electromagnetic waves. The dispersive characteristic of the ionosphere, which is indirectly proportional to the observing frequency, has a big impact on observables like fringe phase, delay, polarization, Given the S-band ($f_S = 2.3$ GHz) and X-band ($f_X = 8.4$ GHz) frequencies typically used in geodetic and astrometric VLBI experiments, the effect due to plasma frequency and electron gyrofrequency can be investigated further. Table 2 summarizes plasma frequency effects, taking into account that the received radio waves also traveled through other plasma than the Earth's ionosphere on its way from the quasar to the receiving antenna. It can be seen from table 2 that there is an order-of-magnitude

Plasma	N_e (e/m^3)	f_P (kHz)	(f_P/f_S)	(f_P/f_X)
Earth	10^{12}	8900	$4 \cdot 10^{-3}$	10^{-3}
Interplanetary	$10^7 - 10^8$	28 - 89	$4 \cdot 10^{-5}$	10^{-5}
Interstellar	10^5	3	$1.2 \cdot 10^{-6}$	$3 \cdot 10^{-7}$

Table 2: Plasma effects (Sovers et al., 1998, [70]).

fall off in the plasma frequency contribution as outward distance from the Earth to interplanetary and interstellar regions increases. In a similar way, effects caused by gyration effects due to the presence of a magnetic field can be described. Table 3 contains representative values of magnetic field strength from each of the regimes in units of Gauss. Comparing table 2 to 3 shows that electron gyrofrequency

Magnetic field	B (Gauss)	f_G (kHz)	(f_G/f_S)	(f_G/f_X)
Earth	0.2	600	$3 \cdot 10^{-4}$	$7.5 \cdot 10^{-5}$
Interplanetary	10^{-4}	0.3	$1.5 \cdot 10^{-7}$	$3.2 \cdot 10^{-8}$
Interstellar	10^{-6}	0.003	$1.5 \cdot 10^{-9}$	$3.2 \cdot 10^{-10}$

Table 3: Electron gyrofrequency effects (Sovers et al., 1998, [70]).

is smaller by a factor of 10 than plasma effects near the Earth and negligible in interplanetary and interstellar environments. The collision frequency depends on the number of free electrons and therefore decreases with height. In a height of 90 kilometers a value of about 10^5 s⁻¹ is found, which gives

$$\frac{\nu}{f_S} \approx 4.3 \cdot 10^{-5} \quad \text{and} \quad \frac{\nu}{f_X} \approx 1.2 \cdot 10^{-5} \quad (4.1)$$

which is about six times smaller than the biggest contribution from gyrofrequency effects. In the following subsection it will be shown that the approximations (equation (2.59)) for the index of refraction μ , neglecting collisions between the electrons and considering terms up to second order are accurate enough for most of the space geodetic techniques. Figure 17 sums up the frequency bands covered by VLBI stations of the European VLBI network (EVN). Additionally, the frequencies observed by the Very Large Array (VLA) and the Very Large Baseline Array (VLBA) are added to this figure. As station Wettzell (Germany) is dedicated to geodetic experiments, it is only equipped with S- and X-band receivers (operating at about 2.3 and 8.4 GHz). Although some of the antennas are equipped with MHz receivers, observations are carried out usually in the GHz domain, which has the big advantage that contributions from the ionosphere are expected to be less than at lower frequencies (the first order approximation of the ionosphere is $\sim 1/f^2$). Satellites of the Global Positioning System (GPS) and the Russian counterpart GLONASS are operating in the L-band at frequencies from 1.2 to 1.5 GHz and therefore their signals are expected to be influenced four times the signals used at the lowest geodetic VLBI frequency. Satellite altimetry missions like Topex/Poseidon, Envisat, ERS have their observation frequency allocations in ranges higher than several GHz and mostly use two different wavelengths, like GPS, to get rid of ionospheric contributions. Also geodetic VLBI makes use of dual frequency observations to eliminate ionospheric influences on the delay observables, which will be discussed in section 4.2.

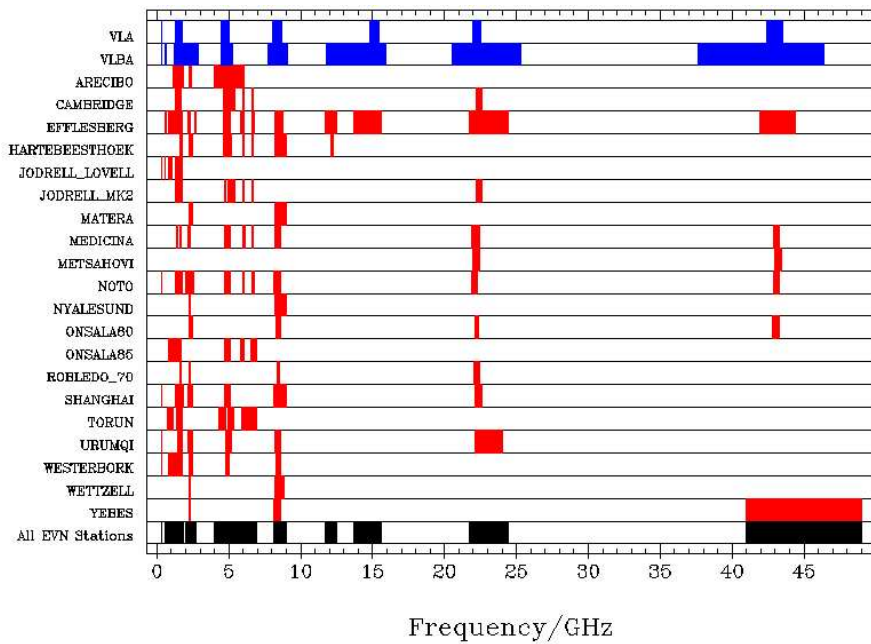


Figure 17: Frequencies covered by the EVN, VLA and VLBA telescopes (EVN, 2005, [20]).

4.1.1 Index of refractivity for phase delay of space geodetic techniques

A simple expression of the ionospheric phase delay was found by equation 2.59, using the so-called isotropic high-frequency approximation, where collision frequency (and synchrotron frequency) effects are neglected as the observing frequency is high enough. This assumption will now be quantitatively verified by using a theoretical height profile of electron density. The location of station Gilcreek ($\varphi_g = 64.83^\circ$, $\lambda_g = 212.51^\circ$) was taken as it is a geodetic VLBI station contributing on a routine basis to global experiments, and because it is located close to the geomagnetic north pole and the Earth's magnetic field strength reaches high values at such latitudes. In addition the magnetic field lines will be nearly parallel to observations in zenith direction, which makes it possible to set the angle $\theta \approx 0$ and subsequently the transversal component \tilde{Y}_T becomes zero, too. Using the IRI 2001 model a theoretical height profile at this station was obtained for June 21st, 2001, 13^h local time (figure 18(a)), and a profile of the Earth's magnetic field strength for that station was gained from a model run using the International Geomagnetic Reference Field (IGRF) (figure 18(b)). The collision frequency ω_C was thought to be decreasing exponentially with height, having a value of 10^5 Hz in a height of 80 km and (nearly) vanishing to 10^2 Hz in 2000 km (figure 18(c)). Together with the fundamental constants the

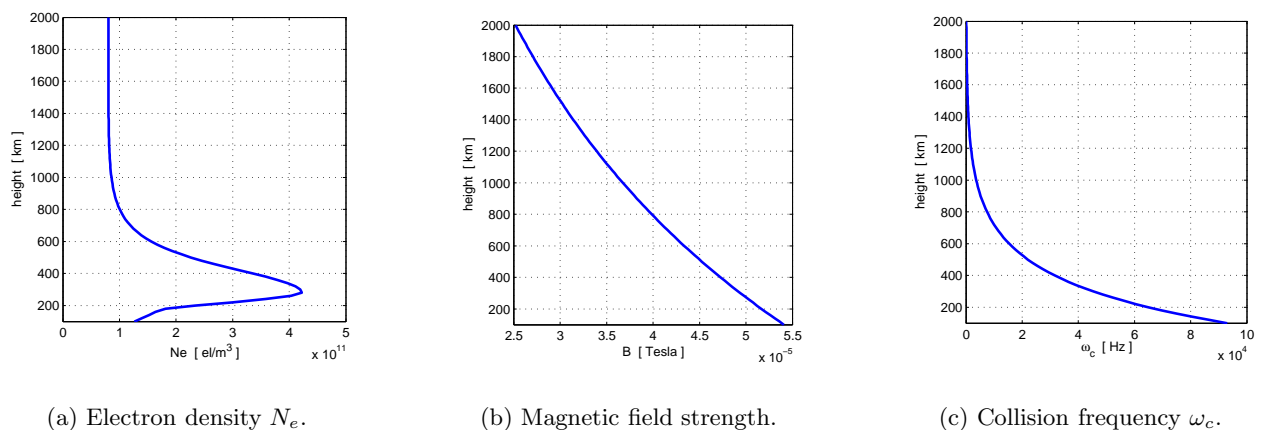


Figure 18: Height profiles for Gilcreek on June 21st, 2001, 13^h local time.

plasma frequency ω_p and the gyrofrequency ω_g can be derived following equations (2.25) and (2.26). Using these parameters, the length of the index vector \vec{n} can be expressed by equation (2.49) split up in real and imaginary part.

$$n^2 = A + jB \quad (4.2)$$

The index of refraction μ and the damping constant χ are related to n^2 in the following way

$$\vec{n} = \vec{\mu} + j\vec{\chi} \Rightarrow n^2 = \vec{\mu} \cdot \vec{\mu} - \vec{\chi} \cdot \vec{\chi} + 2j\vec{\mu} \cdot \vec{\chi} = \mu^2 - \chi^2 + 2j\mu\chi \underbrace{\cos(\vec{\mu}, \vec{\chi})}_C. \quad (4.3)$$

Comparison with equation (4.2) gives

$$A = \mu^2 - \chi^2 \quad B = 2\mu\chi C. \quad (4.4)$$

If it is assumed that the real and imaginary part of the propagation vector \vec{k} are parallel, also $\vec{\mu} \parallel \vec{\chi}$ and $C = \cos(\vec{\mu}, \vec{\chi}) = 1$ are realized. Then the index of refraction and the damping constant are found by solving equation (4.4).

$$\mu^2 = \frac{1}{2} \left(A + \text{sign}(A) \cdot \sqrt{A^2 + B^2} \right) \quad \chi = \frac{B}{2\mu} \quad (4.5)$$

For some applications $B \ll A$ can be assumed, which simplifies the expression of the refractivity index to

$$\mu \approx \sqrt{A} = \sqrt{\mathcal{R}(n^2)}. \quad (4.6)$$

But for the following study equation (4.5) was evaluated for discretized heights and frequencies. As the Appleton-Hartree equation (2.49) gives solutions for the ordinary and extra-ordinary polarized components also the index of refractivity will give two values μ_o and μ_e , whereas the ordinary component is usually treated further, especially if influences from the Earth's magnetic field will be neglected. Figure 19 shows the index of refraction (ordinary and extra-ordinary component) for Gilcreek in zenith direction with the parameters from the theoretical models for frequencies from 1 MHz to 10 GHz. Non-colored regimes represent complex refractivity index, which means that signals with the corresponding frequency will be reflected. For frequencies close to the critical frequency of the medium (at about 10

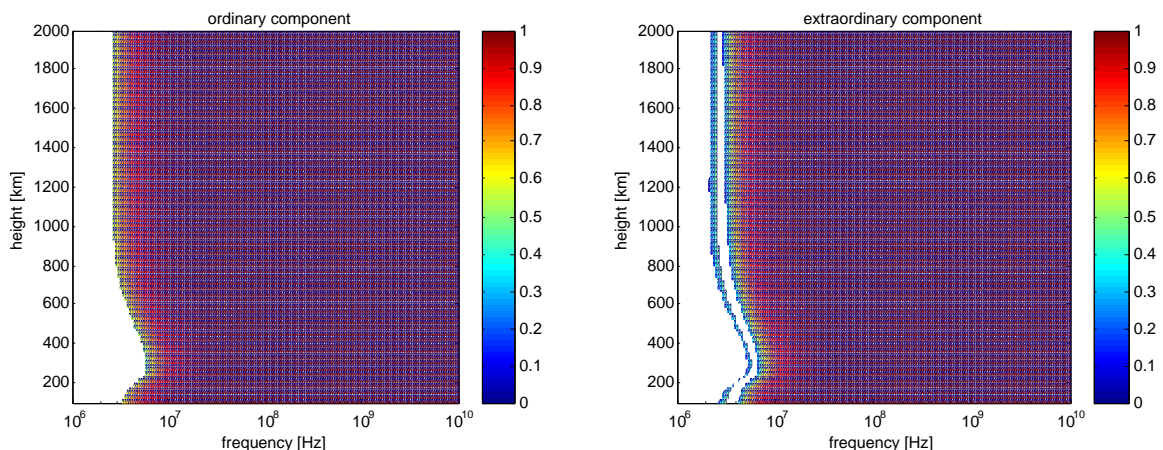


Figure 19: Index of refraction μ for ordinary and extra-ordinary component.

MHz) the ordinary and extra-ordinary solutions for the refractivity index differ, but for frequencies > 100 MHz there seems to be no difference between the ordinary and extra-ordinary component. The impact of an electromagnetic wave propagating at a given frequency $10^6 \leq f \leq 10^{10}$ through this medium will be derived next. Additional delays due to bending of the ray path are neglected and the delay ϱ , expressed in length units, was derived. In fact carrier phases will propagate faster than the speed of light, which is not in contradiction to special relativity, as information can only be sent

by superimposed modulated carrier phases (group delay, see section 4.1.2). The integration over the refraction index

$$\varrho = \int_R^S (\mu - 1) ds \approx \sum_{i=0}^N (\mu_i - 1) \Delta h_i \quad (4.7)$$

was replaced by a summation over using discretized heights. Figure 20 shows the results of this numerical analysis. As expected, for frequencies lower than about 10 MHz no propagation through

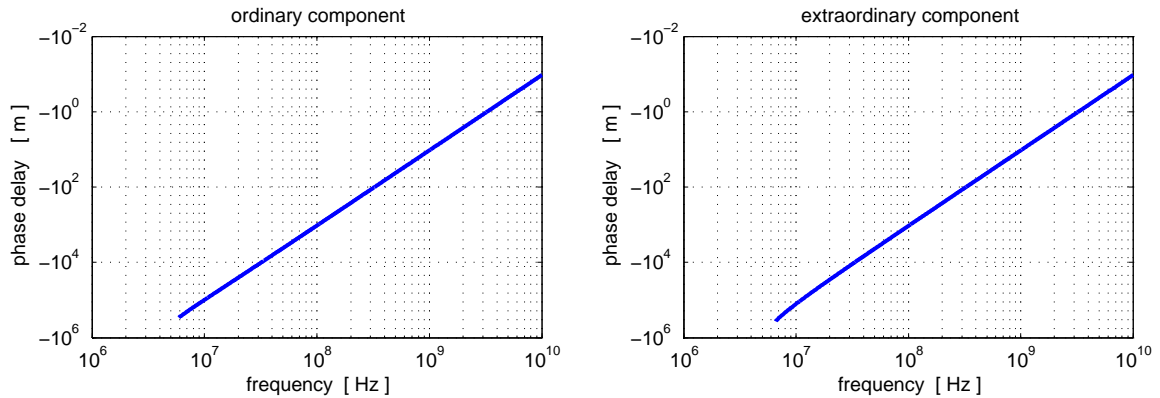
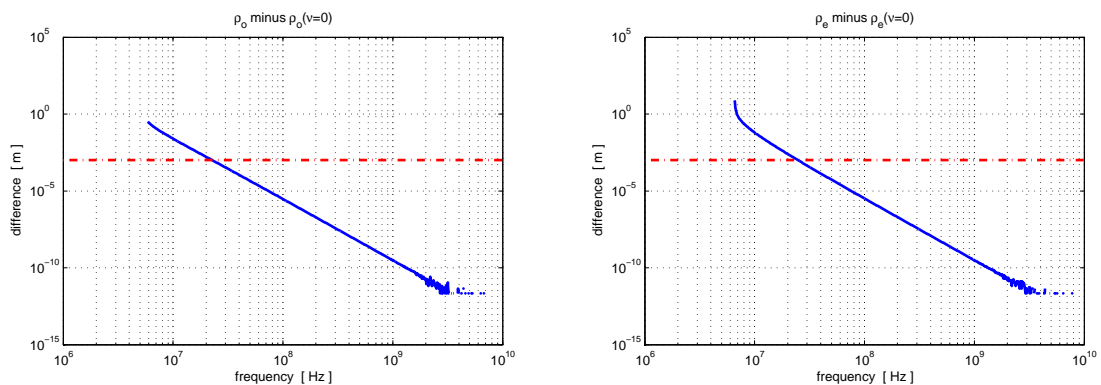


Figure 20: Phase delay expressed in meter for ordinary and extra-ordinary component.

the ionosphere is possible. The logarithmic scale on both axes reveals the strong frequency dependency of the ionospheric delay proportional to $1/f^2$ in a first order approximation, in accordance to equation (2.60). If ionospheric influences would be ignored in GPS measurements, an error of several tens of meters would be introduced. But even for geodetic VLBI the influence would be several meters in S-band and some tens of centimeters in X-band. The theoretical profiles used for this study show rather regular behavior of the ionosphere, so that one can imagine that for a disturbed case the errors can increase by a factor of about three to five. In the next step it was analyzed which parameters of the Appleton-Hartree equation (2.49) can be neglected and how accurate such an approximation would be. For convenience, the delay caused by the ordinary component is abbreviated by ϱ_o , the



(a) Absolute errors in ρ_o .

(b) Absolute errors in ρ_e .

Figure 21: Absolute values of the delay errors due to neglects of the collision frequency for ordinary (left plot) and extra-ordinary components (right plot).

extra-ordinary delay by ϱ_e . In a first step the collision frequency ν is set to zero, the Appleton-Hartree equation is numerically evaluated and plots of figure 21 can be obtained. The dashed red line represent the one millimeter level, which is nowadays the major goal which geometrical geodesy aims to reach by space geodetic observations. It can be seen that even for measurement below 100 MHz the neglect of the collision frequency is permitted and that for frequencies higher than 1 GHz its impact is far

below any detection level. Even computation precision is not sufficient to find a difference between the original and the simplified model for frequencies higher than about 2 GHz, which is the reason for non-continuous lines in the logarithmic plots. When collisions between the electrons and effects due to the presence of magnetic field are neglected the plots of figure 22 are obtained. Again, the dashed

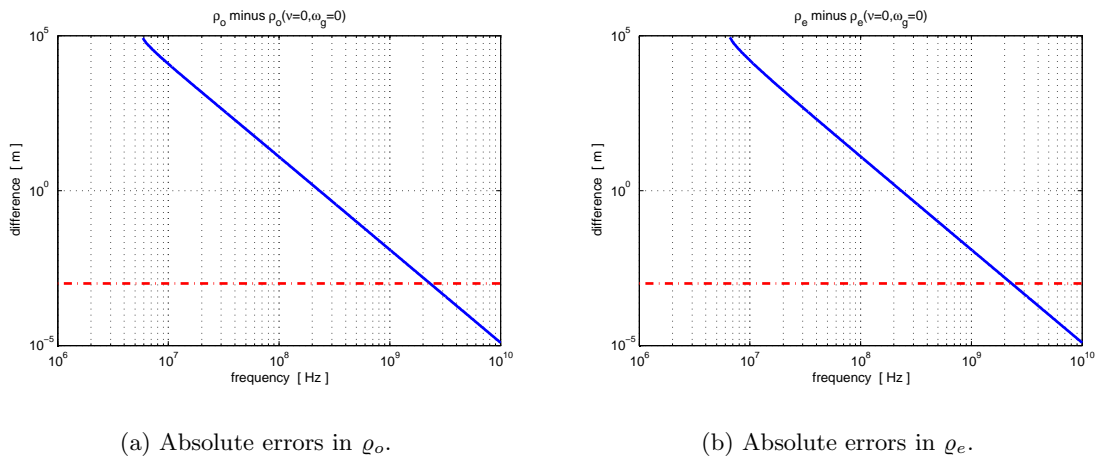


Figure 22: Absolute values of the delay errors due to neglections of collision and gyro frequencies for ordinary (left plot) and extra-ordinary components (right plot).

red lines give the mm-level. This means that the treatment of the ionosphere, taking only plasma frequency effects into account, is not sufficient for space geodetic techniques operating at frequencies below 2 GHz to obtain mm-accuracy. Although in GPS observation some errors are eliminated due to the combination of dual frequency measurements a noticeable amount of ionospheric delay biases still remains in GPS analysis (Kedar et al., 2003, [40]). Hawarey et al. (2005, [30]) have shown that this is not the case for geodetic VLBI observations, and that such effects cannot be distinguished from measurement noise. Finally, as already denoted by equation (2.61), the phase delay can be set in a linear relationship to the slant total electron content (i.e. in this case in the vertical direction, therefore also called vertical total electron content - VTEC) by

$$\varrho \approx -\frac{40.28}{f^2} VTEC \quad (4.8)$$

where

$$VTEC = \int_S^R N_e ds \approx \sum_{k=0}^N N_{e,k} \Delta h_k \quad (4.9)$$

was taken to derive the sum of the electrons along the ray path. Comparison of this approximation to the full Appleton-Hartree expression are given in figure 23. For the extra-ordinary component shown in figure 23(b) a similar result like the one given in figure 22(b) can be detected. For the ordinary component there is a slightly different behavior, which can be explained by the fact that absolute differences are plotted. For frequencies lower than about 1.8 GHz the linear approximation underestimates the delay, but for higher frequencies overestimates it slightly. But generally there is the same conclusion as before (when ν and ω_g were neglected) that for GPS measurements the linear approximation, using only slant total electron content values will be not sufficient for mm-accuracy. This is in accordance to what is mentioned in the paper by Kedar et al. (2003, [40]). On the other hand it is concluded that the simple approximation is sufficient for the analysis of phase delay of geodetic VLBI experiments when frequencies > 2 GHz are used. Even for extreme ionospheric conditions errors due to this approach will not exceed about two millimeters. For the sake of completeness also values of the damping constant χ were derived and plotted in figure 24. This reveals that for space geodetic techniques energy dissipation is rather small for frequencies > 1 GHz, which enables smaller transmitting power for navigation satellites and altimetry missions.

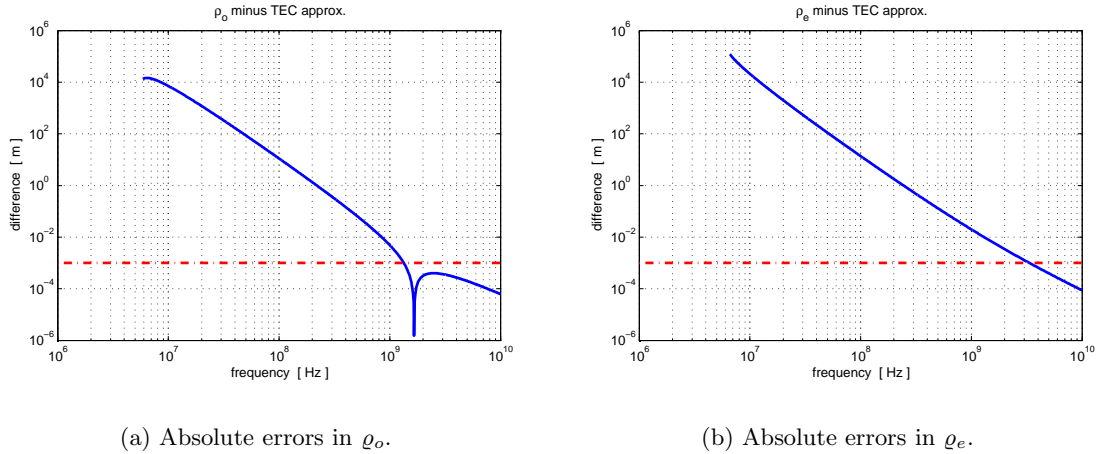


Figure 23: Absolute values of the delay errors due to approximation with total electron content for ordinary (left plot) and extra-ordinary components (right plot).

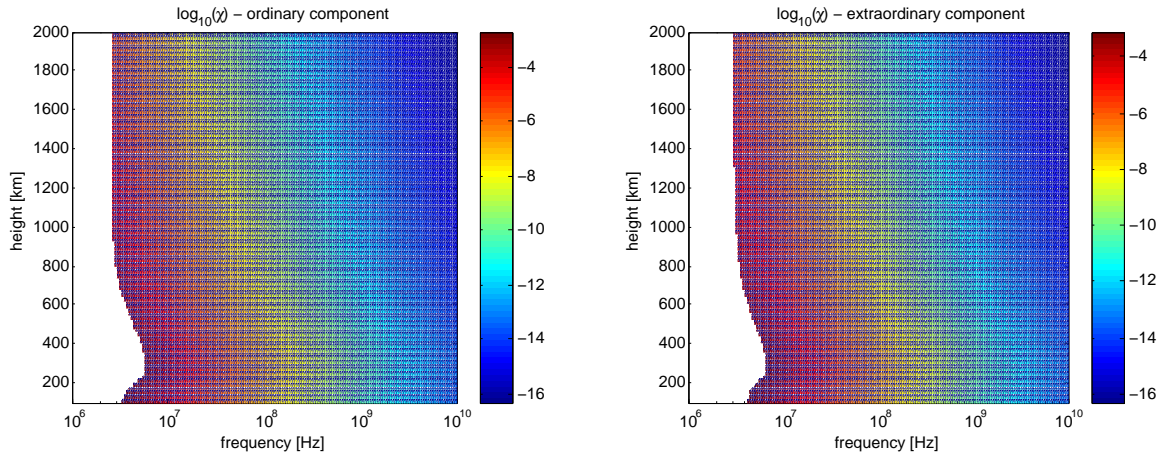


Figure 24: Logarithmic values of the damping constant for ordinary and extra-ordinary component.

4.1.2 Index of refractivity for group delay to space geodetic techniques

If collision and gyro frequency effects are neglected a simple relation for the index of refraction can be found by

$$\mu^2 = 1 - X = 1 - \frac{\omega_P^2}{\omega^2} \quad (4.10)$$

which makes further computations rather simple. The refractivity index for group propagation can be gained when the right side of equation (2.55) is evaluated.

$$\mu\mu_g = \frac{\partial(\omega^2\mu^2)}{\partial(\omega^2)} = 1 \quad (4.11)$$

Therefore the group refractivity index is given by

$$\mu_g = \frac{1}{\mu} = \frac{1}{\sqrt{1 - \frac{\omega_P^2}{\omega^2}}} \approx 1 + \frac{1}{2} \frac{\omega_P^2}{\omega^2} + \frac{3}{8} \left(\frac{\omega_P^2}{\omega^2}\right)^2 + \frac{5}{16} \left(\frac{\omega_P^2}{\omega^2}\right)^3 + \frac{35}{128} \left(\frac{\omega_P^2}{\omega^2}\right)^4 + \dots \quad (4.12)$$

If only first order terms are taken, a similar expression like the one for (phase) refraction index, but with a different leading sign can be found.

$$\mu_g = 1 + \frac{1}{2} \frac{\omega_P^2}{\omega^2} = 1 + \frac{1}{2} \frac{f_P^2}{f^2} = 1 + A \frac{N_e}{f^2} \quad A = 40.28 \quad (4.13)$$

Like in section 2.5.2 it is possible to describe the additional group delay ϱ_g caused by the presence of the ionosphere in linear dependency to the slant total electron content.

$$\varrho_g = \int_S^R (\mu_g - 1) ds = \frac{A}{f^2} \int_S^R N_e ds = \frac{A}{f^2} STEC \quad (4.14)$$

Like in the previous section, this approximation is validated against the full Appleton-Hartree equation, which has to be evaluated, differentiated and numerically integrated using the ionospheric state variables given in figure 18. The refractivity index for group delay observations is always > 1 , which can also be seen in figure 25. The group refractivity index has a smooth behavior and reaches only

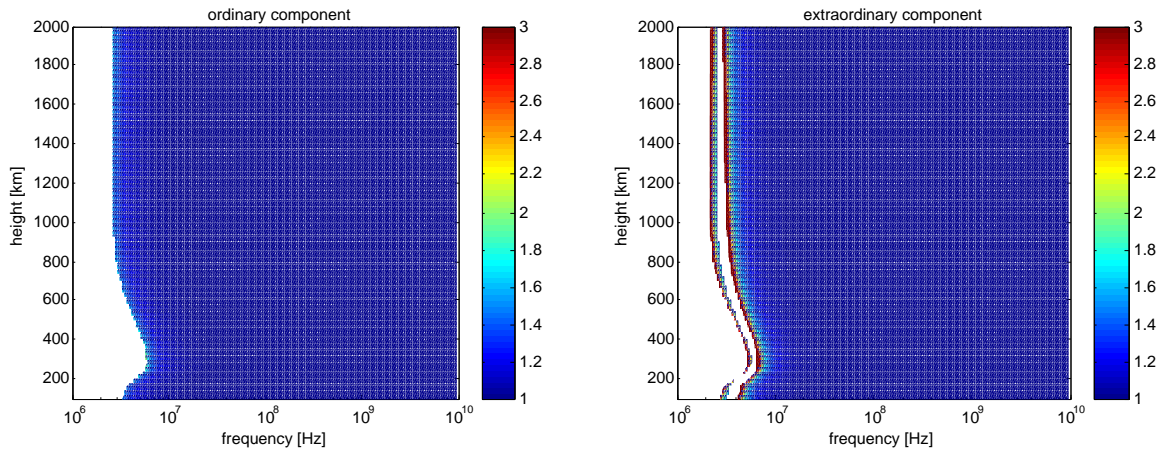
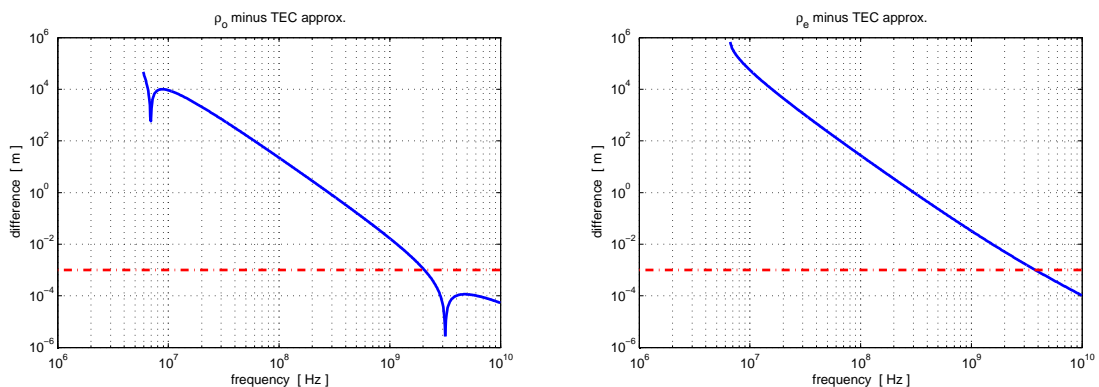


Figure 25: Group refractivity index μ_g for ordinary and extra-ordinary component.

rather small values for the ordinary component of propagation, but the extra-ordinary one shows much higher values and shows that propagation over a short vertical distance is possible at a certain frequency even if at the lower and higher levels it will be prevented by the medium. The comparison between the exact solutions by numerical evaluation of the Appleton-Hartree equation and TEC approximation, defined by equation (4.14), is given in figure 26. As absolute differences are plotted, the



(a) Absolute errors in ϱ_o .

(b) Absolute errors in ϱ_e .

Figure 26: Absolute values of the group delay errors due to approximation with total electron content for ordinary (left plot) and extra-ordinary components (right plot).

left blue curve (ordinary component, figure 26(a)) shows two local minima, caused by underestimation and overestimation of the ionospheric group delay. Again as for phase delay observables, there will be an approximation error of up to one centimeter for the GPS frequencies, whereas geodetic VLBI observations are only influenced at sub-mm level. The accuracy of the approximation for the

extra-ordinary group delay component (figure 26(b)) is slightly worse but still sufficient for geodetic VLBI experiments. Generally it can be stated that the first order approximation, taking only into account the plasma frequency, is sufficient for all VLBI observations, when frequencies are chosen to be higher than 2 GHz. For GPS the neglect of the gyrofrequency (i.e. effect due to the presence of a magnetic field) has a noticeable influence on the measurement accuracy and should be considered in the analysis procedures as described by Kedar et al. (2003, [40]).

4.2 Treatment of ionospheric delays in VLBI processing

4.2.1 Ionospheric contribution to delays

In each channel i fringe phases are rotated by

$$\Delta\phi = -\frac{\alpha}{\omega_i} \quad (4.15)$$

where α is given by

$$\alpha = \frac{40.28}{c} \left(\int N_e ds_1 - \int N_e ds_2 \right) = \frac{40.28}{c} (STEC_1 - STEC_2) . \quad (4.16)$$

The speed of light c is used for conversion to time delay, s_1 and s_2 are the paths of wave propagation from the source to the first and second station of the radio interferometer. This means that VLBI is only sensitive to differences in the ionospheric conditions. Thereafter a fringe phase in channel i can also be described by

$$\phi_i = \tau_{ph}\omega_0 + \tau_{gr}(\omega_i - \omega_0) - \frac{\alpha}{\omega_i} \quad (4.17)$$

where ω_0 is a reference sky frequency. Unknown quantities phase and group delay τ_{ph} and τ_{gr} can be determined from equation (4.17) using a weighted least squares adjustment, where the linearized model is given by

$$\tau_{ph} \frac{\partial\phi_i}{\partial\tau_{ph}} + \tau_{gr} \frac{\partial\phi_i}{\partial\tau_{gr}} = \phi_i + \frac{\alpha}{\omega_i} . \quad (4.18)$$

As described in section 3.3.3 (single band) phase and (multi-band) group delays are obtained within the fringe fitting procedure by finding the values which maximize the delay resolution function. It can be shown (see e.g. Sekido (2001,[65])) that bandwidth synthesis is equivalent to the least squares solution defined in equation (4.18), if the correlation amplitude of each channel corresponds to the weight of phase observables at each channel in the least squares adjustment. Additional residual phases are assumed not to be too large. If residual phases are less than 0.4 rad the difference between both approaches is smaller than 0.1 ps. An explicit expression for the solution of a least squares adjustment problem can be denoted by

$$\hat{\mathbf{x}} = (\mathbf{A}^T \mathbf{P} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{P} \mathbf{l} \quad (4.19)$$

where \mathbf{A} is the design matrix, $\hat{\mathbf{x}}$ are the (unknown) parameters (i.e. phase and group delay), \mathbf{l} are the residuals (or the right side of equation (4.18)) and \mathbf{P} is the weight matrix, having the squared weights \mathbf{r}_i in its diagonal. The necessary partial derivatives can be found easily by comparison between equation (4.17) and (4.18).

$$\frac{\partial\phi_i}{\partial\tau_{ph}} = \omega_0 \quad \frac{\partial\phi_i}{\partial\tau_{gr}} = \omega_i - \omega_0 \quad (4.20)$$

The inversion of the symmetric 2×2 normal equation matrix is formally given by

$$\begin{pmatrix} N_{11} & N_{12} \\ N_{12} & N_{22} \end{pmatrix}^{-1} = \begin{pmatrix} \frac{N_{22}}{\Delta} & -\frac{N_{12}}{\Delta} \\ -\frac{N_{12}}{\Delta} & \frac{N_{11}}{\Delta} \end{pmatrix} \quad \Delta = N_{11}N_{22} - N_{12}^2 \quad (4.21)$$

which makes it possible to set up the necessary parts of the normal equation system. Taking observations from N channels, the components of the normal equation matrix can be explicitly denoted

$$\begin{aligned} N_{11} &= \omega_0^2 \sum_{i=1}^N r_i \\ N_{12} &= \omega_0 \sum_{i=1}^N r_i (\omega_i - \omega_0) \\ N_{22} &= \sum_{i=1}^N r_i (\omega_i - \omega_0)^2 \end{aligned} \quad (4.22)$$

and the solution of equation (4.20) can be stated in the following way.

$$\begin{pmatrix} \tau_{ph} \\ \tau_{gr} \end{pmatrix} = \begin{pmatrix} \frac{N_{22}}{\Delta} & -\frac{N_{12}}{\Delta} \\ -\frac{N_{12}}{\Delta} & \frac{N_{11}}{\Delta} \end{pmatrix} \cdot \begin{pmatrix} \omega_0 \sum_{i=1}^N r_i \phi_i + \alpha \omega_0 \sum_{i=1}^N \frac{r_i}{\omega_i} \\ \sum_{i=1}^N r_i (\omega_i - \omega_0) \phi_i + \alpha \omega_0 \sum_{i=1}^N r_i \frac{\omega_i - \omega_0}{\omega_i} \end{pmatrix} \quad (4.23)$$

After some algebra an explicit solution for phase and group delay can be found.

$$\tau_{ph} = \frac{\sum_{i=1}^N r_i (\omega_i - \omega_0)^2 \cdot \sum_{i=1}^N r_i \phi_i - \sum_{i=1}^N r_i (\omega_i - \omega_0) \cdot \sum_{i=1}^N r_i (\omega_i - \omega_0) \phi_i}{\omega_0 \left(\sum_{i=1}^N r_i \cdot \sum_{i=1}^N r_i (\omega_i - \omega_0)^2 - \left(\sum_{i=1}^N r_i (\omega_i - \omega_0) \right)^2 \right)} - \frac{\alpha}{\omega_{ph}^2} \quad (4.24)$$

$$\tau_{gr} = \frac{\sum_{i=1}^N r_i \sum_{i=1}^N r_i (\omega_i - \omega_0) \phi_i - \sum_{i=1}^N r_i (\omega_i - \omega_0) \cdot \sum_{i=1}^N r_i \phi_i}{\left(\sum_{i=1}^N r_i \cdot \sum_{i=1}^N r_i (\omega_i - \omega_0)^2 - \left(\sum_{i=1}^N r_i (\omega_i - \omega_0) \right)^2 \right)} + \frac{\alpha}{\omega_{ph}^2} \quad (4.25)$$

where the so-called effective ionosphere frequencies ω_{ph} and ω_{gr} are defined by

$$\omega_{ph} = \sqrt{\frac{\omega_0 \frac{\sum_{i=1}^N r_i \cdot \sum_{i=1}^N r_i (\omega_i - \omega_0)^2 - \left(\sum_{i=1}^N r_i (\omega_i - \omega_0) \right)^2}{N}}{\sum_{i=1}^N r_i (\omega_i - \omega_0) \cdot \sum_{i=1}^N r_i \frac{\omega_i - \omega_0}{\omega_i} - \sum_{i=1}^N r_i (\omega_i - \omega_0)^2 \cdot \sum_{i=1}^N \frac{r_i}{\omega_i}}} \quad (4.26)$$

$$\omega_{gr} = \sqrt{\frac{\frac{\sum_{i=1}^N r_i \cdot \sum_{i=1}^N r_i (\omega_i - \omega_0)^2 - \left(\sum_{i=1}^N r_i (\omega_i - \omega_0) \right)^2}{N}}{\sum_{i=1}^N r_i \cdot \sum_{i=1}^N r_i \frac{\omega_i - \omega_0}{\omega_i} - \sum_{i=1}^N r_i (\omega_i - \omega_0) \cdot \sum_{i=1}^N \frac{r_i}{\omega_i}}} \quad (4.27)$$

The effective ionosphere frequencies have a clear physical meaning: if the wide-band signal is replaced by a quasi-monochromatic signal the contribution to group or phase delay will be the same. Equations (4.24) and (4.25) provide the theoretical basis for the treatment of multi-band delays (and their ionospheric contributions) in the same way as it was done for single frequency observations in the prior section. Instead of the observing frequency the effective ionosphere frequency, computed from the frequency distribution, has to be taken to express the ionospheric contribution (measured in TECU) to units of time. Additionally, the linearity of equations (4.24) and (4.25) makes it possible to eliminate ionospheric influences when measurements are carried out at two separate bands.

4.2.2 Ionosphere free linear combinations

Nowadays any geodetic VLBI experiment is carried out at two distinct frequency bands in order to correct for ionospheric influences. Taking the standard bands (X- and S-band) for such experiments gives 4 observables, each of them containing the ionospheric free delay τ_{if} (which will be the input for any precise geodetic analysis) and a contribution α from the ionosphere, scaled by the corresponding effective ionosphere frequencies.

$$\begin{aligned}\tau_{gx} &= \tau_{if} + \frac{\alpha}{\omega_{gx}^2} \\ \tau_{gs} &= \tau_{if} + \frac{\alpha}{\omega_{gs}^2} \\ \tau_{px} &= \tau_{if} - \frac{\alpha}{\omega_{px}^2} \\ \tau_{ps} &= \tau_{if} - \frac{\alpha}{\omega_{ps}^2}\end{aligned}\tag{4.28}$$

Here the first letter in the indices stands for group or phase delay and the second letter represents X- or S-band. Using these equations the unknown parameter α can be eliminated and the ionospheric free delay observable can be obtained. This is carried out by a simple linear combination between two of the expressions, given in (4.28). The most important one is only considering group delay measurements and yields

$$\tau_{if} = \frac{\omega_{gx}^2}{\omega_{gx}^2 - \omega_{gs}^2} \tau_{gx} - \frac{\omega_{gs}^2}{\omega_{gx}^2 - \omega_{gs}^2} \tau_{gs}.\tag{4.29}$$

The right part of equation (4.29) can be considered as the observable, from which all geodetic target parameters (see section 3.4.2) can be determined. One can also form the combinations

$$\tau_{if} = \frac{\omega_{px}^2}{\omega_{px}^2 + \omega_{gs}^2} \tau_{px} + \frac{\omega_{gs}^2}{\omega_{px}^2 + \omega_{gs}^2} \tau_{gs}\tag{4.30}$$

or

$$\tau_{if} = \frac{\omega_{px}^2}{\omega_{px}^2 + \omega_{gx}^2} \tau_{px} + \frac{\omega_{gx}^2}{\omega_{px}^2 + \omega_{gx}^2} \tau_{gx}\tag{4.31}$$

to get the so-called semi-phase observables. This means that the ionospheric influence on phase delay measurements is corrected by group delay observations, using the same (equation (4.31)) or the other band (equation (4.30)). Anyway, in all three cases, remaining ambiguities have to be fixed, which can be done on group delay measurements rather easily as already mentioned in section 3.3.3. Until now, ambiguities fixing for (semi-) phase delay measurements can only be done for rather short baselines, i.e. less than 1000 km. Instead of computing the ionosphere-free linear combination (4.29) one can also compute the ionospheric contribution in X-band

$$\tau_{igx} = \frac{\alpha}{\omega_{gx}^2} = -\frac{\omega_{gs}^2}{\omega_{gx}^2 - \omega_{gs}^2} (\tau_{gx} - \tau_{gs})\tag{4.32}$$

add it to the theoretical delay and thus “correct” or “calibrate” the group delay at X-band. This approach should not be applied as the observable would be corrected using the measurement of itself. In order to avoid this logical pitfall it is preferable to use the concept of the ionosphere-free linear combinations. Anyway, for the sake of completeness also the ionospheric correction (4.32) is usually stored in databases together with all the other information, which is necessary for geodetic analysis.

4.3 Instrumental biases

In fact, real observations do not exactly correspond to equations (4.28), but rather contain an extra delay term caused by instrumental imperfectness. As mentioned by Ray and Corey (1991, [57]) an

additional delay is caused by instrumental delays in the different bands, which change the delay observables to

$$\begin{aligned}
\tau'_{gx} &= \tau_{if} + \frac{\alpha}{\omega_{gx}^2} + \tau_{inst,x} \\
\tau'_{gs} &= \tau_{if} + \frac{\alpha}{\omega_{gs}^2} + \tau_{inst,s} \\
\tau'_{px} &= \tau_{if} - \frac{\alpha}{\omega_{px}^2} + \tau_{inst,x} \\
\tau'_{ps} &= \tau_{if} - \frac{\alpha}{\omega_{ps}^2} + \tau_{inst,s}
\end{aligned} \tag{4.33}$$

When the ionospheric-free linear combination (4.29) is evaluated a biased delay τ'_{if} is obtained.

$$\begin{aligned}
\tau'_{if} &= \frac{\omega_{gx}^2}{\omega_{gx}^2 - \omega_{gs}^2} \tau'_{gx} - \frac{\omega_{gs}^2}{\omega_{gx}^2 - \omega_{gs}^2} \tau'_{gs} \\
&= \tau_{if} + \underbrace{\frac{\omega_{gx}^2}{\omega_{gx}^2 - \omega_{gs}^2} \tau_{inst,x} - \frac{\omega_{gs}^2}{\omega_{gx}^2 - \omega_{gs}^2} \tau_{inst,s}}_{\hat{\tau}}
\end{aligned} \tag{4.34}$$

where the notation $\hat{\tau}$ was used to express overall instrumental delay, caused by a weighted difference between X- and S-band receiving system delays. Although one might think that this would cause a problem in further processing steps, geodetic analysis is not affected by these instrumental delays. As long as instrumental delays do not change between the scans there will be no impact on geodetic results. They can be treated as a constant bias of the delay measurements, independent of azimuth and zenith distance and are absorbed into the clock models (Ray and Corey (1991, [57])). Also the computed ionospheric correction for X-band group delay measurements (4.32) has to be replaced now by the intrinsic one ($\tau'_{ig,x}$), including the receiving system biases.

$$\begin{aligned}
\tau'_{ig,x} &= -\frac{\omega_{gs}^2}{\omega_{gx}^2 - \omega_{gs}^2} (\tau'_{gx} - \tau'_{gs}) \\
&= \frac{\alpha}{\omega_{gx}^2} + \underbrace{\frac{\omega_{gs}^2}{\omega_{gx}^2 - \omega_{gs}^2} (\tau_{inst,s} - \tau_{inst,x})}_{\tau_{inst}}
\end{aligned} \tag{4.35}$$

The scaled difference between S- and X-band instrumental delay, denoted by τ_{inst} , is always contained in the X-band ionospheric correction. Equation (4.35) establishes a basis for determination of ionospheric parameters from X-band ionospheric corrections, which will be derived in section 5. If $\hat{\tau}$ is constant over time, also τ_{inst} will have a fixed, non-varying value and all time-dependent variations in the X-band ionospheric correction can be assigned to variations in α , which is basically the difference of two slant total electron content measurements (equation (4.16)).

4.4 Treatment of interplanetary and interstellar plasma

As discussed in sections 4.1.1 and 4.1.2 it is sufficient in most of the cases, when geodetic VLBI experiments are carried out, that the ionospheric delay can be approximated by the slant total electron content (STEC). From equation (2.61) STEC is defined as the sum of electrons along the ray path from the transmitter to the receiver. In the case of GPS this includes the ionosphere and parts of the magnetosphere, having a height of about 20000 km. For VLBI the traversed ray path does not only include the ionosphere and the magnetosphere but also considers interplanetary and interstellar plasma, as the “sender” is a radio source far away from the Earth. Even if the latter two environments might have lower plasma densities and weaker magnetic fields the path lengths through these regimes will be long enough so that there will be noticeable impact on the propagation speed of the electromagnetic

waves. Without loss of generality the slant total electron content can be split up in parts, which are caused by the Earth's environment (ionosphere, magnetosphere), and other ones taking into account interplanetary and interstellar plasma. Then STEC is given by

$$STEC' = \int_R^{B_0} N_e ds + \sum_{i=1}^M \int_{L_i}^{U_i} N_e ds = STEC + \sum_{j=1}^M STEC_j^* \quad (4.36)$$

where M regions treated as plasma, with an extent from L_i to U_i along the ray path have been considered and $STEC'$ represents the slant total electron content in the Earth's surrounding. VLBI is only sensitive to differences in the delay, as expressed by equation (4.16), the parameter α can be denoted by

$$\begin{aligned} \alpha &= \frac{40.28}{c} (STEC'_1 - STEC'_2) \\ &= \frac{40.28}{c} \left[(STEC_1 - STEC_2) + \sum_{j=1}^M (STEC_{1,j}^* - STEC_{2,j}^*) \right]. \end{aligned} \quad (4.37)$$

Figure 27 represents only one interstellar plasma region and two VLBI stations separated by the maximum distance of the Earth's diameter. As the extension of interstellar plasma is usually much

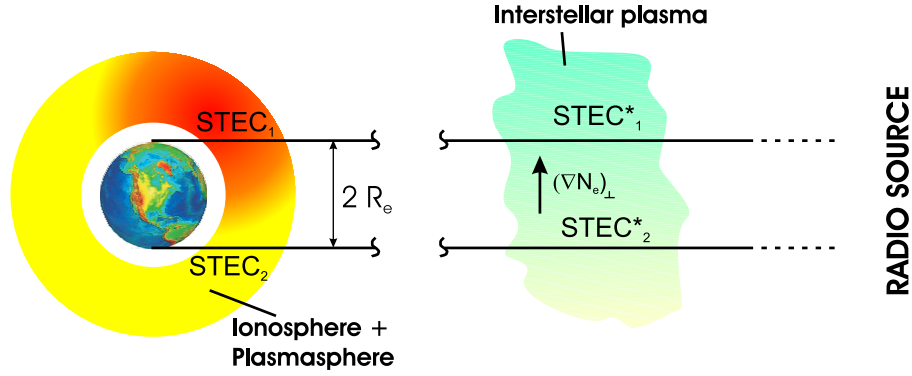


Figure 27: Interstellar plasma and its contribution to the measured ionospheric delay.

larger than one solar radius it can be assumed that both ray paths enter and leave the medium at approximately the same boundaries L and U , which gives for the situation described in figure 27 the following expression.

$$\begin{aligned} \alpha &= \frac{40.28}{c} \left[(STEC_1 - STEC_2) + \int_{L_1}^{U_1} N_{e,1} ds - \int_{L_2}^{U_2} N_{e,2} ds \right] \\ &= \frac{40.28}{c} \left[(STEC_1 - STEC_2) + \int_L^U (N_{e,1} - N_{e,2}) ds \right] \quad \text{for } U = U_1 \approx U_2, L = L_1 \approx L_2 \end{aligned} \quad (4.38)$$

The component of the electron density gradient (inside the interstellar plasma) orthogonal to the propagation direction and inside the plane, spanned by both ray paths, is notated by $(\nabla N_e)_\perp$ (see figure 27). Then the last integral in equation (4.38) will disappear if

$$(\nabla N_e)_\perp \cdot 2R_e \approx 0 \quad (4.39)$$

is fulfilled. This means that interstellar plasma effects can be neglected if electron densities don't change noticeably within the distance of the Earth's diameter. This is fulfilled for nearly all interplanetary and interstellar regimes and therefore the expression α contains only contributions from the Earth's ionosphere and the magnetosphere.

$$\alpha = \frac{40.28}{c} (STEC_1 - STEC_2) \quad (4.40)$$

Nevertheless, the interplanetary and interstellar environments have a noticeable impact on radio astronomical observations as described in Thompson et al. (2001, [75]).

4.5 The thin-shell approximation and its mapping function

Most of the regional and global models, which were determined from space geodetic measurements, consist of two-dimensional maps of the total electron content along the vertical direction (so-called VTEC). For the following discussions the upper limit of the ionosphere, located at height H_1 and the lower limit situated at height H_2 have been set. Figure 28 represents the basic geometry, for deriving the thin-shell model. The signal comes from any outer region, enters the ionosphere at S_1 , leaves it at S_2 , and reaches the receiving station E at a zenith distance z . If it is now assumed that all electrons inside the ionosphere can be condensed to an infinitesimally thin layer at height H (dashed line in figure 28), the signal will intersect with this shell at point P . This intersection is called ionospheric pierce point and its projection on the Earth's surface is named sub-ionospheric point Q . If the thin-

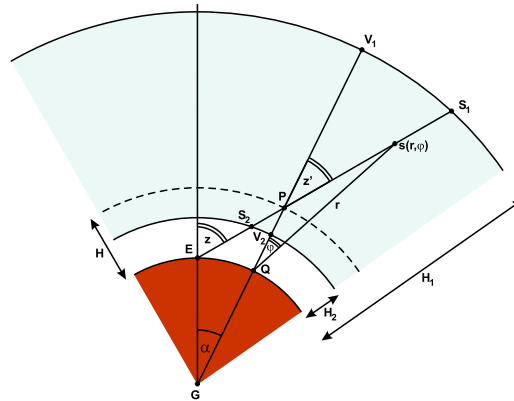


Figure 28: Path of the signal along the ionosphere to the ground receiver.

shell approximation holds, it will be possible to find a relation between the STEC measured at station E and the vertical total electron content between V_1 and V_2 above the sub-ionospheric point. The relation between STEC and VTEC is given by a mapping function

$$M = \frac{STEC}{VTEC} = \frac{\int_{S_1}^{S_2} N_e ds}{\int_{V_1}^{V_2} N_e dh} \quad (4.41)$$

in accordance with the situation described in figure 28. Introduction of a polar coordinate system with its origin in the sub-ionospheric point Q and an X -axis direction towards the ionospheric pierce point P makes the following considerations more transparent. Any point on the oblique path $S_1 - S_2$ can be described by $s(r, \theta)$ and the electron density at this location will be denoted by $N_e(r, \theta)$. Therefore the coordinate r describes the changes along the vertical direction, while the coordinate θ reveals horizontal gradients of the N_e . The first of these variations is more significant than the second, which permits an approximation of $N_e(r, \theta)$ by a Taylor expansion up to first order in θ .

$$N_e(r, \theta) = N_e(r, 0) + \left. \frac{\partial N_e(r, \theta)}{\partial \theta} \right|_{\theta=0} \theta + \dots = N_e(r, 0) + f(r)\theta + \dots \quad (4.42)$$

where $f(r)$ is the derivative of $N_e(r, \theta)$ with respect to θ , evaluated for $\theta = 0$. This approximation can be used to calculate the integral in the numerator of equation (4.41).

$$STEC = \int_{S_1}^{S_2} N_e(r, \theta) ds = \int_{S_1}^{S_2} N_e(r, 0) ds + \int_{S_1}^{S_2} f(r)\theta ds + \dots \quad (4.43)$$

From figure 28 it can be concluded that

$$\frac{V_1 - V_2}{S_1 - S_2} \approx \cos z' \quad (4.44)$$

where z' is the zenith distance at the ionospheric pierce point P . This means that

$$\int_{S_1}^{S_2} N_e(r, 0) ds = \frac{1}{\cos z'} \int_{V_1}^{V_2} N_e(r, 0) ds = \frac{1}{\cos z'} VTEC \quad (4.45)$$

which leads to

$$STEC = \frac{1}{\cos z'} VTEC + \int_{S_1}^{S_2} f(r) \theta ds + \dots \quad (4.46)$$

If the contribution of the horizontal gradients to the electron density distribution is neglected, the following approximation for the mapping function is finally obtained.

$$M = \frac{STEC}{VTEC} \approx \frac{1}{\cos z'} \quad (4.47)$$

If signal bending can be neglected, z' can be expressed as a function of the zenith distance of the observed signal z , using simple geometry on the triangle GEP (figure 28).

$$\sin z' = \frac{R_e}{R_e + H} \sin z \quad (4.48)$$

which gives

$$M(z) = \frac{1}{\sqrt{1 - \left(\frac{R_e}{R_e + H} \sin z \right)^2}} \quad (4.49)$$

using the mean Earth's radius R_e and the height of the single layer H . The later one is crucial for determination of VTEC values, as the observed delay is compared to the modeled delay (which has to be adjusted) multiplied by the mapping function. Then, using a wrong height of the ionospheric thin layer (greater or lower than the correct value) produces a wrong estimation of the VTEC. E.g. an increase in H produces a decrease in the mapping function and an increase in the estimated VTEC. These variations depend on the zenith distance of the observed signal and what can lead to systematic variations of the residuals correlated with the zenith distance, and therefore changing the mean square residual error. Comparisons by Coco et. al (1988, [15]) have shown that the mapping function as introduced in equation (4.49) has an error of 10% (RMS) at a zenith distance of 80° , and a maximum error of 20% at the same angle. At 60° of zenith distance the RMS error was already less than 3%, with a maximum of 6%. The authors conclude that the mapping function can be used with a wide variety of ionospheric profiles without introducing large errors. Usually the height of the thin-shell layer is chosen between 350 and 500 km approximately equal to the height of the F2 layer. Schaer (2005, [62]) has proposed to use the so-called Modified Single-Layer Model Mapping function, which was derived from a best fit with respect to the Jet Propulsion Laboratory (JPL) extended slab model (ESM) mapping function. The height of the single layer is set $H = 506.7$ km and a scaling parameter $\alpha = 0.9782$ is introduced so that the new mapping $M'(z)$ function is given by

$$M'(z) = M(\alpha z) = \frac{1}{\sqrt{1 - \left(\frac{R_e}{R_e + H} \sin \alpha z \right)^2}} \quad (4.50)$$

5 Determination of ionospheric parameters from group delay measurements

As denoted by equation (4.35) in section 4.2.2 each ionospheric delay correction consists of two contributions

- a time-variable part, from the variation of the ionosphere. It is expressed by the portion α (equation (4.40)) and gives the difference of the ionospheric characteristics between the stations.
- a (constant) instrumental bias for each baseline, evoked by instrumental conditions at each station.

Therefore the ionospheric correction τ'_{igx} on a baseline between station m and n can be rewritten, when instrumental biases are assigned to each station.

$$\tau'_{igx,mn}(t) = \frac{40.28}{c\omega_{gx,mn}^2} [STEC_m(t) - STEC_n(t)] + \tau_{inst,m} - \tau_{inst,n} \quad (5.1)$$

One has to be aware that the effective ionospheric frequency ω_{gx} has to be computed for each scan and baseline, as the loss of channels at any station will influence the value of ω_{gx} . Figure 29 describes the situation denoted by equation (5.1), which can be seen as the basic principle for the following

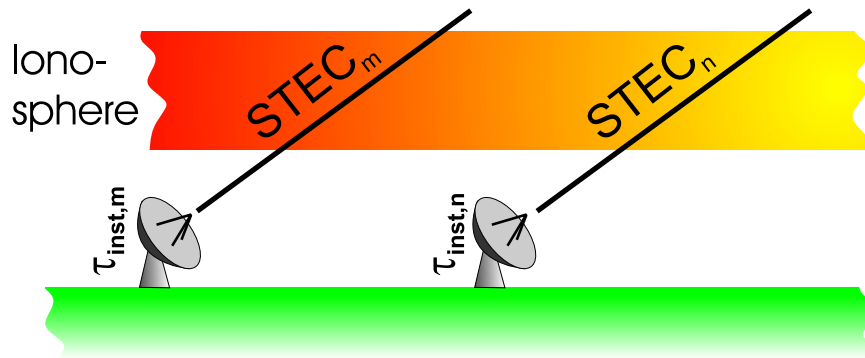


Figure 29: Basic situation for the determination of TEC values from VLBI observations.

subsections. As the measurements contain only relative values of total electron content the main goal will be now the separation of the instrumental biases from ionospheric delays and the development of a method that determines absolute values.

5.1 Methods and algorithms

Assuming a single layer model of the ionosphere, equation 5.1 changes to

$$\tau'_{igx,mn}(t) = \frac{40.28}{c\omega_{gx,mn}^2} [M(z_m) \cdot VTEC_m(t) - M(z_n) \cdot VTEC_n(t)] + \tau_{inst,m} - \tau_{inst,n} \quad (5.2)$$

using the zenith distances z_m and z_n and the corresponding mapping functions $M(z)$ to find a relationship to values of vertical total electron content. As observations are carried out in different azimuth and zenith angles $VTEC_m(t)$ and $VTEC_n(t)$ are not only a function of time. Therefore one can write

$$VTEC_i(t) = VTEC(\lambda_i^*, \phi_i^*, t^*) \quad (5.3)$$

where i stands for any arbitrary station, λ_i^* and ϕ_i^* represent the geographic coordinates of the ionospheric pierce-point, which differ from the station position (λ_i, ϕ_i) . For considerations which will follow

in section 5.1.1 the time of observation is denoted by t^* . From an algebraic point of view equation (5.2) clarifies the problem that without any model assumptions it will be impossible to solve for ionospheric parameters as each new measurement (represented by the left side of this equation) corresponds to two new unknown values of vertical total electron content. The only way to overcome the problem (and to gain an overdetermined system) is to set up an analytical model, which parameterizes the behavior of the vertical electron content in a way that the number of unknowns is less than the number of observations. The most straightforward approach would be to set up a global model which represents VTEC at any point on the globe and to estimate the necessary model parameters in a least-squares adjustment. This was carried out by Bergstrand and Haas (2004, [5]) using a sun-fixed reference frame and a spherical harmonics expansion to model vertical total electron content on a global scale. As the station distribution of typical geodetic VLBI experiments is rather sparse, only small parts of the Earth are covered with observations. In 24 hour-experiments, the usage of a sun-fixed reference frame has the advantage that long bands in the longitude domain are covered, but their latitude extent does not exceed three to four degrees. In order to obtain reasonable values of vertical total electron content one has to introduce artificial observations (so-called constraints) which fill gaps in regions where no data were taken and stabilize the design matrix used in the follow-up adjustment process. These constraints are given an arbitrary weight, which should be less than the weight of "real" observations. The problem is now to find proper weights which enhance numerics but don't contribute too much to the solution. Furthermore, it is not an easy task to determine the absolute value of the constraints. One might use a "plain surface" approach, where all constraints are given the same value, but it might also be worthwhile to use a theoretical model of the artificial observations. Nevertheless, it was pointed out by the authors that it is principally possible to determine global maps of vertical total electron content, which agree to results from GPS on a level of about ± 10 TECU. In another approach, using global spherical harmonics, described by Todorova et al. (2004, [77]), VLBI data are used to set up the normal equation system, which is stacked with results from other techniques like GPS and satellite altimetry. The latter technique provides useful measurements over the oceans, where neither VLBI nor GPS stations exist. Besides an improved model of the ionosphere, especially over the oceans, it is expected that instrumental biases, can be determined more reliably. One of the restrictions using this method is given by the fact that nearly all satellite altimetry missions have orbits, which vary very little in a sun-fixed reference system during one revolution. Another problem is caused by the fact that the number of GPS measurements exceeds the number of VLBI and satellite altimetry measurements by at least a factor of hundred. Therefore the weights of the measurements have to be determined carefully before they are combined with normal equation matrices from GPS. Besides the strategies described above using equation (5.2), one can also solve for station-dependent models, which describe the time-dependent variations of the vertical total electron content. In a paper by Kondo (1993, [44]) it was already mentioned that the ionosphere is not only an obstacle which has to be negotiated in processing of geodetic VLBI sessions, but also that station-dependent ionospheric parameters can be obtained. In an earlier paper by Kondo (1991, [43]) it was explained in detail how station-dependent VTEC values can be obtained. It was suggested to describe the time-variation of the total electron content above the station by the following approach

$$VTEC(\lambda_i^*, \phi_i^*, t^*) \approx VTEC(\lambda_i, \phi_i, t^*) = a_{i0} + \sum_{k=1}^4 \left[a_{ik} \cos\left(\frac{kt^*\pi}{12}\right) + b_{ik} \sin\left(\frac{kt^*\pi}{12}\right) \right] + c_i t^* \quad (5.4)$$

where spatial gradients around the station were neglected. This means that vertical total electron content values related to the ionospheric pierce point are assumed to be the same as those measured directly above the station. The proposed model takes into account harmonics up to the 4th order (i.e. 6-hour period), which has been found sufficient to describe the main characteristic time-variations of vertical total electron content. More than ten years later another approach, based on the idea of neglecting spatial gradients, too, was presented using piece-wise linear functions (Hobiger and Schuh, 2004, [33]). The two approaches were compared internally and against results from GPS measurements. It was shown that they provide nearly the same outcomes, but the approach using piece-wise linear functions has the advantage that it is not sensitive to data gaps and can detect ionospheric variations with much smaller periods as no periodicity is set by default. The following section describes a

station-dependent model, that takes spatial gradients around the stations into account.

5.1.1 The station-dependent model

Instead of modeling the ionosphere on a global scale or neglecting spatial variations around the observing stations a new approach will be described which uses the physical nature of the ionosphere to relate the vertical total electron content at the ionospheric pierce point to the one that would be observed directly above the station. Figure 30 already takes the conversion between STEC and VTEC into account as it was described by equations (5.1) and (5.2). When the thin-shell approximation is

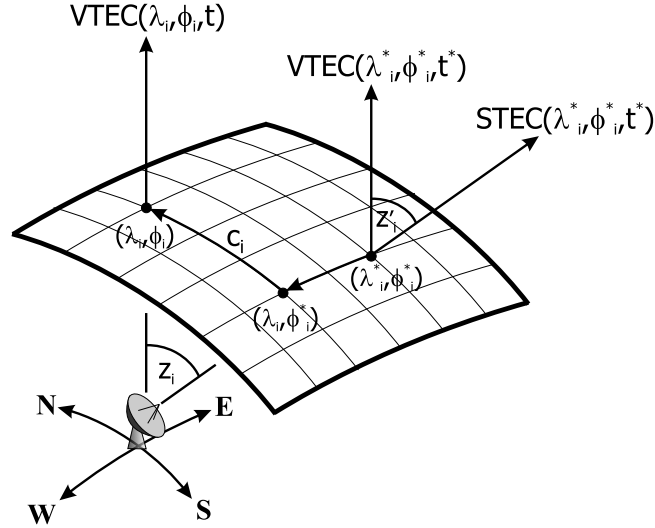


Figure 30: Relation between VTEC at the observing station at the intersection point.

used a simple mapping function as described in section 4.5 is sufficient to obtain values of vertical total electron content $VTEC(\lambda_i^*, \phi_i^*, t^*)$ where the star denotes time and location of the observation. As the ionosphere is co-rotating with the sun, the ionospheric pierce point will be rotated too and will cross the meridian of the station at time t . The time difference (in hours) between the observation and this event is given by

$$t - t^* = \frac{\lambda^* - \lambda}{15} \quad (5.5)$$

where λ denotes the longitude of the station. In other words, it can be assumed that the same VTEC observation is made at location (λ, ϕ^*) at time

$$t = t^* + \frac{\lambda^* - \lambda}{15}. \quad (5.6)$$

If the ionospheric pierce-point is located eastwards of the station, the fraction on the right side of equation (5.6) will be positive and the time has to be increased, for a westward location the time must be decreased, respectively. Therefore it is possible to “rotate” any arbitrary observation into the meridian of the station. When a constant latitudinal gradient c_i is assumed it is possible to deduce the VTEC value at the station by

$$VTEC(\lambda_i, \phi_i^*, t) = (1 + c_i \Delta\phi) VTEC(\lambda_i, \phi_i, t) \quad \text{with} \quad \Delta\phi = \phi_i^* - \phi_i. \quad (5.7)$$

The north-south gradient c_i at each station can be treated as an unknown quantity and can be estimated together with the other target parameters. Therefore the relation between the observed value of vertical total electron content and the modeled value at the station is given by equation (5.7) when the observation time is modified according to expression (5.6). Finally the station-dependent VTEC model can be parameterized by any suitable model (e.g. the Fourier-approach described in the prior section). Although not mentioned explicitly, the coordinate system in the equations above can be set to a geographic or geomagnetic one. But as the ionospheric variations can be described better (i.e. the latitudinal gradients c_i are nearly constant) in a geomagnetic coordinate system such a model

space is strongly recommended. Recapitulating the necessary computation steps before adjustment yields the following procedure:

1. calculation of the ionospheric pierce point in a geographical coordinate system,
2. transformation of the pierce point and station location into a geomagnetic coordinate system,
3. “correction” of the observation time by computing the (geomagnetic) longitude differences according to equation (5.6),
4. computation of the (geomagnetic) latitude differences needed to set up equation (5.7).

As already mentioned in the prior section, a piece-wise linear function model will be taken for the variations of the vertical total electron content at each station.

5.1.2 Piece-wise linear function

Piece-wise linear functions are first order polynomials, which are strung together, so that a continuous curve is granted. From a mathematical point of view it can be treated as a first order spline function, and when t represents the independent variable, any value can be derived by

$$VTEC(t) = offset + \sum_{i=0}^K rate_i(t_{i+1} - t_i) + rate_M(t - t_M) \quad (5.8)$$

where M satisfies $t_M < t$ and $M = K + 1$.

The parameters offset and rate define the shape of the curve and can be easily estimated within a least-squares adjustment as they don't have to be linearized. The piece-wise linear function is often used for parameterizing time-dependent variations within geodetic VLBI analyses. Typical examples are clock and troposphere models, which vary significantly within one experiment. Usually equally spaced time intervals are chosen for geodetic applications as the results will be combined with or compared to other analysis centers. To overcome singularities in the design matrices, artificial observations (constraints)

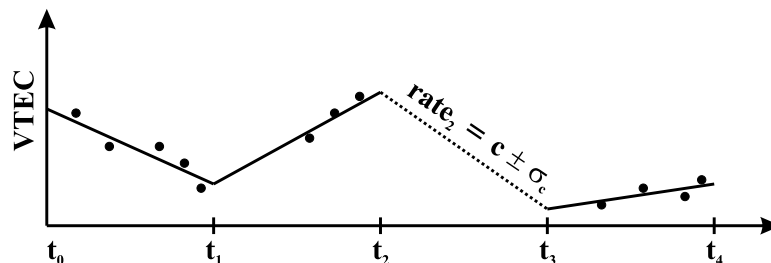


Figure 31: Equally spaced piece-wise linear function, constraints are needed for data gaps.

are introduced. Figure 31 illustrates the problem, when no data are available to contribute to the third interval. Therefore constraints c to the rates

$$rate_i = c \pm \sigma_c \quad (5.9)$$

are added to the design matrix to get a reliable solution. The reciprocal formal error of the constraint ($1/\sigma_c$) defines the weight of the additional observation and has to be chosen very carefully. To avoid artificial observations whenever possible a very simple method will be used for all computations. Instead of equally spaced intervals, interval lengths are defined by a given number of observations (N_{obs}). Interval end-points are calculated by the arithmetic mean of the last value of one group and the first value of the following one. Figure 32 shows how the piece-wise linear function will look like, when this strategy is applied to the test data set. Generally, the time intervals between the VLBI scans vary according to geometry, slewing times of the antennas, flux densities of radio sources and other reasons. Thus the time steps denoted in equation (5.8) should be set that the given number N_{obs} of observations defines the length of each interval. In the analysis a value of $N_{obs} = 8$ observations, contributing to

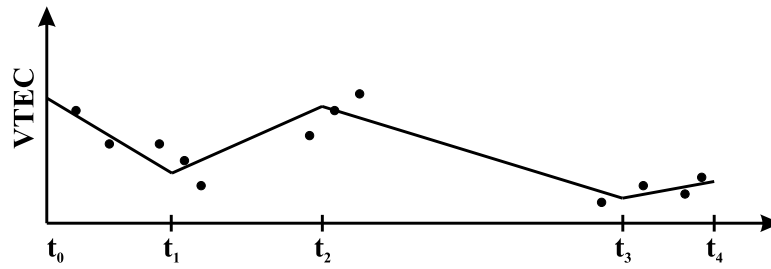


Figure 32: Piece-wise linear function with adaptive interval lengths.

the estimation of one interval, was chosen, which corresponds to an average temporal resolution of the outcomes of about 30 minutes. This value defines also the lower limit of the detectable ionospheric periods. Changing the value N_{obs} to a lower number of observations, increases time resolution, but decreases the redundancy of the estimation process and weakens the stability of the matrices used in the adjustment.

5.1.3 Instrumental offsets and their treatment

Before any adjustment of the unknown parameters can be carried out, singularity of the design matrix, caused by the constant terms of the instrumental offsets (last two terms in equation (5.2)), has to be eliminated. One method to overcome this problem is to set the instrumental offset at one station within the network to zero and obtain the other values relatively to that one. Another approach - equivalent to the first one - sets the sum of all offsets within a network of N stations to zero (Sekido et al., 2003, [67]).

$$\sum_{i=1}^N \tau_{inst,i} = 0 \quad (5.10)$$

One has to be careful not to over-parameterize the station-dependent VTEC models. Let the number of model intervals set up for station i , be int_i then the redundancy R , using n_{obs} observations and N stations is given by

$$R = n_{obs} - \left(3N + \sum_{i=1}^N int_i \right) + 1. \quad (5.11)$$

The last term in equation (5.11) results from the effect that one artificial observation (equation (5.10)) increases the number of observations by 1. As the number of observations carried out by VLBI within one session is rather small compared to global GPS measurements, redundancy should be high enough to be able to detect outliers within the data. In some experiments it is necessary to set up more than one instrumental offset per station, as some instrument changes within the session shift this value. In geodetic analysis such an event is treated as a “clock-jump”, which is modeled by a second set of clock parameters for that station, too. As mentioned by Li (2005, [50]) this effect cannot be handled easily when VLBI is used to track satellite signals in order to determine their accurate position. Figure 33 shows an example of a geodetic experiment (EUROPE-52, Dec. 13th, 1999) where one station (Crimea, Ukraine) experiences a power outage² after the first hours of the experiment. Therefore two instrumental offsets, before and after the jump will be parameterized for this station when values of vertical total electron content are solved for. Without this setting it would become impossible to obtain reasonable VTEC values for this particular station and even all other stations would be influenced as this discontinuity is visible on all baselines connecting Crimea. As seen by this example it would be necessary to check all experiments for such discontinuities and mark them, before adjustment processes can be applied to all available geodetic VLBI databases.

²documented in the station log file (<http://lupus.gsfc.nasa.gov/data10/sessions/1999/euro52/euro52sm.log.com.html>)

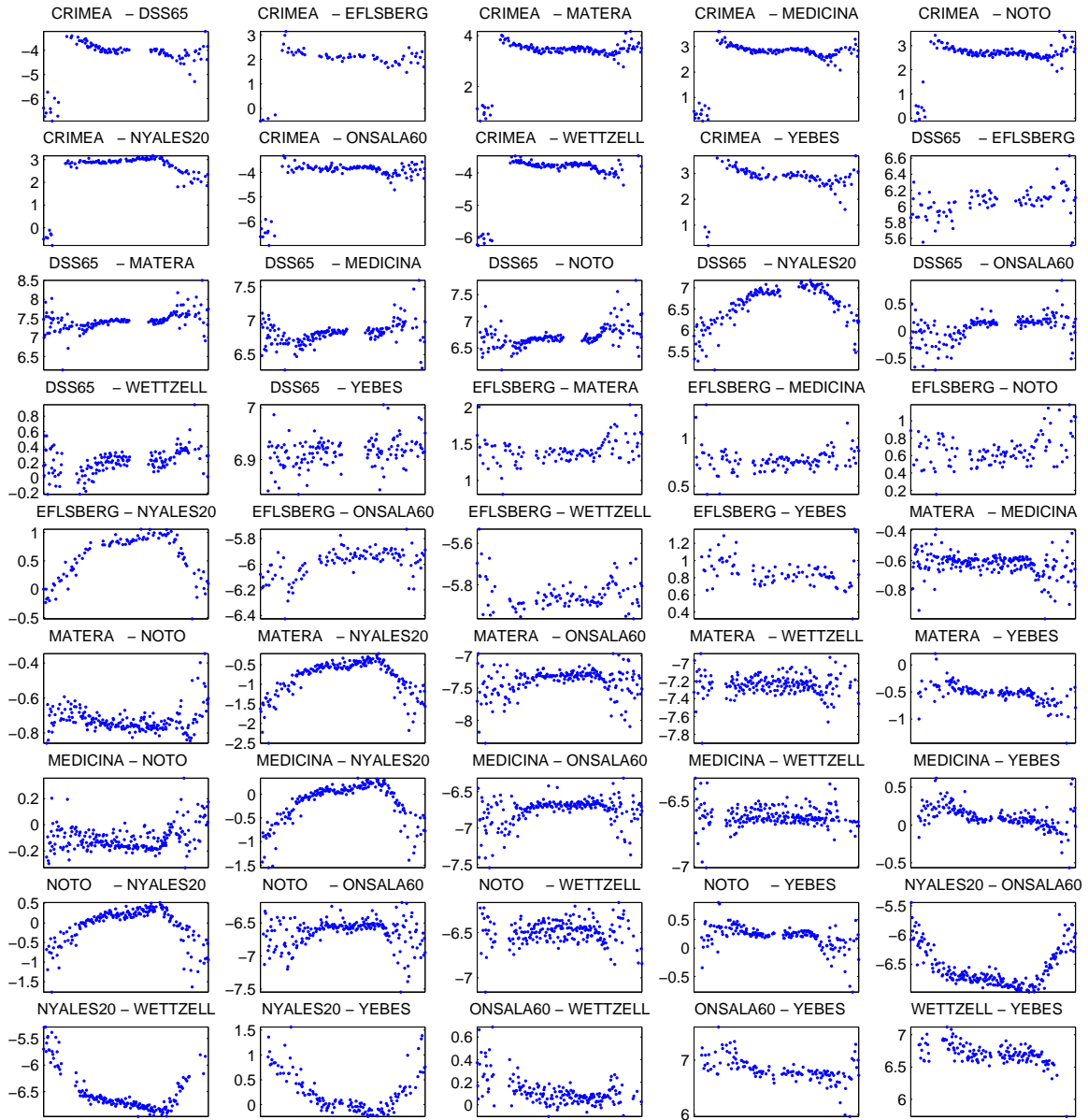


Figure 33: Ionospheric delay corrections at X-band in nanoseconds for EUROPE-52 experiment, Dec. 13th, 1999. Due to instrumental changes at station CRIMEA (Ukraine), ionospheric corrections for baselines connecting this station have a significant jump after the first hours of operation.

5.1.4 Proper height of the thin-shell layer

As discussed in section 4.5, the height H of the single layer used for the mapping function $M(z)$ should be chosen approximately at the height of the F2 layer. When using GPS observations, the errors introduced by this approach have been mentioned, too. In the case of VLBI, the situation is different, as each station has only one observation at a time, in contrast to GPS where each station tracks at least four satellites in different azimuth and elevation angles, simultaneously. Therefore an accurate mapping function will become essential for the determination of station-dependent VTEC values. Hence the single layer height should be as close as possible to the “real” height of the F2 peak. A rough estimate of the error, introduced by a wrong value of H , can be found using the partial derivative of equation (4.49) with respect to H

$$\frac{\partial M(z)}{\partial H} = - \frac{R_e^2 \sin^2 z}{(R_e + H)^3 \left[1 - \left(\frac{R_e}{R_e + H} \sin z \right)^2 \right]^{3/2}} \quad (5.12)$$

and calculating the relative error

$$\frac{\Delta M(z)}{M(z)} = \frac{\partial M(z)}{\partial H} \frac{\Delta H}{M(z)} = - \frac{R_e^2 \sin^2 z}{(R_e + H)^3 \left[1 - \left(\frac{R_e}{R_e + H} \sin z \right)^2 \right]} \Delta H \quad (5.13)$$

of the mapping function. Setting the radius of the earth to $R_e = 6370$ km and assuming that the height of the single layer was chosen to be $H = 450$ km, figure 34 gives the relative error in % of the corresponding mapping function for different zenith distances and different heights. It can be seen

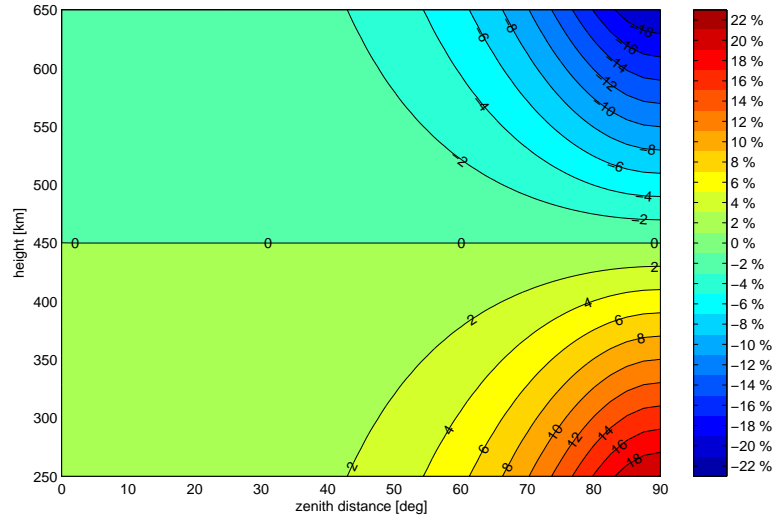


Figure 34: Error of the mapping function in percent, when the actual height differs from the single layer model height ($H = 450$ km).

that for observations at very low elevations the mapping function error increases rather fast when the height of the single layer is not set properly. In the case of GPS this is not such a big problem as satellites are tracked continuously and many different elevation angles are involved. In order to find out how the actual height of F2 peak varies a model run using the IRI2001 (Bilitza, 2001, [6]) was carried out. Figure 35 shows a spatial distribution of this height parameter, computed for year 2000, DOY 125 at 12 h universal time (UT). If the height of the single layer model is set to 450 km in this

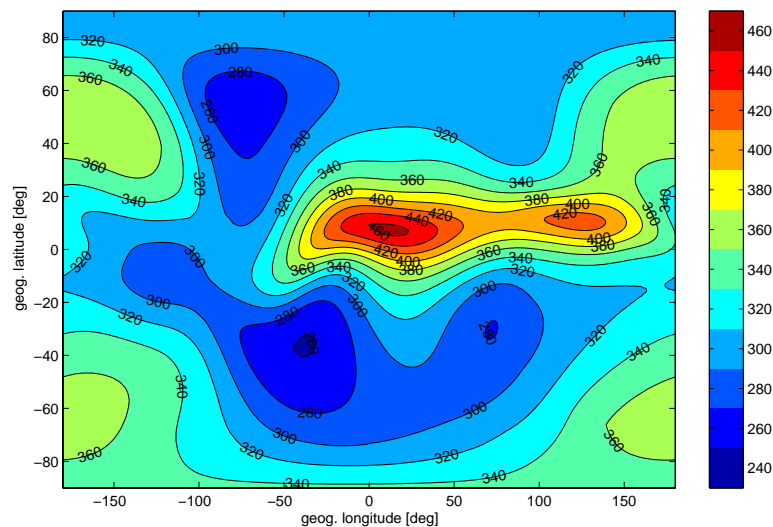


Figure 35: Geographical distribution of the height of the F2 peaks (hmF2). Computed from IRI2001 model, for 2000, DOY 125, 12h UT.

case, the mapping function can introduce errors (depending on the zenith distance of the observation)

in many regions of the world, as the peak height of the F2 layer is well below 450 km. But as the pattern of hmF2 is similar to that one of a daily VTEC distribution the effect of such a mis-modeling is reduced. The differences between actual and model heights will be bigger in regions where VTEC shows lower numbers, but as the relative error is multiplied by smaller numbers the effect on values of total electron content is lowered, too. Anyway, if someone wants to obtain accurate VTEC values from VLBI observations it will be necessary to take such effects into account. Principally, there are two ways of handling this problem: Estimating the height of the ionospheric layer together with the other unknowns in the adjustment process or introducing the height from external measurements or models. The first approach has the advantage that no external information is needed, but the gained heights will be correlated with the other parameters, which can weaken the matrices and can complicate the determination of accurate VTEC values. One has to be aware of over-parameterizing the whole system, too, when setting up the height models for the stations. When applying the second approach, using a height of the F2 peak from external data sources, the yielded parameters cannot be called “VLBI-only” values anymore. But when a theoretical model, like IRI, is used, the main information, i.e. the values of vertical total electron content at each station, will still be a pure VLBI product. Therefore it was decided to take values of hmF2 from IRI2001 to consider variations of the mapping functions within one experiment. For this purpose, time-series of hmF2, sampled every hour, were computed for all VLBI stations, using the IRI2001 model. These values can be interpolated to the time of observation and applied for the calculation of the mapping function, when the matrices are set up for the adjustment process. The situation can be compared to the estimation of tropospheric parameters, where numerical weather models are applied, to obtain better values of the tropospheric mapping function and to improve station coordinates (Böhm and Schuh, 2003, [8]).

5.1.5 Adjustment and non-negative constraints

Usually a general least-squares process is applied to solve a set of equations for a set of unknowns. The only requirement is that there are at least as many equations as unknowns. If the equations are linear, the least-squares process will produce a direct solution for the unknowns. Otherwise an initial approximation to the unknowns is required, and improvements of the initial parameters are obtained. This is repeated until the results converge. The linear case is an adjustment using zero as initial guess of all parameters. The process requires a set of equations with the unknowns on one side and some known quantity on the other. Let x_i be the set of unknowns, and let the equations be of the form

$$F_i(x_1, x_2, x_3, \dots) = o_i \quad (5.14)$$

where o_i is the observation (value) of which the least-squares error will be minimized. Since there are more equations than unknowns, the solution of the unknowns will not be exact. Evaluating $F_i(x_1, x_2, x_3, \dots)$ will not generate the exact observation value o_i . The Jacobian or design matrix \mathbf{A} is the matrix of the partial derivatives of each equation with respect to the unknowns. That is

$$\mathbf{A} = \begin{pmatrix} \frac{\partial F_1}{\partial x_1} & \frac{\partial F_1}{\partial x_2} & \frac{\partial F_1}{\partial x_3} & \dots \\ \frac{\partial F_2}{\partial x_1} & \frac{\partial F_2}{\partial x_2} & \frac{\partial F_2}{\partial x_3} & \dots \\ \frac{\partial F_3}{\partial x_1} & \frac{\partial F_3}{\partial x_2} & \frac{\partial F_3}{\partial x_3} & \dots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}. \quad (5.15)$$

Furthermore, let the vector $\Delta\mathbf{Y}$ be the vector of the differences between the observations and the model, calculated from the initial values. That is

$$\Delta\mathbf{Y} = \begin{pmatrix} o_1 - F_1(x_{10}, x_{20}, x_{30}, \dots) \\ o_2 - F_2(x_{10}, x_{20}, x_{30}, \dots) \\ o_3 - F_3(x_{10}, x_{20}, x_{30}, \dots) \\ \vdots \end{pmatrix} \quad (5.16)$$

where x_{i0} is the initial guess for the unknown x_i . The initial guesses can have any finite real value, but in order to avoid numerical instabilities and to let the system converge faster the guesses should be close to the solution. \mathbf{x}_0 is the vector of these initial guesses and $\Delta\mathbf{x}$ represents the vector of improvement as obtained from the adjustment. Taking the residuals

$$\mathbf{v} = \mathbf{A}\Delta\mathbf{x} - \Delta\mathbf{Y} \quad (5.17)$$

for minimization of their weighted squared sum, using the (weight-)matrix \mathbf{P} of the observations yields

$$\begin{aligned} \min_{\Delta\mathbf{x}} \mathbf{v}^T \mathbf{P} \mathbf{v} &= \min_{\Delta\mathbf{x}} ((\Delta\mathbf{x}^T \mathbf{A}^T - \Delta\mathbf{Y}^T) \mathbf{P} (\mathbf{A}\Delta\mathbf{x} - \Delta\mathbf{Y})) \\ &= \min_{\Delta\mathbf{x}} \left(\Delta\mathbf{x}^T \underbrace{\mathbf{A}^T \mathbf{P} \mathbf{A}}_{\mathbf{H}} \Delta\mathbf{x} - 2 \underbrace{\mathbf{A}^T \mathbf{P} \Delta\mathbf{Y}}_{-\mathbf{f}^T} \Delta\mathbf{x} + \underbrace{\Delta\mathbf{Y}^T \mathbf{P} \Delta\mathbf{Y}}_{const>0} \right). \end{aligned} \quad (5.18)$$

Finding the minimum of equation (5.18) by setting the derivative with respect to $\Delta\mathbf{x}$ to zero would be equal to the Gauss-Markov model (Koch, 1997, [42])

$$\Delta\mathbf{x} = -\mathbf{H}^{-1} \mathbf{f}^T \quad (5.19)$$

used in classical geodetic adjustments. \mathbf{H} is the normal equation matrix and $-\mathbf{f}^T$ represents the ‘‘right-side’’ of the equation system. If the vector of improvements underlies some additional restrictions, a solution according to equation (5.19) cannot be obtained, but another method has to be used. One has to go back to expression (5.18) and find another minimizing algorithm, which takes the additional constraints into account. In the case of estimating VTEC values a simple relation can be found as follows. Due to the physical nature of the ionosphere negative values of vertical total electron content are not possible. This fact can be introduced to the estimation algorithm by adding in-equation conditions to expression (5.18). If none of the VTEC values at the interval end-points is less than zero, it is ensured that all values described by the corresponding piece-wise linear functions are positive or at least zero. This yields a set of conditions

$$\begin{aligned} offset &\geq 0 \\ offset + rate_1(t_1 - t_0) &\geq 0 \\ offset + rate_1(t_1 - t_0) + rate_2(t_2 - t_1) &\geq 0 \\ &\vdots \\ &\vdots \end{aligned} \quad (5.20)$$

on the parameters which describe the piece-wise linear function. As these parameter should be estimated within an adjustment process they can be described by the sum of initial guess (denoted by subscript 0) and improvement from the adjustment (denoted by Δ). Therefore equation (5.20) can be written as

$$\begin{aligned} (offset_0 + \Delta offset) &\geq 0 \\ (offset_0 + \Delta offset) + (rate_{1,0} + \Delta rate_1)(t_1 - t_0) &\geq 0 \\ (offset_0 + \Delta offset) + (rate_{1,0} + \Delta rate_1)(t_1 - t_0) + (rate_{2,0} + \Delta rate_2)(t_2 - t_1) &\geq 0 \\ &\vdots \\ &\vdots \end{aligned} \quad (5.21)$$

Applying the non-negative constraint to all points at the interval boundaries yields an equation that ensures that all values will be positive after the adjustment. If all conditions are summed up using matrix notation, the vector of initial guess \mathbf{x}_0 and the vector of improvements $\Delta\mathbf{x}$ equation (5.22) can be substituted.

$$\underbrace{\begin{pmatrix} 1 & 0 & 0 & 0 & \dots \\ 1 & t_1 - t_0 & 0 & 0 & \dots \\ 1 & t_1 - t_0 & t_2 - t_1 & 0 & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}}_{\mathbf{B}} \mathbf{x}_0 + \underbrace{\begin{pmatrix} 1 & 0 & 0 & 0 & \dots \\ 1 & t_1 - t_0 & 0 & 0 & \dots \\ 1 & t_1 - t_0 & t_2 - t_1 & 0 & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}}_{\mathbf{B}} \Delta\mathbf{x} \geq 0 \quad (5.22)$$

Although it would be possible to multiply by \mathbf{B}^{-1} the matrix \mathbf{B} will be kept for sake of completeness. Re-arranging equation (5.22) yields a simple condition between the vector of improvements and the vector of initial values.

$$\mathbf{B} \cdot \Delta \mathbf{x} \geq -\mathbf{B} \cdot \mathbf{x}_0 \quad \Rightarrow \quad \mathbf{B} \cdot \Delta \mathbf{x} \geq \mathbf{C} \quad \text{where} \quad \mathbf{C} = -\mathbf{B} \cdot \mathbf{x}_0 \quad (5.23)$$

As the constant part in equation (5.18) can be ignored, the minimization problem with conditions reduces to

$$\min_{\Delta \mathbf{x}} \left(\frac{1}{2} \Delta \mathbf{x}^T \mathbf{H} \Delta \mathbf{x} + \mathbf{f}^T \Delta \mathbf{x} \right) \quad (5.24)$$

$$\mathbf{B} \cdot \Delta \mathbf{x} \geq \mathbf{C}.$$

Equation (5.24) represents a quadratic programming problem, which can be solved by algorithms described in Gill et al. (1981, [26]) or in a paper by Coleman and Li (1996, [16]). Appendix A gives a brief overview of the algorithm which solves expression (5.24). If all observations made by VLBI would be perfect it would not be necessary to solve the quadratic programming problem but only take the simple Gauss-Markov model (equation (5.19)). But as outliers in the data easily evoke negative values of vertical total electron content, this method was chosen to obtain a robust solution. It can be shown, that if the Gauss-Markov solution does not contain negative values for the piece-wise linear functions, the solution of the quadratic programming problem is identical.

5.1.6 Stochastic model, weighting of the data

The weight matrix \mathbf{P} described in the prior section is a $N_{obs} \times N_{obs}$ matrix defined as a diagonal matrix

$$P_{ij} = \begin{cases} \left(\frac{1}{\sigma_i} \right)^2 & (i = j) \\ 0 & (i \neq j) \end{cases}. \quad (5.25)$$

using the formal error of each observation σ_i , neglecting correlations between the observations. The stochastic model described in equation (5.25) does not take into account any geometric concerns. As described in Hobiger and Schuh (2004, [33]), observations at low elevations are necessary to de-correlate vertical total electron content from station-dependent instrumental offsets, but may introduce errors in the VTEC values above the station because of the imperfectness of the station-dependent model. In such a case the ionospheric pierce point is far away from the station and the transformations described in section 5.1.1, should be appropriate to describe spatial gradients around the station. On the other hand, observations made at high elevations have ionospheric pierce-points close to the station location, which means that the observed VTEC and the value above the station will have nearly the same value. The drawback of such observations is that it is difficult to separate ionospheric contributions from station-dependent instrumental offsets as the mapping function will be close to one. The situation gets better for large VLBI networks as zenith distances will differ more, as the baseline gets longer. Therefore a factor, taking the geometry of each observation into account, can be considered for setting up the stochastic model (Hobiger and Schuh, 2004, [33]). Error propagation (neglecting instrumental offsets) of equation (5.2) gives

$$\sigma_{\Delta STEC} = \sqrt{M(z_m)^2 \sigma_{VTEC_m}^2 + M(z_n)^2 \sigma_{VTEC_n}^2} \quad (5.26)$$

which simplifies to

$$\sigma_{\Delta STEC} = \sqrt{2} \sigma_{VTEC} \quad (5.27)$$

when it is assumed that $\sigma_{VTEC} = \sigma_{VTEC_m} = \sigma_{VTEC_n}$ and that both antennas point into zenith direction. Although such a situation will never happen for any station and source combination (due to the curvature of the Earth), expression (5.27) can be used to describe the relation between the ideal zenith observations and any other pointing combination. Additionally this ratio can be taken to the k -th power, where k is an arbitrary integer number, which yields equation (5.28). In fact the

parameter k determines how much weight is given to low elevation observations in the adjustment process.

$$w^k(z_m, z_n) = \left(\frac{M(0^\circ)}{\sqrt{\frac{1}{2} [M(z_m)]^2 + \frac{1}{2} [M(z_n)]^2}} \right)^k$$

$$= \left(\frac{2 \cdot \left[1 - \left(\frac{R_e}{R_e + H} \right)^2 \sin^2 z_m \right] \cdot \left[1 - \left(\frac{R_e}{R_e + H} \right)^2 \sin^2 z_n \right]}{2 - \left(\frac{R_e}{R_e + H} \right)^2 (\sin^2 z_m + \sin^2 z_n)} \right)^{k/2} \quad (5.28)$$

The impact of the parameter k can be outlined creating several contour plots with different values for k . Figure 36 shows plots of the weights of the mapping function for $k = -6, -3, -2, -1, 0, 1, 2, 3, 6$, normalized by their maximum value. Hobiger and Schuh (2004, [33]) concluded from the comparison

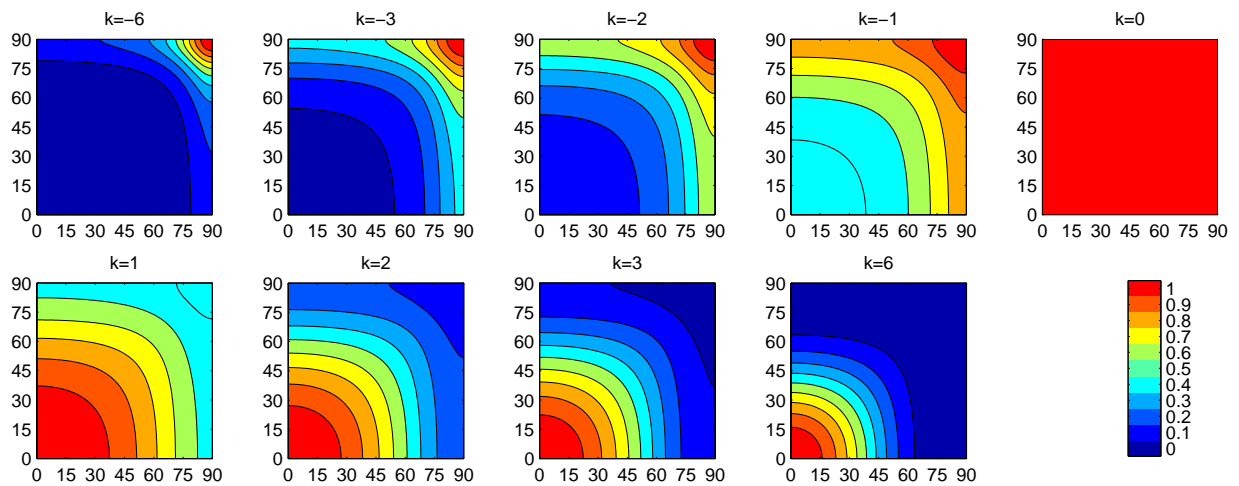


Figure 36: Weighting of measurements as a function of zenith distances (z_1, z_2), with different values for the steering parameter k . The height of the ionospheric single layer was set to $H = 450$ km and the radius of the Earth was assumed to be $R_e = 6370$ km. All plots are normalized, to a maximum value of 1.

between VLBI and GPS derived VTEC values that a value $k = 4$ represents the most suitable weight model. This value was found using a simple station model. Therefore a new comparison was made, applying data from a continuous observation campaign, in order to obtain a proper value of k , considering the station model described in section 5.1.1. Table 7 in chapter 5.5.1 gives an overlook of the differences between VLBI and GPS, for that continuous campaign and it is recommended to use $k = 6$ for all computations. Finally, the adapted stochastic model can be described by

$$P_{ij} = \begin{cases} \frac{w^k(z_m, z_n)}{\sigma_i^2} & (i = j) \\ 0 & (i \neq j) \end{cases} \quad (5.29)$$

5.1.7 Formal error of the results

Together with the unknown parameters their formal errors can be obtained from the adjustment of an overdetermined system. The a-posteriori variance-factor σ_0 can be calculated from the residuals \mathbf{v} (equation (5.17)) by

$$\sigma_0 = \sqrt{\frac{\mathbf{v}^T \mathbf{P} \mathbf{v}}{R}} \quad (5.30)$$

where the redundancy R is derived according to expression (5.11). If the solution is obtained from an adjustment, following the Gauss-Markov model (5.19), the variance-covariance matrix for the unknowns can be computed with

$$\mathbf{Q}_{\mathbf{xx}} = \sigma_0^2 \mathbf{H}^{-1} \quad (5.31)$$

using the normal equation matrix \mathbf{H} from equation (5.18). But as the quadratic problem (5.24) is solved under inequality side conditions, the underlying probability distributions will no longer be of Gaussian type (normal distribution), which principally makes it impossible to apply such a simple relation as (5.31) to conclude on the formal errors of the unknown parameters. But as a straight forward approach it can be assumed that the underlying statistics can be approximated by a Gaussian distribution. Therefore precision numbers of the unknown parameters can be computed based on equation (5.31). As mentioned above, if the inequality side conditions are not utilized, the results of the quadratic programming problem will be the same as those of the Gauss-Markov model anyway. If inequality conditions affect the unknown parameters, the residuals will get bigger than in the Gauss-Markov model adjustment, and therefore the a-posteriori variance-factor σ_0 will increase, too. As expected this means, if negative values of the piece-wise linear function (representing values of vertical total electron content) are prevented by the algorithms, the formal error of the parameters will get bigger, too. The variance-covariance matrix of the adjusted observations $\mathbf{Q}_{\mathbf{oo}}$ can be described by

$$\mathbf{Q}_{\mathbf{oo}} = \sigma_0^2 \mathbf{A} \mathbf{H}^{-1} \mathbf{A}^T \quad (5.32)$$

using the same approximation as above. One simple measure for the detection of gross outliers can be found by calculation of the standardized residuals

$$\gamma_i = \frac{v_i}{\sigma_{v,i}} \quad (5.33)$$

utilizing the variance-covariance matrix of the residuals

$$\mathbf{Q}_{\mathbf{vv}} = \mathbf{P}^{-1} - \mathbf{Q}_{\mathbf{oo}}. \quad (5.34)$$

5.2 Results for globally distributed stations

Using the whole available database of the International VLBI Service for Geodesy and Astrometry (2005, [38]), time-series for 143 network stations plotted in figure 37 were created. One can see that the

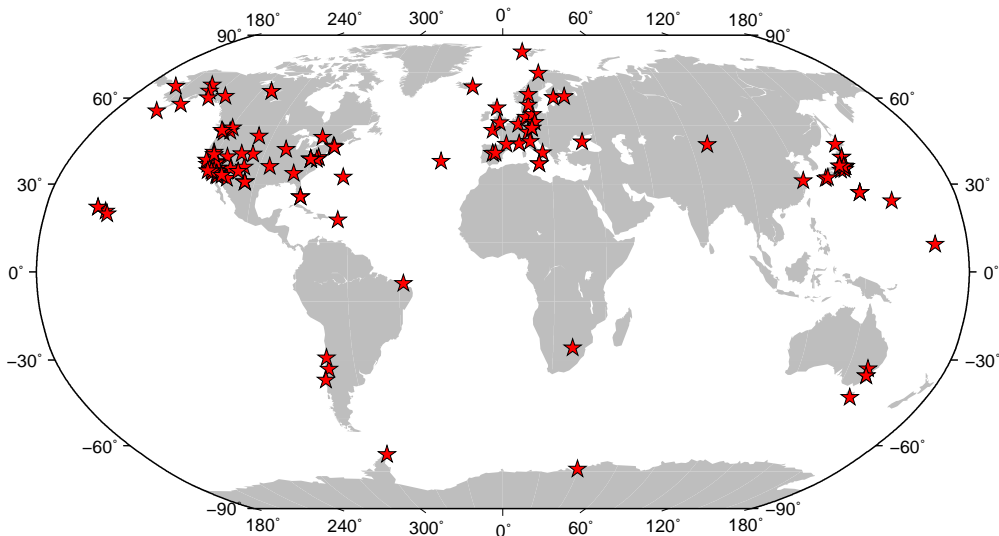


Figure 37: Map of VLBI stations for which VTEC time-series were computed.

station distribution is not uniform, but dominated strongly by stations on the northern hemisphere. Nevertheless, nearly all ionospheric regions, as discussed in section 2.2.2 are covered by at least one station. The series obtained are of unequal lengths as stations were constructed and dismantled at

different epochs in time. Furthermore some antennas contributed less than others, as they are not mainly dedicated to astrometric/geodetic observations and/or are not equipped permanently with dual-frequency receivers. All data were checked manually in order to flag jumps or to remove gross outliers from the raw data. During this time-consuming process several databases were found which show jumps due to some instrumental changes within the experiments. As many of these findings were not parameterized with a “clock-jump” (Titov et al., 2001, [76]) in geodetic analysis, feed-back to the VLBI analysis was given. The adapted steering parameter k for re-weighting of the data (section 5.1.6) was obtained from several solutions with different setups, which were compared to GPS results (section 5.5.1). The so-called intensive sessions, which should determine $dUT1$ on a routine basis (Fischer et al., 2004, [21]), are excluded from the analysis since they last only one hour and use mainly stations WETTZELL (Germany) and TSUKUBA (Japan). Although it is possible to gain meaningful values of VTEC from these sessions, the outcomes will not be robust against outliers and the short experiment duration makes it difficult to determine reliable values for the instrumental biases. The number of hours and the total time-span covered with data is given in table 4 for the ten most active VLBI stations, which performed astrometric/geodetic experiments. It can be seen that the three most

station name	hours of VTEC	first obs.	last obs.
WETTZELL	45964	16/11/1983	30/12/2004
GILCREEK	43696	07/07/1984	23/12/2004
WESTFORD	38654	13/05/1981	14/12/2004
MOJAVE12	31174	26/06/1983	01/09/1992
KOKEE	22668	08/06/1993	30/12/2004
HRAS085	20010	11/04/1980	31/10/1990
FORTLEZA	17144	22/04/1993	30/12/2004
RICHMOND	17051	04/01/1984	21/08/1992
NYALES20	13798	04/10/1994	30/12/2004
HARTRAO	13270	09/01/1986	07/12/2004

Table 4: The ten most active geodetic VLBI stations.

active stations are still contributing to regular observing programs, whereas stations MOJAVE12, HRAS085 and RICHMOND have stopped their participation in geodetic experiments more than ten years ago. Nevertheless several stations cover more than one solar cycle, which makes it possible to detect long-term trends in the ionosphere. All other space geodetic techniques do not cover such a long time-span of data collected.

5.2.1 Results for stations located in different ionospheric regions

Two stations from the mid latitude region, one in the northern hemisphere (WETTZELL (Germany)) and the other one from the southern hemisphere (HARTRAO (South Africa)) are presented here. Figures 38 and 39 show the results of the estimated vertical total electron content values at those stations. As station HARTRAO is closer to the geomagnetic equator than WETTZELL, the obtained VTEC value are slightly higher. The South African station had joined several experiments already in the mid 80ies and started its routine contribution to geodetic VLBI observations around 1990. Both stations clearly show an annual variation due to the Earth’s revolution around the Sun. The solar cycle strongly dominates the overall magnitude of the results, which can be seen in both plots. The plots also contain zero values or values close to zero, which can be found in about 2% of all sessions. Those values are evoked by outliers, occurring at local night-time, and are forced to be positive by the non-negative constraint of the adjustment process. As shown in section 5.2.2 such effects disappear when daily mean values are computed. Station KOKEE (Hawaii, U.S.A), located north of the geomagnetic equator ($\phi_m = 27.33^\circ$), is presented here as a station representing the equator region of the ionosphere. Figure 40 presents the VTEC values for that station. Due to the presence of the equator anomaly large latitudinal gradients of vertical total electron content can occur, which increases the demand on modeling spatial gradients for such stations. In section 5.3 it can be seen that for stations located in

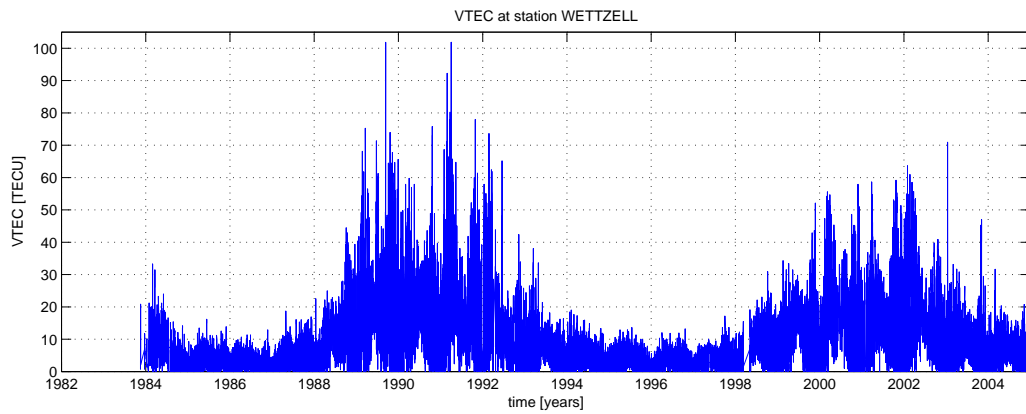


Figure 38: VTEC at station WETTZELL (Germany) in 44.24° geomagnetic latitude.

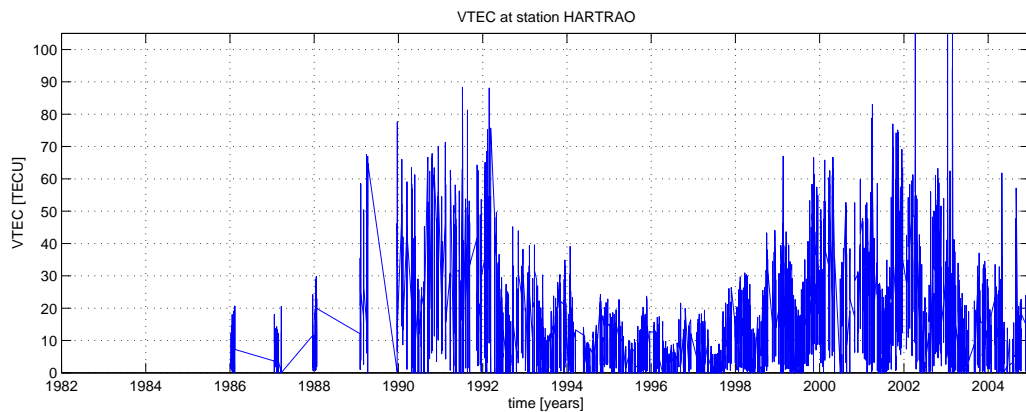


Figure 39: VTEC at station HARTRAO (South Africa) at -31.77° geomagnetic latitude.

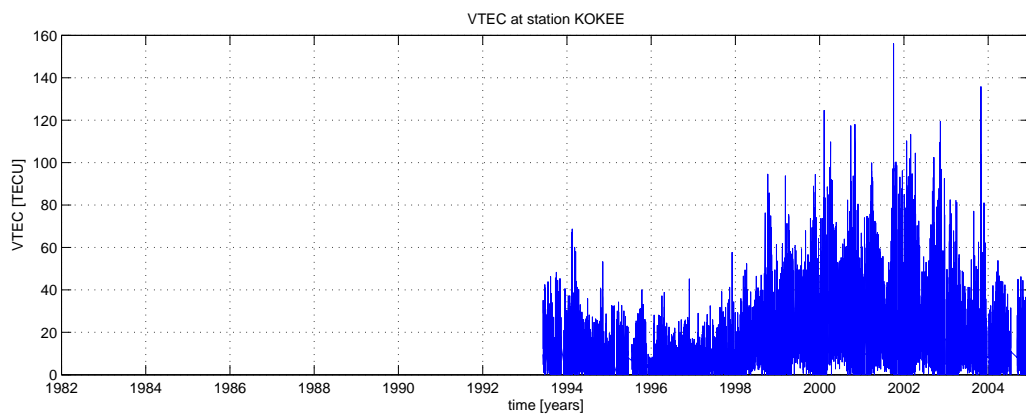


Figure 40: VTEC at station KOKEE (Hawaii, U.S.A) at 27.33° geomagnetic latitude.

the equatorial region, the gradient-like parameter c_i gets larger than in other regions. As depicted in figure 40, the estimated values of vertical total electron content are much larger than in the two other examples shown before, since this station is located in a part of the Earth where electron density values reach their maxima. In section 5.3 it will be shown that for stations located very close to the geomagnetic equator, the mean value of the estimated gradient-like parameter c_i will become larger. This can be explained by larger values of vertical total electron content north as well as south of the geomagnetic equator (equator anomaly). Station NYALES20, located on Spitsbergen (Norway), is plotted in figure 41, representing a station in the northern auroral region. Although one VLBI station (SYOWA, Antarctica) is located in the southern auroral region, it is not presented here as this station contributed only to very few global 24 hours experiments. The obtained VTEC values at

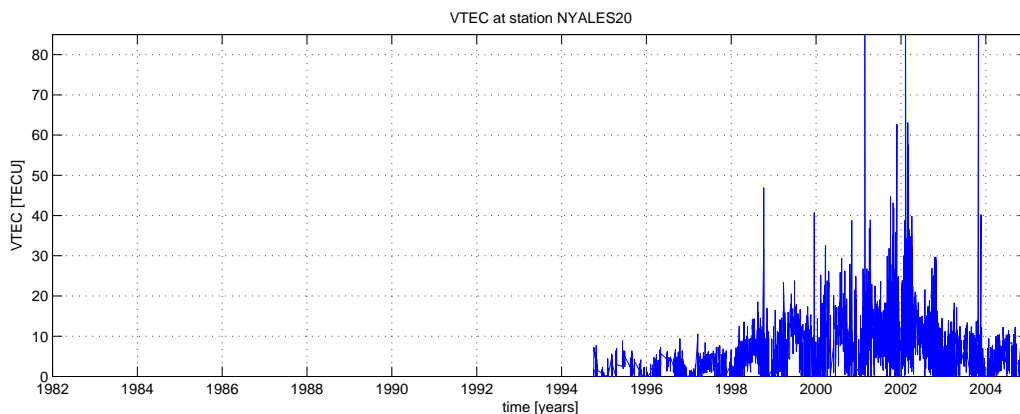


Figure 41: VTEC at station NYALES20 (Norwegian) at 73.27° geomagnetic latitude.

NYALES20 reach the smallest magnitudes amongst the stations presented here. Even at high solar activity, the estimated vertical total electron content seldom exceeds 40 TECU. The requirements of the station-dependent model are less than in other regions as the total error introduced by a wrong spatial gradient model scales down with the magnitude of VTEC. However, a significant mean gradient can be found in the data, as shown in section 5.3. Another feature of the results can already be seen by naked eye, when looking at the plots: the obtained time-series are not equally sampled. Since experiments are not carried out every day, antennas are not always dedicated to geodetic experiments and station contribution even changes between two consecutive experiments of the same type due to scheduling, observation time is distributed arbitrarily. Additionally maintenance of the antenna and the equipment causes large gaps (up to several months) in the time-series. This has to be considered when the frequency spectrum is computed. As discussed in section 5.1.7 the formal error of each estimated parameter can be obtained after the adjustment process. Values of standard deviation (1σ level) for the VTEC values presented here are given in chapter 5.2.3 together with an in-depth discussion about the time-dependent behavior of the parameter precision.

5.2.2 Daily mean values of globally distributed stations

As for the analysis only experiments with a duration of 24 hours or more were selected, a daily mean VTEC for each station can be computed. If observation time exceeds one day, the outer ends of the VTEC time-series for that experiment are cut-off, to obtain unbiased 24 hours averaged values. Such measures are thought to be free of any diurnal variations and long-periodic signals in the ionosphere can be identified more clearly. Figures 42 to 45 show the computed daily averaged values for stations WETTZELL, HARTRAO, KOKEE, and NYALES20. One can clearly see the annual variation of

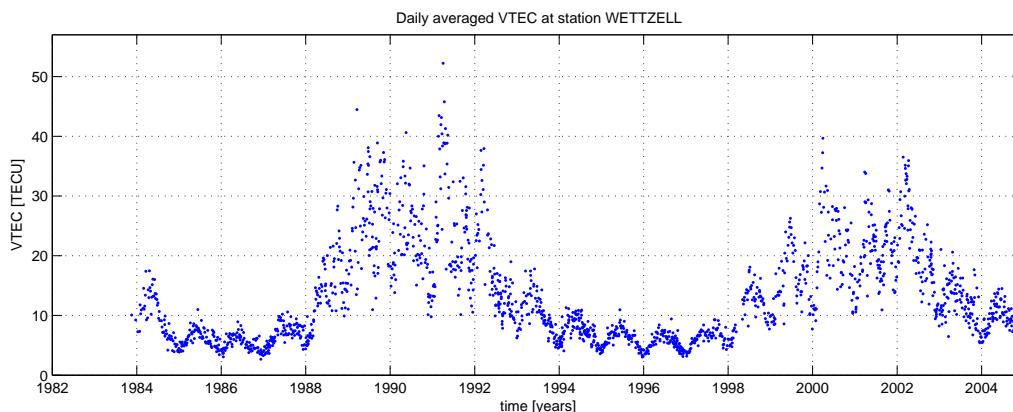


Figure 42: Daily mean values of VTEC at station WETTZELL (Germany) at 44.24° geomagnetic latitude.

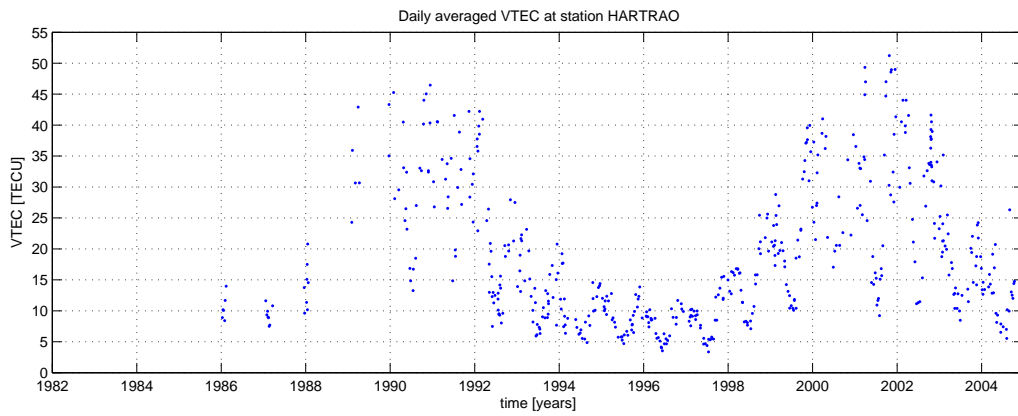


Figure 43: Daily mean values of VTEC at station HARTRAO (South Africa) at -31.77° geomagnetic latitude.

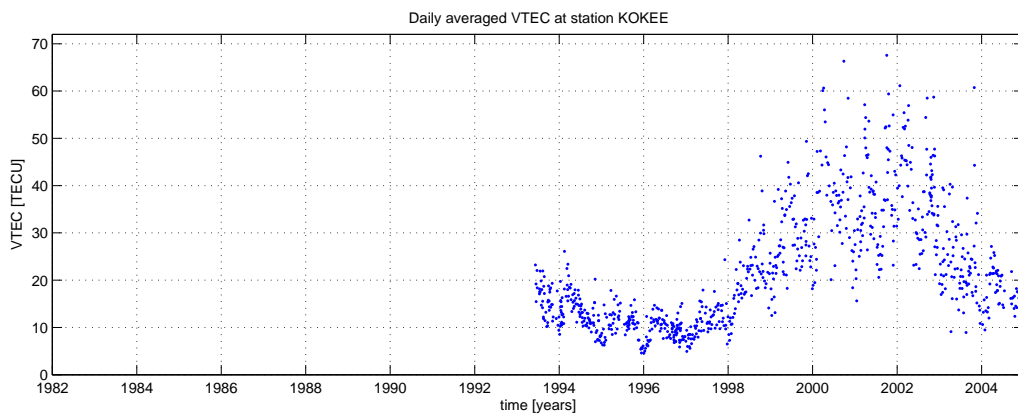


Figure 44: Daily mean values of VTEC at station KOKEE (Hawaii, U.S.A) at 27.33° geomagnetic latitude.

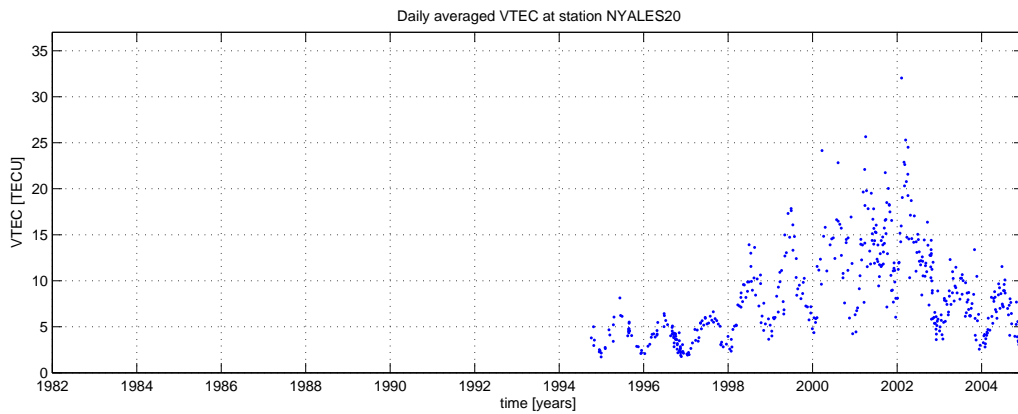


Figure 45: Daily mean values of VTEC at station NYALES20 (Norway) at 73.27° geomagnetic latitude.

the averaged VTEC values. As stated in the previous subsection, the obtained results for stations located in the equator region are significantly higher than in mid latitude and polar regions. The annual signal itself is modulated by the solar cycle, having its maximum values around 1990 and 2001. Consequently VLBI is the only space geodetic technique observing on radio frequencies (up to now) which can determine periods of the last two solar cycles.

5.2.3 Formal error of the estimated VTEC

Taking the variance-covariance matrix of the unknowns (equation (5.31)) $\mathbf{Q}_{\mathbf{xx}}$ and computing the vector \mathbf{p}^T , which contains the partial derivatives of the piece-wise linear function

$$\begin{aligned} \mathbf{p}^T &= \left(\frac{\partial VTEC(t)}{\partial offset}, \frac{\partial VTEC(t)}{\partial rate_0}, \frac{\partial VTEC(t)}{\partial rate_1}, \dots, \frac{\partial VTEC(t)}{\partial rate_i} \right) \\ &= (1, (t_1 - t_0), (t_2 - t_1), \dots, (t - t_i)) \end{aligned} \quad (5.35)$$

one can obtain the formal error of the estimated VTEC at time t by

$$\sigma_{VTEC(t)} = \sqrt{\mathbf{p}^T \mathbf{Q}_{\mathbf{xx}} \mathbf{p}}. \quad (5.36)$$

Figure 46 shows an exemplary plot of estimated VTEC values with their formal error, for station WETTZELL of the second day of the CONT02 campaign (October 17th, 2002). The formal errors of

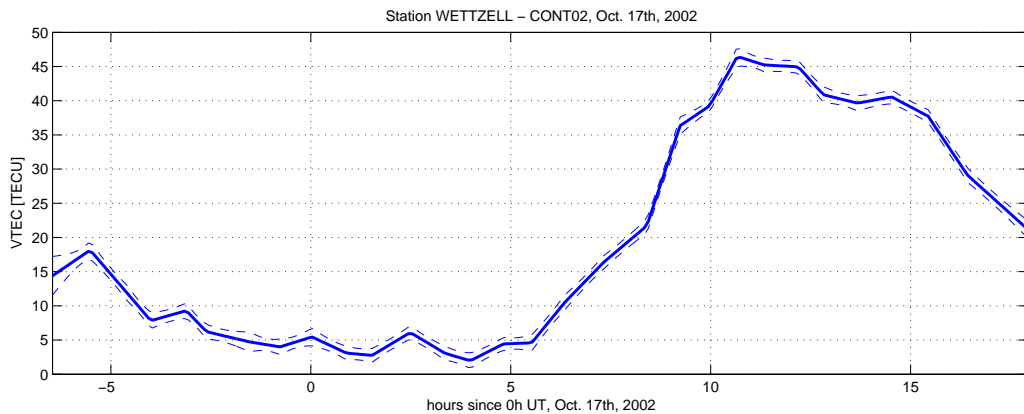


Figure 46: VTEC together with its formal error at station WETTZELL, Oct. 17th, 2002.

the estimated VTEC values, represented by piece-wise linear functions, were computed after the adjustment process as well. As an example figure 47 shows how the precision of the obtained parameters changes with time. Generally, the formal error scales with the magnitude of the vertical total electron

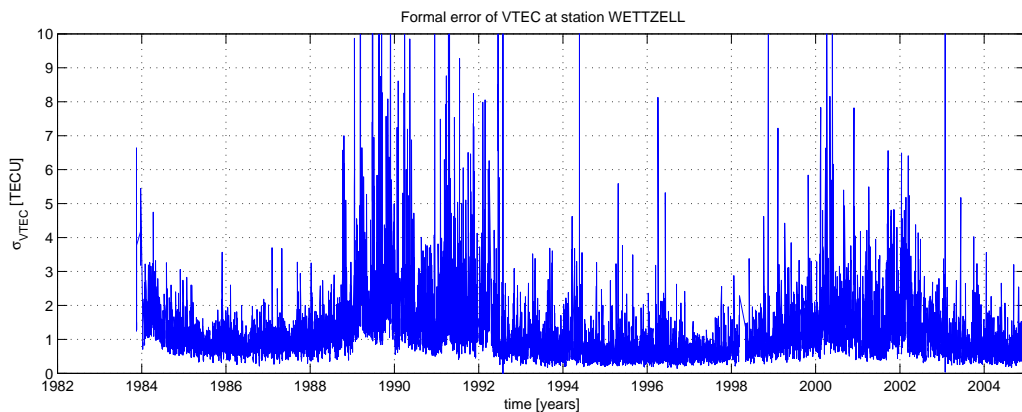


Figure 47: Formal error (1σ level) of the VTEC values at station WETTZELL.

content and therefore shows a strong correlation with the solar cycle. Computing the mean value for all sessions at station WETTZELL gives an averaged formal error of 1.03 TECU.

5.3 Gradient-like parameter c_i

The station-dependent model, introduced in section 5.1.1, relates distant measurements to the coordinates of the station. After “rotating” the observation to the same geomagnetic meridian a gradient-like constant c_i , set up for each station, takes into account latitudinal variations around the site. If

$VTEC(\phi_i)$ is the modeled vertical total electron content at an arbitrary station i at geomagnetic latitude ϕ_i , then the VTEC at latitude ϕ_i^* can be expressed by

$$VTEC(\phi_i^*) = (1 + c_i \Delta\phi_i) VTEC(\phi_i) \quad \text{with} \quad \Delta\phi_i = \phi_i^* - \phi_i. \quad (5.37)$$

Solving for c_i in equation (5.37) gives

$$\begin{aligned} c_i &= \frac{VTEC(\phi_i^*) - VTEC(\phi_i)}{\underbrace{VTEC(\phi_i)}_{\kappa}} \frac{1}{\Delta\phi_i} \\ &= \frac{\kappa}{\Delta\phi_i} \end{aligned} \quad (5.38)$$

where κ describes the relative change of VTEC when the ionospheric pierce point is thought to vary by $\Delta\phi_i$. Therefore the coefficient $100 \cdot c_i$ can be interpreted as the percentage increase/decrease of vertical total electron content by one degree of geomagnetic latitude. Thus, it can be treated similar to a gradient, having a latitudinal component c_i and longitudinal component equal to zero in a geomagnetic system. From the adjustment of all geodetic VLBI sessions also time-series of the parameter c_i with its formal error are obtained. This information is used to compute a weighted average value for each station, excluding stations which have contributed to less than 50 sessions. For a better representation the components of the gradient-like values are transformed in a geographic coordinate system and plotted as vectors at each station (figure 48). Positive values of c_i are plotted towards the geomagnetic north-pole, negative ones are pointing to the geomagnetic south-pole. The leading

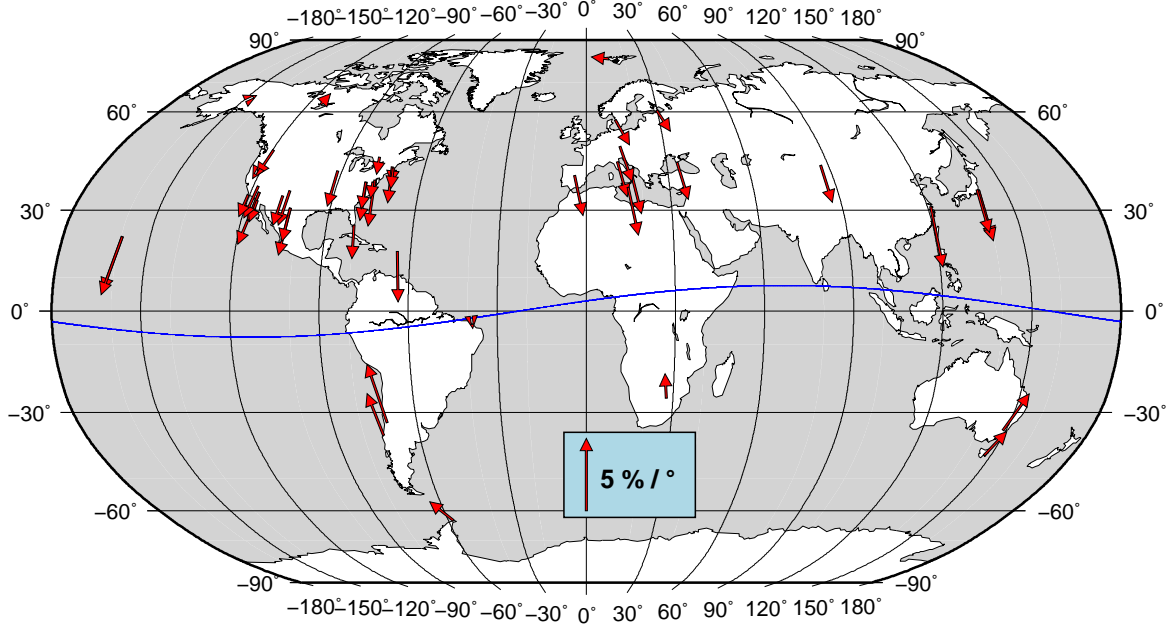


Figure 48: Global map of the averaged gradient-like parameters c_i .

sign of the parameter c_i depends whether the station is located north or south of the geomagnetic equator (blue line). Whereas stations at geomagnetic latitudes bigger than zero get negative gradients (southward directed), stations with geomagnetic latitudes less than zero reveal positive gradients (northward directed). Furthermore it can be noticed that gradients usually get larger for stations closer to the geomagnetic equator. The systematics of the result can be explained by the characteristics of the vertical total electron content distribution according to the geographical regions of the ionosphere. For locations in mid latitude regions the observed VTEC values get bigger around the geomagnetic equator. Stations in equatorial regions show up with large latitudinal gradients of VTEC, whereas the direction is determined by the station position. It can be seen very clearly that both Hawaiian stations get large southward directed gradients, as these sites are located slightly north of the northern equatorial anomaly. Station Fortaleza (Brazil, $\phi_m = -1.85^\circ$), located very close to the geomagnetic equator, is

supposed to get nearly symmetric latitudinal variations, which should result in a zero gradient. This is verified when an averaged value c_i for this station is computed. Stations in the auroral region get gradients directed to the geomagnetic poles, which is also consistent with phenomenological models of the ionosphere. Generally the magnitude of the obtained values of c_i agrees well for stations located close to each other, which can be seen in figure 48, too. Figure 49 shows the obtained parameters c_i depending on geomagnetic latitude. It can be seen that the leading sign of the parameter c_i changes

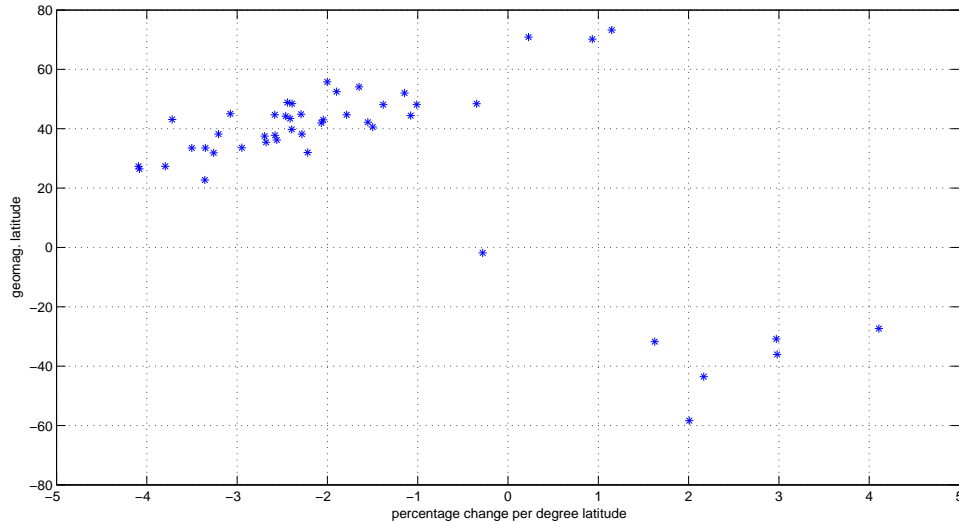


Figure 49: Averaged gradient-like parameters c_i depending on geomagnetic latitude, the units are percentage change of VTEC per degree latitude.

between the northern and southern hemisphere. Maximum gradients exist in regions around 30 degree around the geomagnetic equator and a percentage increase/decrease per degree of latitude of up to 7% can be found. Generally it can be stated that for stations located at mid latitude regions all mean gradients point towards the geomagnetic equator, which is consistent to theoretical models and measurements by other techniques.

5.4 Frequency characteristics of the results

The obtained VTEC time-series cannot be Fourier-transformed by standard algorithms like the Fast Fourier Transform (FFT) (Bracewell, 1986, [10]) as the data points are given at unequally sampled points. One can use interpolation techniques to transform the time-series to equally spaced samples to utilize such techniques as FFT, but artifacts in the frequency spectrum (so-called ghosts) will show up. The explanation for such effects is found in the scheduling of stations and experiments, which pretends artificial frequencies in the obtained spectra. The usual situation for VTEC values obtained from geodetic VLBI looks like this: data are collected within a session with an average sampling interval from about 10 minutes over one session, but between the sessions data gaps of three or more days happen rather often. Before a suitable algorithm for the computation of a frequency spectrum will be presented (section 5.4.1) the problem will be denoted mathematically. When $x(t)$ is a square-integrable continuous function, its spectrum can be obtained by the Fourier transformation *FT*.

$$X(\omega) = FT\{x(t)\} = \int_{-\infty}^{\infty} x(t)e^{-2\pi j\omega t} dt \quad (5.39)$$

In the discrete case, where N data points are given at arbitrary times t_i , only samples of the function $x(t_i)$ are available. This discretizing process can be described by a sampling function

$$s(t) = \frac{1}{N} \sum_{i=1}^N \delta(t - t_i) \quad (5.40)$$

using the Dirac delta function $\delta(t)$. Therefore the sampled signal can be expressed by $x_s(t) = x(t)s(t)$ and its Fourier transform is given by

$$X_s(\omega) = FT\{x_s(t)\} = FT\{x(t)s(t)\} = X(\omega) \otimes S(\omega) = \int_{-\infty}^{\infty} X(\omega')S(\omega - \omega') d\omega' \quad (5.41)$$

where \otimes is the convolution operator. Expression $S(\omega)$ is called the ‘‘spectral window function’’ and can be denoted by

$$S(\omega) = \int_{-\infty}^{\infty} s(t)e^{-2\pi j\omega t} dt = \frac{1}{N} \sum_{i=1}^N e^{-2\pi j\omega t_i}. \quad (5.42)$$

$X_s(\omega)$ is named the ‘‘dirty spectrum’’

$$X_s(\omega) = \int_{-\infty}^{\infty} f_s(t)e^{-2\pi j\omega t} dt = \frac{1}{N} \sum_{i=1}^N f_i e^{-2\pi j\omega t_i} \quad (5.43)$$

as the spectral content of the signal is contaminated by the spectral window function after convolution. For evenly spaced data with a sampling interval Δt the relation between sampling function and spectral window function is given by

$$s(t) = \frac{1}{N} \sum_{i=-\infty}^{\infty} \delta(t - i\Delta t) \quad \xleftrightarrow[FT^{-1}]{FT} \quad S(\omega) = \frac{1}{\Delta t} \sum_{i=-\infty}^{\infty} \delta\left(\omega - \frac{i}{\Delta t}\right). \quad (5.44)$$

After introducing the Nyquist-frequency $\omega_N = 1/2\Delta t$ the dirty spectrum for evenly distributed data can be written by

$$X_s(\omega) = \frac{1}{\Delta t} \left(X(\omega) + \sum_{i=1}^{\infty} X(\omega - 2i\omega_N) + X(\omega + 2i\omega_N) \right) \quad (5.45)$$

Equation (5.45) implicitly defines the sampling theorem as $x_s(t)$ is completely recoverable from $X_s(\omega)$ when the function $X(\omega)$ is zero for $|\omega| \geq \omega_N$. When the data are sampled by a box-function

$$s(t) = \begin{cases} 1 & \text{for } t_0 \leq t \leq t_N \\ 0 & \text{otherwise} \end{cases} \quad (5.46)$$

the spectral window function is given by

$$S(\omega) = \frac{\sin(\pi\omega T)}{\pi\omega} e^{-\pi j\omega(t_0+t_N)} \quad T = t_N - t_0. \quad (5.47)$$

This means that the obtained spectrum is smeared (spectral leakage) having a width (frequency resolution) determined by the data length T ($\delta\omega \approx 1/T$). Those two simple cases show that even for equally sampled data difficulties can occur, when the obtained dirty spectrum has to be interpreted. Appendix B shows the theory of Fourier transformations and presents the Lomb-Scargle algorithm, which can be utilized to compute spectra of irregularly sampled time-series. Another very promising algorithm, discussed in the next sub-section was presented by Baisch and Bokelmann (1999, [4]) in the field of geosciences.

5.4.1 Treatment of irregularly sampled data - the CLEAN algorithm

If data points of the time-series x_i are missing, the gained spectrum $X_s(\omega)$ is equivalent to the dirty spectrum of a data set, in which the missing data points have zero values. This effect can be removed, when the spectral window function is eliminated from the dirty spectrum. It is not possible to carry out a deconvolution in frequency domain, as the sampling function has mostly zero values (Roberts et al., 1987, [59]). Furthermore the authors recommended to estimate the (complex) amplitude of a

cosinusoidal, then remove its influence on the dirty spectrum together with the impact from the side lobes and start the next iteration. Taking a single harmonic with frequency $\tilde{\omega}$

$$x(t) = A \cos(2\pi\tilde{\omega}t + \Phi) \quad (5.48)$$

with (real) amplitude A and frequency Φ gives the spectrum

$$X(\omega) = a\delta(\omega - \tilde{\omega}) + a^*\delta(\omega + \tilde{\omega}) \quad (5.49)$$

where $*$ denotes the complex conjugate and $a = (A/2)e^{j\Phi}$. When $f(t)$ is sampled at N discrete times and the convolution theorem is applied to this time-series, the dirty spectrum of expression (5.48) becomes

$$X_s(\omega) = aS(\omega - \tilde{\omega}) + a^*S(\omega + \tilde{\omega}). \quad (5.50)$$

If $S(0) = 1$ the amplitude at the peak frequency $\tilde{\omega}$ can be given by $F_s(\tilde{\omega}) = a + a^*S(2\tilde{\omega})$. Getting a similar expression for F_s^* and substituting a^* in the prior expression makes it possible to determine the amplitude a of the peak by

$$a(\tilde{\omega}) = \frac{F_s(\tilde{\omega}) - F_s^*(\tilde{\omega})S(2\tilde{\omega})}{1 - |S(2\tilde{\omega})|^2} \quad (5.51)$$

if the frequency $\tilde{\omega}$ is known. The so-called CLEAN algorithm uses expression (5.51) to find the (complex) amplitude of a cosine signal and remove its contribution to the spectrum F_s together with all its side lobes by equation (5.50). Roberts et al. (1987, [59]) gave the following iteration scheme:

1. Compute the dirty spectrum $F_s(\omega)$.
2. Start the iteration with the initial residual spectrum $R^0 = F_s$.
3. On the i -th iteration find the maximum frequency ω_{peak} in the previous residual spectrum R^{i-1} and calculate its complex amplitude using equation (5.51).
4. Use expression (5.50) to obtain the contributions of $a(\omega_{peak})$ to the dirty spectrum and form the residual spectrum R^i by subtracting a fraction g ($0 < g < 1$) of R^{i-1} :

$$R^i = R^{i-1} - g \left[a^i S(\omega - \omega_{peak}) + (a^i)^* S(\omega + \omega_{peak}) \right] \quad (5.52)$$

the subtracted fraction ga^i is stored in a clean component array at the locations ω_{peak} and $-\omega_{peak}$ and the next iteration step is started.

5. The iteration will stop if R^i drops below a certain threshold, which should be set to the noise level (Baisch and Bokelmann (1999, [4])) and the clean component array is convolved with a Gaussian function to obtain a reasonable frequency resolution. Finally, the residual spectrum of the last iteration is added.

One limitation of the CLEAN algorithm is, that according to equation (5.51) the spectrum of the sampling function must be known up to twice the frequency of the dirty spectrum. Thus, cleaning of the dirty spectrum is possible only up to half of the maximum frequency. Using the algorithm above CLEANed spectra of the VLBI outcomes were derived and compared to GPS results (details about the time-series are given in section 5.5). Furthermore values of solar radio flux at 10.7cm (F10.7) back to January 1st, 1984 were downloaded from the World Data Center for Solar-Terrestrial Physics, Chilton and analyzed, too. Again results for station WETTZELL will be presented here, as this site can provide the longest time-series. As values of F10.7 contain some data gaps the CLEAN algorithm also had to be applied to this time-series. Finally all outcomes were normalized by the biggest amplitudes in order to compare TEC numbers to F10.7. Figure 50 shows the gained spectra from VLBI, GPS, and F10.7. Only VLBI and F10.7 reveal a sharp peak which can be assigned to the main solar cycle. The maximum amplitude from VLBI results is found at 10.5 years, which is identical to the maximum from F10.7. As also expressed by figure 50 it can be seen that the ratio of the amplitudes of the diurnal, semi-annual and annual variations is about the same for VLBI and GPS. As expected, GPS cannot

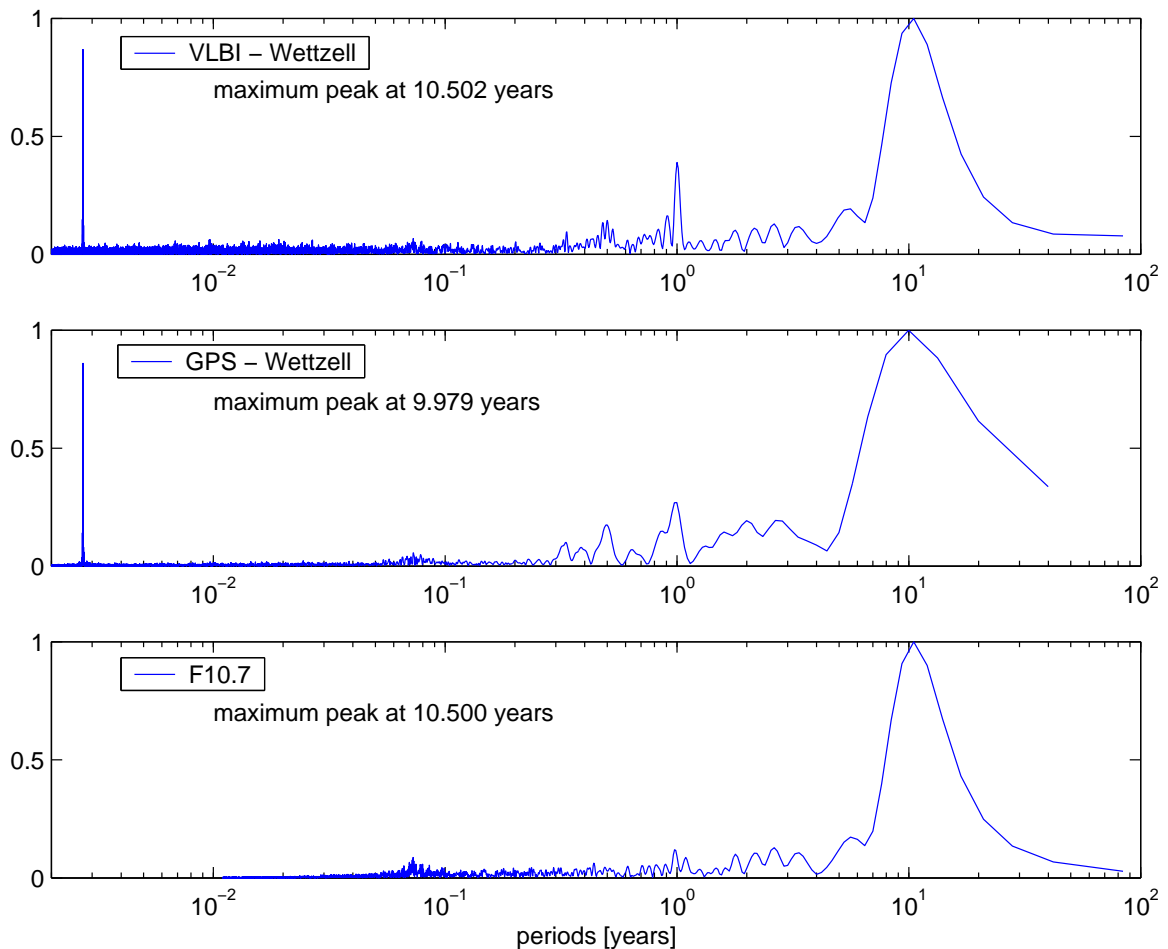


Figure 50: CLEANed spectra of VLBI and GPS (both WETTZELL, Germany) and F10.7.

detect longer periods than about 5 years in the ionosphere as time-series are not long enough. Station HARTRAO (South Africa) was taken as a representative station for the southern hemisphere. The estimated VTEC time-series at this station covers 19 years, whereas the first five years of observation contain huge data gaps up to one year (see figure 39). Using common Fourier methods like FFT, many artificial periods caused by interpolation to an equally sampled series, showed up in the spectrum. The results from the CLEAN algorithm are given in figure 51, together with the spectrum obtained from GPS. Again, only VLBI shows both, the annual and solar cycle period, in the gained spectrum. Even the large data gaps in the first 5 years of observation did not affect the maximum peak of the solar cycle a lot. The corresponding period is found to be at 10.8 years, which is also in quite good agreement to the value obtained from F10.7 data. Table 5 shows values of the dominant cleaned periods of the ten most active stations (see table 4) around the expected periods, which dominate the behavior of the ionosphere. It has to be mentioned, that both lunar tides around 28 days and differential rotation of the Sun (it rotates every 27 days at the equator and takes progressively longer to rotate at higher latitudes, up to about 36 days at the poles, (Stenflo, 1989, [71])) can be a possible explanation for peaks around 28 days in the spectra. Additionally a limitation is given due to scheduling of some VLBI stations, as several sessions are observed with gaps of three or four days, which can cause ghost signals around 27 and 28 days. Schaer (1999,[61]) uses a fixed 14.77 days period in his prediction algorithm, but his amplitude spectra of the mean GPS TEC do not show significant periods in this range. According to Střeštík (1998, [72]) the synodic period of solar rotation (Carrington rotation) is given by 27.2753 days and does not evoke a higher harmonic around 13.5 days. The latter period can be explained by a two-sector structure of the solar wind and the interplanetary magnetic field (IMF), respectively. The detected periods in the VLBI VTEC time-series are in good agreement to this theory. Station WETTZELL does not show significant (i.e. highest amplitude is 30% bigger than the 2nd highest one) periods in the 10 to 30 days search window, but reveals several peaks around

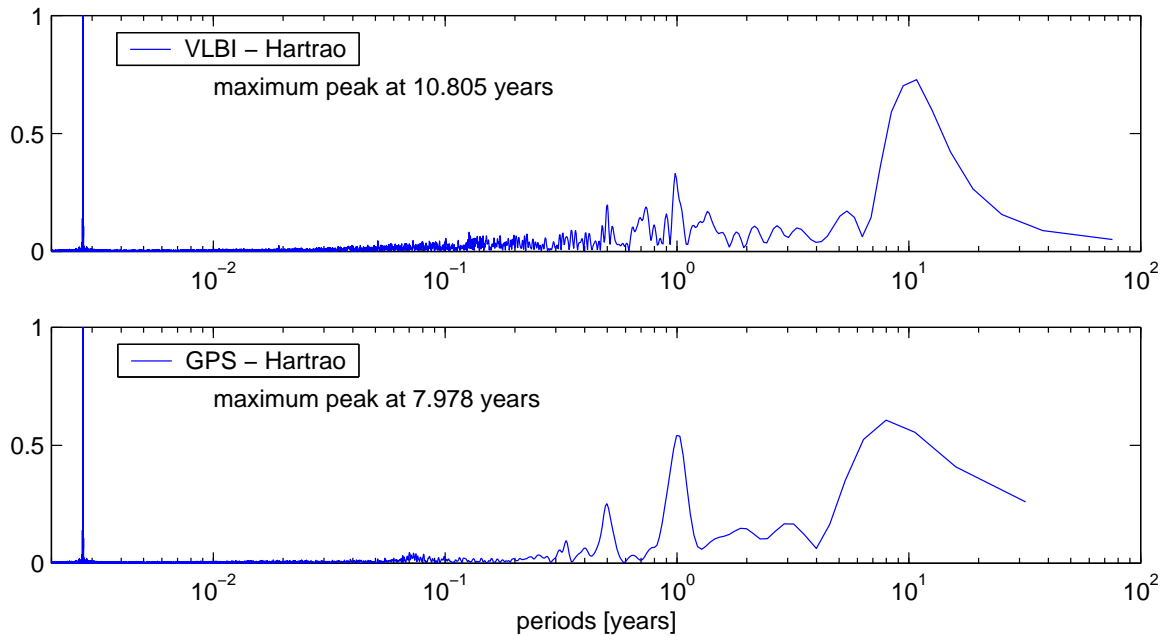


Figure 51: CLEANed spectra of VLBI and GPS (both HARTRAO, South Africa).

station	10 - 15 days	26 - 29 days	175-190 days	350-380 days	9 - 13 years
FORTLEZA	NS	NS	183.67	363.44	11.69
GILCREEK	NS	27.15	175.86	369.10	10.23
HARTRAO	13.99	26.74	181.75	363.50	10.81
HRAS085	13.52	28.93	183.55	367.10	X
KOKEE	13.99	27.29	181.62	375.36	11.56
MOJAVE12	13.25	26.10	178.90	353.10	X
NYALES20	13.44	27.60	189.36	364.87	X
RICHMOND	13.47	27.23	180.09	350.17	X
WESTFORD	13.96	26.31	182.34	370.57	10.48
WETTZELL	NS	NS	182.59	365.36	10.50

Table 5: Dominant periods at the ten most active geodetic VLBI stations, “X” marks periods which cannot be detected as the time-series is too short. “NS” represents spectra with no significant periods inside the search window.

13.5 and 27.3 days. The diurnal and sub-diurnal period domain are investigated using FFT spectra, whereas the data were interpolated to be equally spaced. The spectra for each station are computed and a mean spectrum for all VLBI sites is computed. Figure 52 shows the spectral content between 3 and 26 hours. The diurnal signal is dominant, having its maximum period at exactly 24 hours (the spectral resolution around the 24 hours period is about 5 seconds). Also the higher harmonics at 12, 8 and 6 hours can be found as significant peaks in the spectrum. This is in complete agreement with the periods detected by Kondo (1991, [43]), before he set up the parameterization of his Fourier approach for the determination of VTEC values from VLBI measurements. The apparent peak around 16 hours seems to be a higher harmonic artifact of the 8 hour period and was not detected by the CLEAN algorithm.

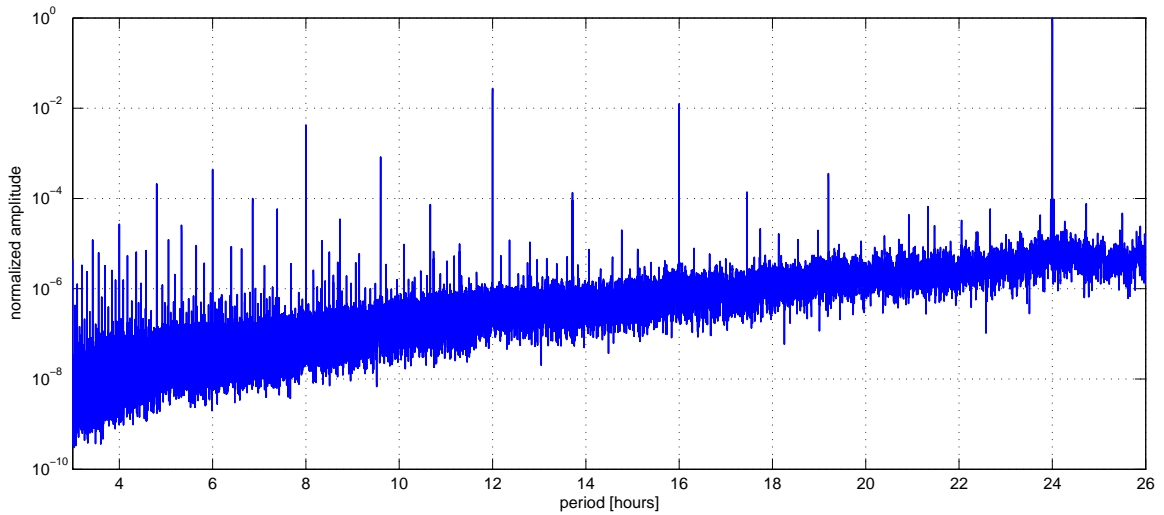


Figure 52: Normalized mean spectral content of all time-series in the diurnal and sub-diurnal domain.

5.4.2 Wavelet analysis

According to Keller (2004, [41]) a function $\psi(s)$ is called Wavelet function when its Fourier-transform $\Psi(\kappa)$ fulfills the condition

$$C_{\Psi} = \int_{-\infty}^{\infty} \frac{|\Psi(\kappa)|^2}{|\kappa|} d\kappa < \infty. \quad (5.53)$$

Then the continuous Wavelet transform $X(a, b)$ of a signal $x(t)$ is given by

$$X(a, b) = |a|^{-1/2} \int_{-\infty}^{\infty} x(t) \psi^* \left(\frac{t-b}{a} \right) dt \quad (5.54)$$

where a is the scale-parameter, which can be related to the frequency ω (Schmidt, 2001, [64]). The parameter b steers translations of the wavelet on the time-axis. Algorithms for fast computations of the Wavelet transform, so-called Multi-resolution analysis (MRA), can be found in Kaiser (1994, [39]) and Schmidt (2001, [64]). Although Foster (1996, [24]) proposed an algorithm to estimate Wavelet spectra from unevenly sampled time-series, the VLBI VTEC data were interpolated to regularly spaced intervals and standard algorithms as described in Schmidt (2001, [64]) were carried out. As the annual period domain was investigated only, artifacts due to interpolation are not expected to show up significantly in the range between 1 and 1000 days. For the following analysis the Morlet-Wavelet

$$\psi(s) = e^{j\kappa_0 s} \left(e^{-\frac{s^2}{2\sigma^2}} - \sqrt{2} e^{-\frac{\kappa_0^2 \sigma^2}{4}} e^{-\frac{s^2}{\sigma^2}} \right) \quad \kappa_0 = \pi \sqrt{\frac{2}{\ln 2}} \quad (5.55)$$

was used with the form-parameter σ , which was set to 1. Figure 53 shows the obtained Wavelet scalograms from VLBI and GPS data for station WETTZELL. The influence of the last two solar cycles can be clearly seen in the VLBI scalograms as the amplitude of the annual signal corresponds to the long-term variations of the activity of the sun. Scalograms from GPS results exactly confirm this variation in the annual and semi-annual period domain. Taking only the epoch covered by VLBI and GPS measurements and computing the quadratic cross-scalogram $W_{VLBI, GPS}(a, b) = W_{VLBI, VLBI}(a, b) \cdot W_{GPS, GPS}(a, b)$ shows the agreement of both time-series in the time frequency domain, whereas $W(a, b) = X(a, b) \cdot X^*(a, b)$ denotes the absolute magnitude of the Wavelet-transform $X(a, b)$. The right plot of figure 54 represents the normalized quadratic cross-scalogram between VLBI and GPS time-series and its normalized Wavelet coherency is given in the left part. It can be seen that the agreement is very good for the semi-annual and annual period domain and coherency is very high (> 0.8) for periods larger than 100 days. The results from station KOKEE park were selected

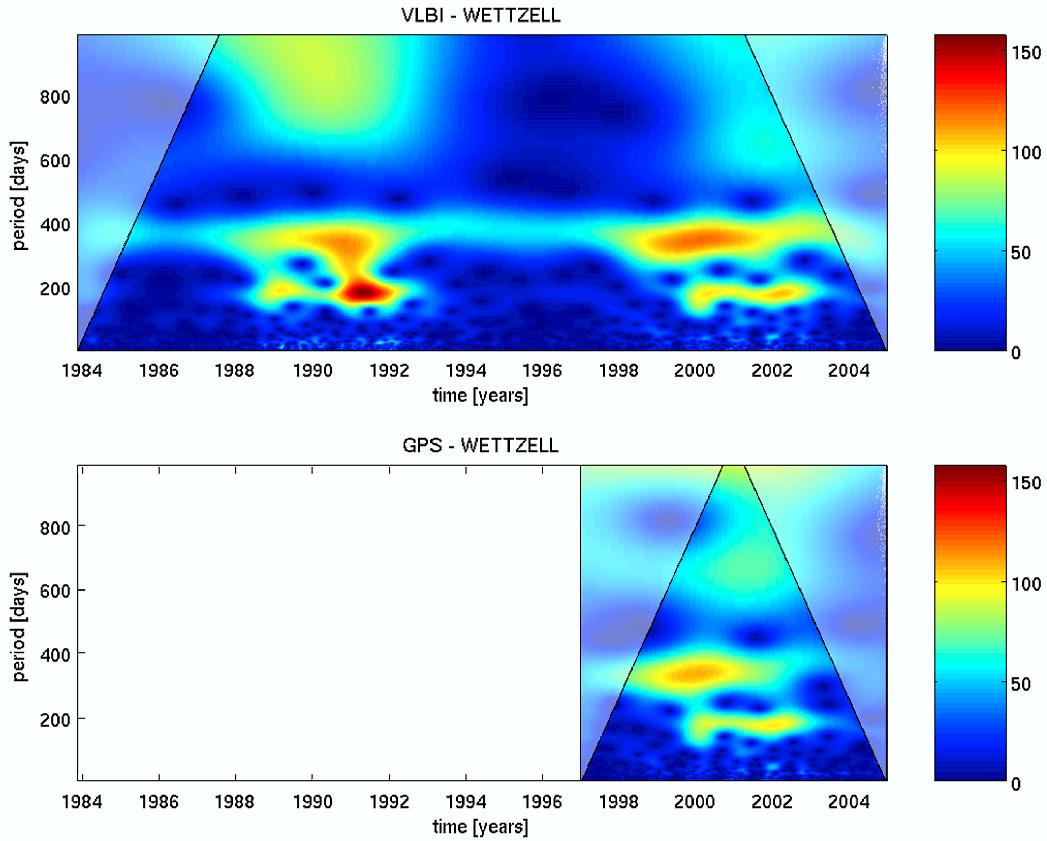


Figure 53: Wavelet spectra of VLBI and GPS (both WETTZELL).

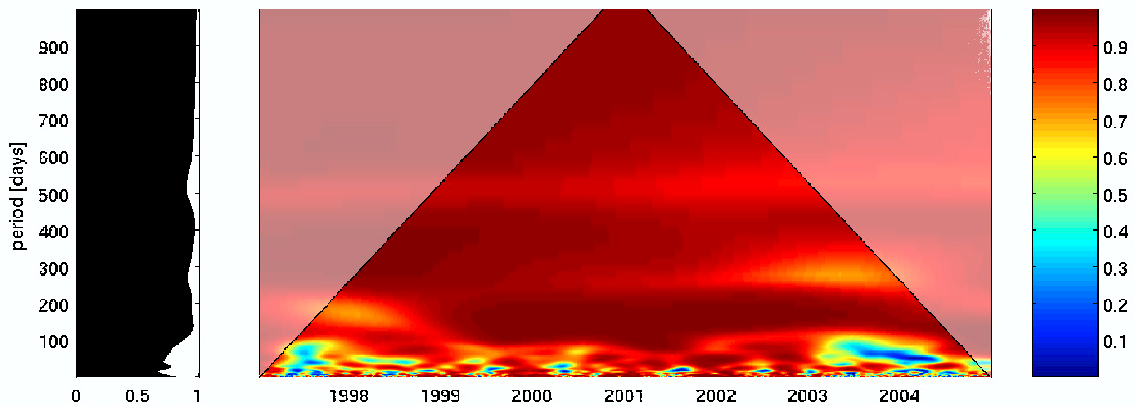


Figure 54: Quadratic cross-scalogram $|W_{VLBI,GPS}(a,b)|^2$ (right) and normalized Wavelet coherency (left) at station WETTZELL).

because this station is located in the equatorial region and shows high values of VTEC. Figure 55 displays the obtained Wavelet spectra and reveals a clearly annual signal from both techniques. But VLBI reveals only one strong peak, whereas the Wavelets scalogram of the GPS time-series shows an extended maximum. The reason for this is that VLBI does not observe continuously and station KOKEE did not carry out experiments on days of high TEC occurrence. For example the extremely high CME (coronal mass ejection) on 29 and 30 October 2003 had a remarkable impact on the ionospheric TEC, and peak VTEC levels clearly above 200 TECU could be observed. A similar event happened on November 20th, 2003 but station KOKEE started its observation at 18^h UT, so no noticeable impact in the VTEC time-series could be found. Therefore the VLBI Wavelet scalogram shows less energy

around that period. Thus, it is not a surprise that the cross-scalogram and the Wavelet coherence

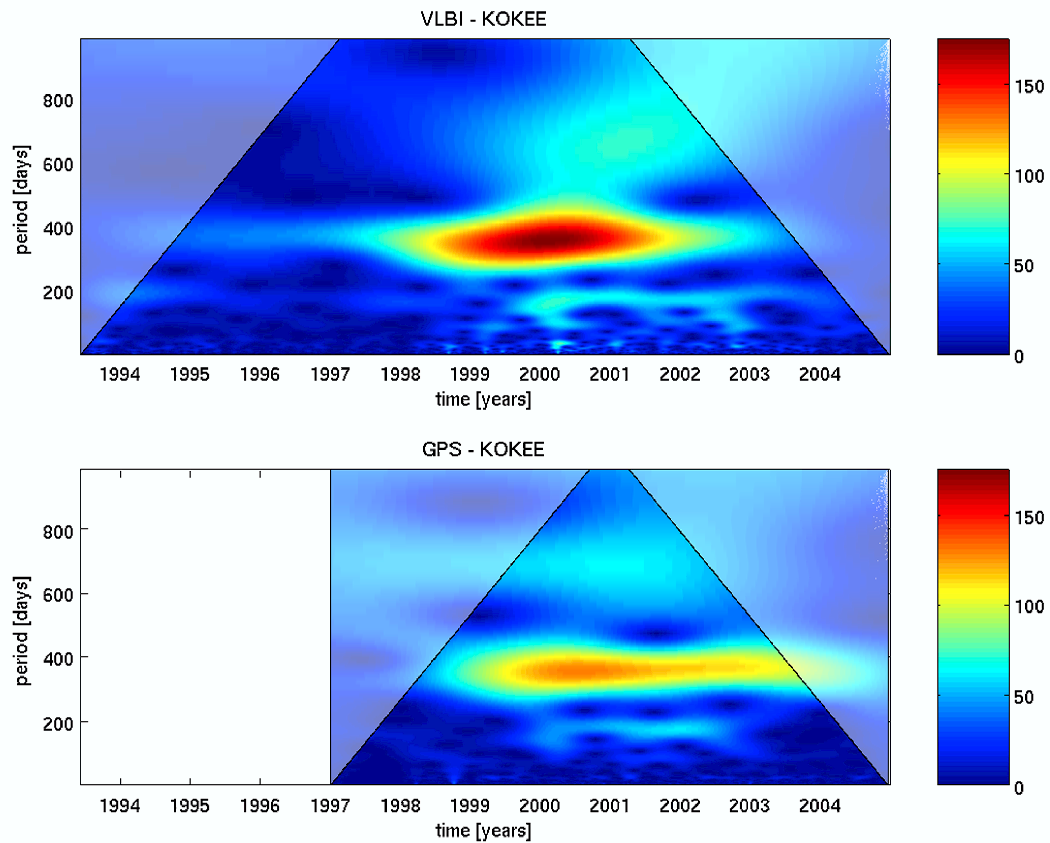


Figure 55: Wavelet spectra of VLBI and GPS (both KOKEE).

agree very well, except at the period when VLBI did not observe high TEC values as the station was not scheduled at that time. Figure 56 depicts the obtained cross-scalogram and the corresponding Wavelet coherence. Furthermore it can be seen that the coherency plots for WETTZELL and for KOKEE show a peak around 30 days, which can be related to the effects discussed in the prior section.

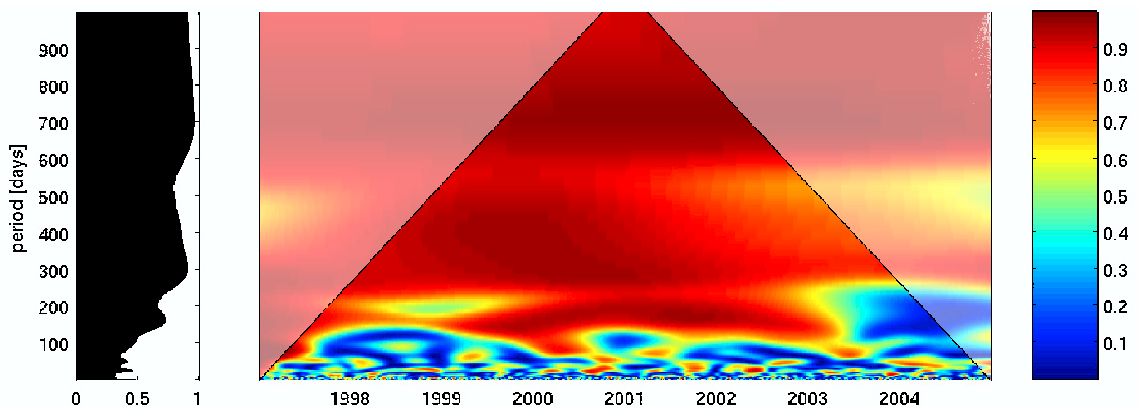


Figure 56: Quadratic cross-scalogram $|W_{VLBI,GPS}(a,b)|^2$ (right) and normalized Wavelet coherency (left) at station KOKEE.

5.5 Comparison to ionospheric models derived from GPS measurements

For a cross-technique comparison results from a worldwide GPS station network were used. The International GNSS Service (IGS) provides Global Ionospheric Models (GIM) of VTEC values on a geographical grid ($\Delta\lambda = 5^\circ$, $\Delta\phi = 2.5^\circ$) with a time resolution of two hours (IGS, 2005, [37]). These maps are stored in a format called IONEX (Schaer, 1998, [63]) together with the formal error evaluated at each grid point. The official IGS product is a weighted combination of the results from different analysis centers and is validated by satellite altimetry measurements. The accuracy of the so-called “Final Ionospheric TEC grid” is expected to be between 2 and 8 TECU, whereas the VTEC maps have a formal error of 1 to 2 TECU. As the official combined product has only been available since Dec. 15th, 2002 it was not considered for comparisons with VLBI results. Instead, the solution from the Center for Orbit Determination in Europe (CODE), Astronomical Institute of the University Berne, which is one of the 5 analysis centers contributing to the official IGS IONEX solution, was selected. Using a proper interpolation method, suggested by Schaer et al. (1998, [63]) values of VTEC can be obtained for each VLBI station and GPS time-series from Jan.1st, 1997 until Dec. 31st, 2004 were generated. The GPS results were used for finding an optimal weight parameter k by comparing them with selected VLBI sessions before the whole IVS database was processed (section 5.5.1). Furthermore, the VTEC values obtained from VLBI were compared to the GPS times-series in order to detect biases between the techniques (section 5.5.2) and to compare the spectral characteristics as already discussed in section 5.4.

5.5.1 Finding the optimum weight parameter k from CONT02 measurements

Before all available geodetic VLBI databases were analyzed a test run was made using the first nine days of CONT02³, which was a two-week campaign of continuous VLBI sessions scheduled from October 16th until October 31st. As on October 25th one station (NYALES20) suffered from a power outage and on the following days other problems occurred, only sessions until October 24th were taken, analyzed and compared to VTEC values from the International GNSS Service (IGS), provided in IONEX format. The parameter k , which steers the weighting according to the zenith distances (equation (5.29)) was set to integer values of $k = -6, -5, \dots, 7, 8$. Table 6 gives the a-posteriori variance-factor σ_0 (equation (5.30)) for different values of k . When k gets larger, the a-posteriori

k	Oct.16	Oct.17	Oct.18	Oct.19	Oct.20	Oct.21	Oct.22	Oct.23	Oct.24	mean
-6	32.33	27.46	27.70	30.28	30.97	30.56	30.02	28.58	27.06	29.44
-5	24.27	21.65	20.73	24.60	24.90	23.42	23.37	22.07	20.77	22.86
-4	18.62	15.02	15.53	18.46	17.67	16.35	18.89	16.58	15.19	16.92
-3	14.73	11.83	12.11	13.17	13.52	13.13	13.97	13.10	11.25	12.98
-2	12.73	9.43	9.88	10.19	9.73	10.00	10.79	10.40	8.76	10.21
-1	9.66	7.40	7.53	7.94	7.95	7.84	8.22	7.94	6.41	7.88
0	7.39	5.96	5.96	6.28	6.72	6.35	6.21	6.38	5.06	6.26
1	5.37	4.50	4.41	4.91	5.22	4.51	4.82	4.77	3.80	4.70
2	3.83	3.72	3.52	3.80	3.81	3.53	3.64	3.65	2.88	3.60
3	3.04	2.79	2.65	2.86	2.87	2.59	2.81	2.79	2.20	2.73
4	2.27	1.99	2.02	2.21	2.06	1.92	2.10	2.21	1.68	2.05
5	1.73	1.48	1.51	1.65	1.48	1.52	1.58	1.68	1.24	1.54
6	1.28	1.13	1.14	1.22	1.17	1.18	1.20	1.29	0.93	1.17
7	0.94	0.85	0.88	0.95	0.87	0.89	0.89	0.95	0.70	0.88
8	0.70	0.62	0.66	0.67	0.63	0.66	0.67	0.70	0.58	0.66

Table 6: The a-posteriori variance-factor σ_0 for different values of k , applied to CONT02 data.

variance-factor gets down, which is confirmed by figure 36. Big values of k downweight observations

³contributing stations: ALGOPARK, GILCREEK, KOKEE, HARTRAO, NYALES20, ONSALA60, WESTFORD, WETTZELL

at low elevations and therefore decrease the errors introduced by the model assumption. If k is set to large negative values, more weight is given to low elevation observations, which increases the residuals (and therefore the a-posteriori variance-factor) when the transformations described in section 5.1.1 cannot entirely describe the spatial gradients around the station. Figure 57 shows the mean value of σ_0 as a function of k . According to Koch (1997, [42]) a value of $\sigma_0 = 1$ (i.e. between $k = 6$ and $k = 7$)

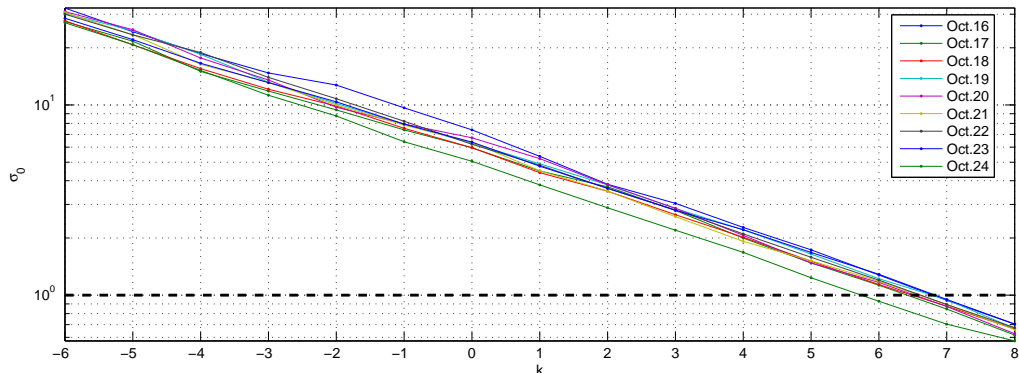


Figure 57: Dependency of the mean a-posteriori variance-factor σ_0 on the parameter k for the first 9 days of CONT02.

means that the stochastic model of the adjustment problem was chosen properly. As a measure of accuracy the mean VTEC difference between VLBI and GPS was computed for each parameterization k . Table 7 gives these differences in TECU together with the mean value for all 9 days. From this

k	Oct.16	Oct.17	Oct.18	Oct.19	Oct.20	Oct.21	Oct.22	Oct.23	Oct.24	mean
-6	0.55	0.49	1.62	-0.65	0.96	1.73	0.87	0.49	0.21	0.70
-5	1.00	1.05	2.18	-0.14	0.43	1.63	1.06	0.70	0.03	0.88
-4	1.41	0.69	2.02	-0.17	0.69	2.46	1.23	1.33	-1.06	0.96
-3	1.63	0.65	2.27	0.32	0.38	1.70	1.10	1.03	-1.07	0.89
-2	1.50	0.61	1.97	-0.24	0.61	1.87	1.12	0.94	-0.26	0.90
-1	0.87	0.39	1.91	-0.43	1.19	1.27	1.30	0.96	-1.46	0.67
0	0.99	0.19	1.91	0.00	1.08	1.58	0.93	0.27	-0.63	0.70
1	0.71	0.53	1.87	-1.19	1.48	0.68	1.02	0.87	-1.74	0.47
2	0.87	0.40	1.73	-0.65	1.09	1.00	1.05	0.35	-1.36	0.50
3	-0.08	1.26	1.82	-0.12	1.14	0.14	0.97	0.28	-1.60	0.42
4	0.13	1.12	1.49	0.27	0.17	-0.04	1.03	-0.10	-1.74	0.26
5	-0.34	0.95	1.21	0.34	0.06	0.17	0.67	-1.02	-2.17	-0.01
6	-0.19	1.34	0.88	0.82	-0.10	0.04	0.38	-1.45	-2.66	-0.11
7	-0.76	1.28	0.72	1.08	-0.56	-0.36	0.17	-1.76	-2.89	-0.34
8	-1.20	1.27	0.66	0.59	-0.62	-0.55	0.03	-1.70	-3.03	-0.51

Table 7: Mean difference between GPS and VLBI computed for all stations within CONT02.

table it can be seen that the differences become smaller when k increases (the minimum absolute difference can be found for $k = 5$). Again, downweighting of low elevation observations reduces the errors which might be introduced from the station-dependent model. If k is set to a high value (> 10) it becomes more difficult to solve for reliable instrumental offsets as the constant term of the piece-wise linear function correlates highly with these instrumental biases and de-correlation is only granted by low elevation observations, which have a mapping function value $M(z)$ significantly larger than one. Besides the differences between both techniques, also the root mean square values was analyzed. It reveals how good the time-series of each technique agree according to their shape. Table 8 gives the R.M.S. values for different choices of k . It can be seen that for values of $k = 4$ and $k = 5$ the overall R.M.S. is smallest. As a trade off it was decided to use $k = 6$ to run all sessions that are available from the IVS database. In the paper by Hobiger and Schuh (2004, [33]) a value $k = 4$ was recommended,

k	Oct.16	Oct.17	Oct.18	Oct.19	Oct.20	Oct.21	Oct.22	Oct.23	Oct.24	mean
-6	10.03	6.23	7.96	7.21	7.17	6.36	6.51	8.34	6.53	7.37
-5	10.32	5.96	7.95	6.87	6.85	6.63	6.30	7.93	6.31	7.24
-4	9.99	6.28	7.81	6.62	6.60	6.10	5.87	7.64	5.60	6.95
-3	9.55	5.41	7.60	6.56	6.32	6.10	6.36	7.76	5.73	6.82
-2	8.85	5.92	7.49	6.69	5.89	6.16	5.68	6.99	6.00	6.63
-1	8.33	5.30	7.30	6.69	5.77	5.93	5.59	6.89	5.43	6.36
0	8.00	5.29	6.97	6.17	5.93	5.36	5.41	7.24	5.65	6.22
1	7.35	4.93	6.42	6.33	5.55	5.30	4.97	6.82	5.28	5.88
2	7.28	4.90	6.05	5.72	5.13	4.76	4.87	6.81	5.26	5.64
3	6.48	4.90	5.50	5.50	5.32	4.76	4.66	6.75	5.27	5.46
4	6.00	4.78	5.07	5.10	5.12	4.69	4.63	6.62	5.09	5.23
5	6.52	4.75	4.87	4.93	5.38	4.51	4.77	6.14	5.17	5.23
6	6.71	5.19	5.03	4.87	5.39	4.76	5.03	6.00	5.21	5.35
7	6.90	5.22	5.15	4.85	5.73	4.82	5.25	5.83	5.22	5.44
8	7.21	5.48	5.31	4.84	6.26	5.05	5.54	5.75	5.31	5.64

Table 8: R.M.S. of the difference between GPS and VLBI computed for all stations within CONT02.

but at that time the station model did not take into account spatial gradients around each station, which can also be seen as a reason why the mean differences between GPS and VLBI, presented in that paper, were larger than here.

5.5.2 Differences between GPS and VLBI results

All VTEC values from VLBI were compared to GPS, by interpolating the CODE time-series to the VLBI epochs. As already discussed at the beginning of this section, comparisons could only be made for observations after Jan. 1st, 1997, since IONEX maps are not available for earlier measurements. Due to the fact that only vertical total electron content values were compared, no mapping functions, which can introduce relative errors up to 20 % as discussed in section 5.1.4, were needed. In the following plots and tables the difference between both techniques is always computed in the sense of “GPS minus VLBI”. Figures 58 and 59 show histograms of differences between both techniques, for station WETTZELL and HARTRAO, i.e. in mid latitude regions. At both stations the agreement

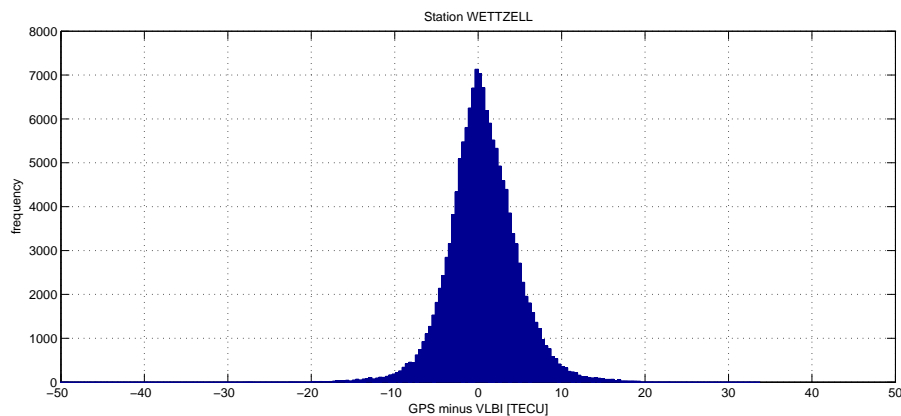


Figure 58: Differences of VTEC at station WETTZELL between GPS and VLBI.

between VTEC values from GPS and VLBI is very good, showing a mean difference of only 0.5 TECU and a standard deviation of about ± 4 TECU for station WETTZELL. For HARTRAO the mean difference is found to be 0.6 TECU having a standard deviation of ± 8.5 TECU. The scattering of the differences is related to the geomagnetic latitude, which will be shown later. Again, station KOKEE is taken as a station from the equator region, and VTEC values of both techniques were compared against each other. Figure 60 shows the histogram of the discrepancies with a mean difference of 2.6

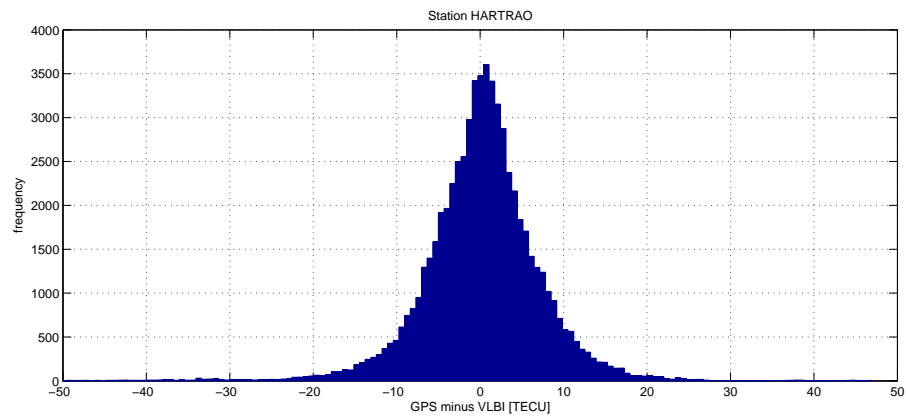


Figure 59: Differences of VTEC at station HARTRAO between GPS and VLBI.

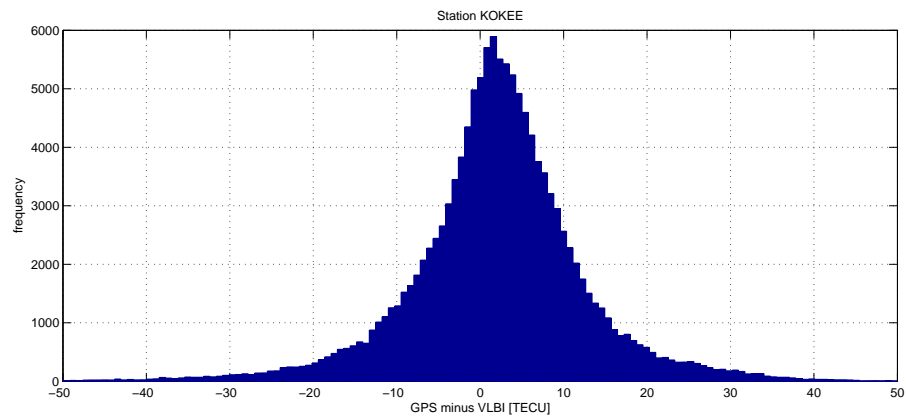


Figure 60: Differences of VTEC at station KOKEE between GPS and VLBI.

TECU and a standard deviation of ± 10.2 TECU. VLBI seems to underestimate the ionosphere slightly at this site, which can be explained by assumptions made for the station-dependent model. As this VLBI antenna is located in a region of high ionospheric activity, TEC values will reach high values and any error, introduced by the model, which transforms between values at the ionospheric pierce point and the station coordinates, will scale proportionally to TEC. Station NYALES20 is representative for a station in the auroral region. A histogram of the differences between VTEC from GPS and VLBI at that station is given in figure 61. The mean difference for that site is 1.6 TECU with a standard

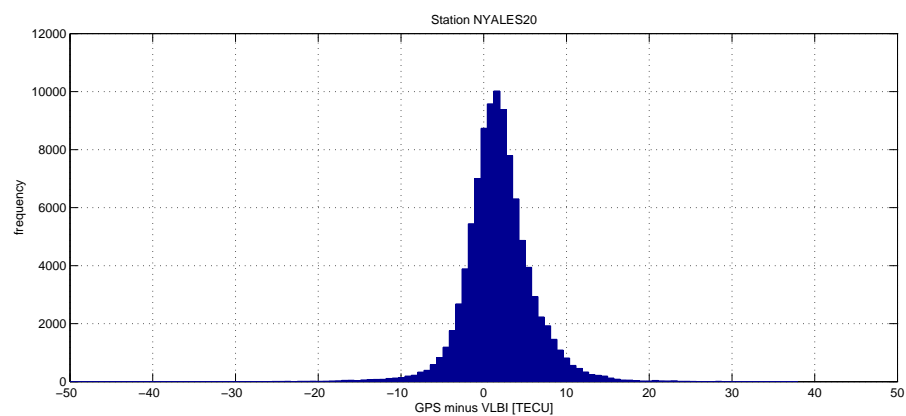


Figure 61: Differences of VTEC at station NYALES20 between GPS and VLBI.

deviation of ± 4.6 TECU. Table 9 contains the statistical information about the discrepancies between GPS and VLBI for all stations. Several stations are missing as they were already dismantled or assigned to other than geodetic VLBI purposes before the year 1997, when GPS VTEC series became available.

Figure 62 illustrates by a histogram the differences between GPS and VLBI, calculated for all stations.

Station	GPS-VLBI	Std. dev.	Station	GPS-VLBI	Std. dev.
AIRA	6.12	11.54	MATERA	1.46	4.74
ALGOPARK	-0.16	4.71	MEDICINA	0.71	3.52
BR-VLBA	0.11	1.95	MIZNAO10	8.24	11.39
CHICH10	6.55	18.72	NL-VLBA	1.06	2.35
CRIMEA	1.26	6.06	NOTO	1.34	4.85
DSS15	1.86	7.01	NRAO20	0.51	2.71
DSS45	-2.36	5.07	NYALES20	1.59	4.60
DSS65	1.04	3.70	OHIGGINS	-4.44	8.04
EFLSBERG	0.35	5.06	ONSALA60	-0.43	3.58
FD-VLBA	2.80	8.16	PIETOWN	0.93	3.80
FORTLEZA	3.19	10.54	SC-VLBA	3.30	5.41
GGAO7108	0.10	5.32	SESHAN25	4.89	9.02
GILCREEK	0.63	4.25	SVETLOE	-0.59	3.43
HARTRAO	0.55	8.49	TIGOCONC	-0.33	11.25
HN-VLBA	0.32	5.90	TIGOWTZL	-0.25	4.58
HOBART26	-1.92	8.58	TSUKUB32	4.00	8.79
KASHIM34	1.21	10.10	URUMQI	2.24	5.71
KASHIMA	1.40	12.51	WESTFORD	2.27	4.78
KOKEE	2.60	10.24	WETTZELL	0.50	4.04
KP-VLBA	2.05	3.77	YEBES	0.36	3.87
LA-VLBA	1.56	3.00	YLOW7296	0.74	4.46

Table 9: Mean differences in TECU between GPS and VLBI and corresponding standard deviation computed for all stations.

The overall mean difference is found to be 1.03 TECU, with a standard deviation of ± 7.63 TECU. Figure 63 shows the differences between both techniques as a function of geomagnetic latitude. One

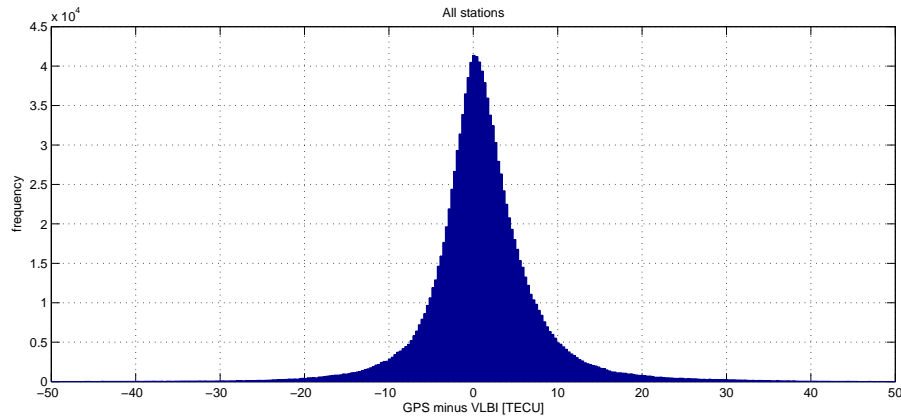


Figure 62: Differences of VTEC at all used stations between GPS and VLBI, mean difference 1.03 TECU, standard deviation ± 7.63 TECU.

can see that for mid latitude regions both agree well and differences and their corresponding standard deviations get larger when stations are located in the equator region. Another feature for stations in high latitude regions can be denoted in figure 61. Total electron content at those regions reaches only rather small values, so outliers in the VLBI data can evoke negative estimates of VTEC. But these are avoided by the non-negative constraints of the adjustment process and forced to be at least zero. Therefore the histogram of the differences between stations in high latitude regions is skewed as it can be seen for station NYALES20. As GPS satellites are orbiting the Earth in a height of about 20000 km VLBI measurements should provide slightly higher TEC estimates since additional delays in

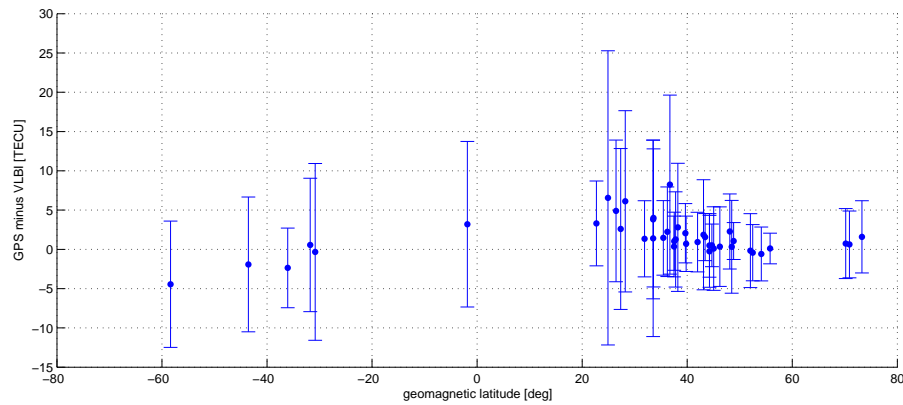


Figure 63: VTEC differences in dependency of the geomagnetic latitude.

the plasmasphere increase these values. If the influences caused by longer way-paths through plasma are assumed to be less than 2 TECU, it can be stated that VLBI is underestimating the ionosphere by about 3 TECU. When looking at the differences and their standard deviation (e.g. figure 63) one can see that both are roughly scaled by the total electron content which is strongly related to geomagnetic latitude. The only linear scaling parameter in the estimation process is the mapping function, which leads to the following explanation for the bias between both techniques. When the height of the mapping function is reduced, the mapping function itself gets larger. As in the adjustment the model parameters must fit to the observed values, VTEC is estimated with smaller values. The CODE team estimates global TEC on the basis of spherical harmonics expansions dealing with a fixed height of the single layer at $H = 506.7$ km and the modified mapping function, described by equation (4.50). For the estimation of vertical total electron content values from VLBI measurements a time-varying height model, using values from the IRI2001, model was applied. Comparing the global mean of the theoretical heights with the values used in CODE's GPS processing reveals differences up to 100 km. If VTEC values were computed with the GPS single layer heights, the differences would become smaller than those found here. But due to the fact that VLBI can only observe in one direction at a time, the geometry is weak and the proper height of the single layer can have a huge impact on the obtained VTEC values. Additionally it has to be mentioned, that the cut-off elevation angle for VLBI measurements is lower than the one of GPS and many experiments are scheduled to observe at low elevations to distinguish height, clock and tropospheric parameters better within the adjustment process. Another portion of the underestimation of VTEC might be the fact that plasmaspheric contribution is canceled out between two VLBI stations even though they are several 1000s of km apart, as the plasmasphere is more uniform and homogeneous in space and time than the ionosphere. The plasmaspheric contribution cannot be recovered completely by VLBI modeling strategies and therefore GPS provides slightly higher values since plasmaspheric contributions up to 20000 km (i.e. the orbit height) are contained in the measurements. Nevertheless the agreement between both techniques is astonishing and confirms the results of GPS within their formal error.

5.6 Comparison to satellite altimetry data

Several altimetry satellites are equipped with dual-frequency transmitters/receivers to correct measurements for ionospheric delays. Thus, this information can also be used to compute values of total electron content. Since the transmitted signal is reflected from sea surface and received by the satellite again, assuming a perpendicular reflection, only VTEC values can be measured. Comparison to other techniques is complicated as all conventional measurement systems are based on land, but altimeter foot-prints can only be found on the oceans or in big lakes. IGS uses measurements from altimeter missions to validate the quality of their global ionospheric maps over the oceans (Hernández-Pajarez, 2005, [31]) or regional GPS VTEC models can be used to calibrate altimeter instruments as shown by Ping et al. (2004, [55]). Station KOKEE, located on Hawaii, is the only VLBI site, which has a long history of observations and is completely surrounded by the ocean. Therefore all measurements

with corresponding footprints within a distance of 0.5° around station KOKEE (spatial gradients are expected to be small within this radius) were obtained through the Altimeter Database System (ADS, 2005, [3]). Time-series fulfilling the spatial filter criteria were obtained for TOPEX/POSEIDON and JASON-1 measurements. Figure 64 shows the footprints of the observations which can be used for comparison to VLBI results for station KOKEE. TOPEX/Poseidon is a joint mission of the Na-

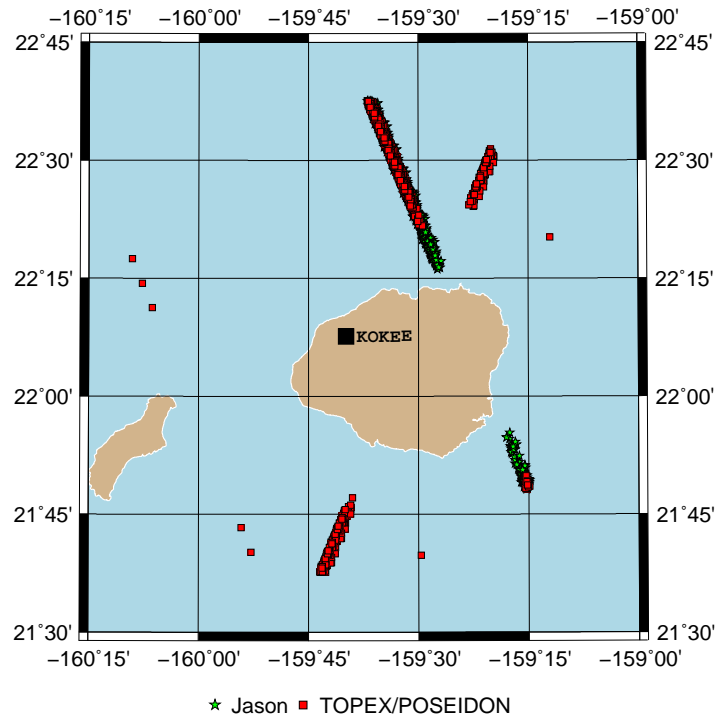


Figure 64: Footprints of different altimetry satellites, within a radius of 0.5 degrees from KOKEE.

tional Aeronautics and Space Administration (NASA) and the French space agency, Centre National d'Etudes Spatiales (CNES) using a state-of-the-art dual-frequency radar altimeter and three independent precision orbit determination systems. It was launched on August 10, 1992 and the satellite orbits the Earth at an altitude of 1336 km with an inclination angle of 66 degrees and a period of 112 minutes and is still in orbit. The altimeter operates at two separate frequencies, at 13.6 GHz (Ku-band) and 5.3 GHz (C-band), which makes it possible to determine VTEC values. JASON-1 is the first satellite in a series designed to ensure continued observation of the oceans for several decades. It is the follow-on mission of TOPEX/POSEIDON, operating on the same frequency bands, and was launched on December 7, 2001. The orbital height is equal to TOPEX/POSEIDON's height and the satellite is operated by the same two organizations. The number of available measurements for comparison is reduced further as epochs of VLBI observations must coincide with those of altimetry within the spatial search radius of 0.5° . Figure 65 shows how VTEC from TOPEX/POSEIDON agree with values derived from VLBI. The left side (figure 65(a)) illustrates discrepancies by a histogram plot, revealing a mean difference of 1.6 TECU and a standard deviation of ± 6.7 TECU. The right side of figure 65 shows VLBI VTEC values, plotted against TOPEX/POSEIDON measurements. The comparison with JASON-1 measurement is shown in figure 66 in the same manner, yielding a mean difference of -0.4 TECU and a standard deviation of ± 11.4 TECU. Since the plasma above the altimeter orbit is not considered in TOPEX/POSEIDON and JASON-1 measurements, the true difference is expected to be slightly higher (1 - 3 TECU) than the values mentioned above. After taking this effect into account, approximately the same bias as for GPS can be found. As discussed in section 5.5.2, VLBI tends to underestimate the ionosphere slightly, which can be seen especially in equatorial regions.

5.7 Comparison to a theoretical model - IRI2001

The two prior subsections showed that the VTEC values obtained from VLBI measurements are in very good agreement with results from other space geodetic techniques. Thus, the outcomes can also

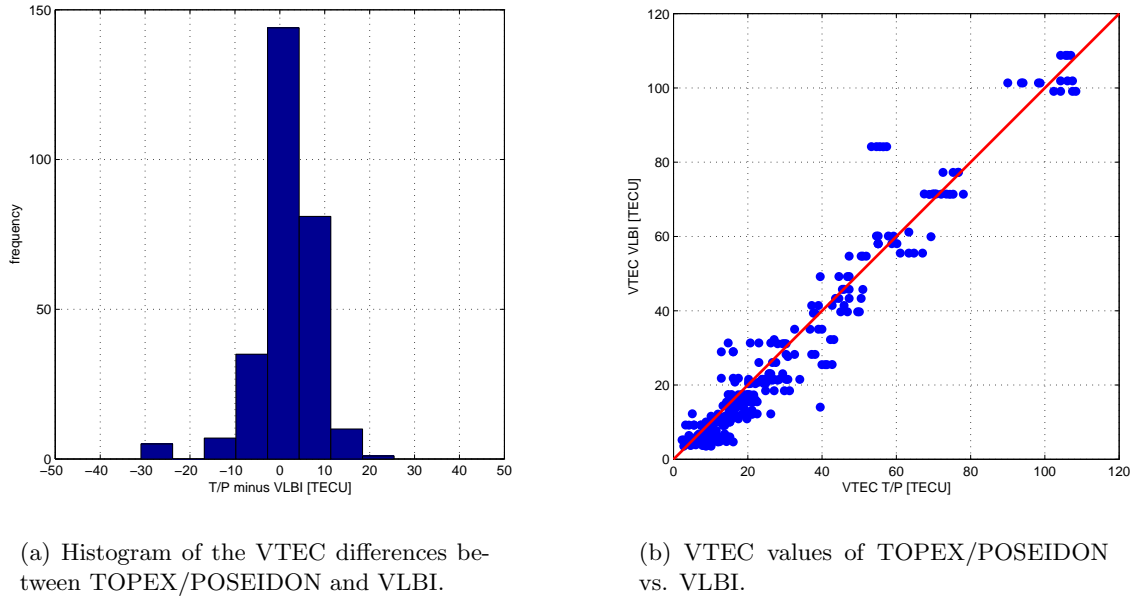


Figure 65: Comparison to TOPEX/POSEIDON VTEC for station KOKEE.

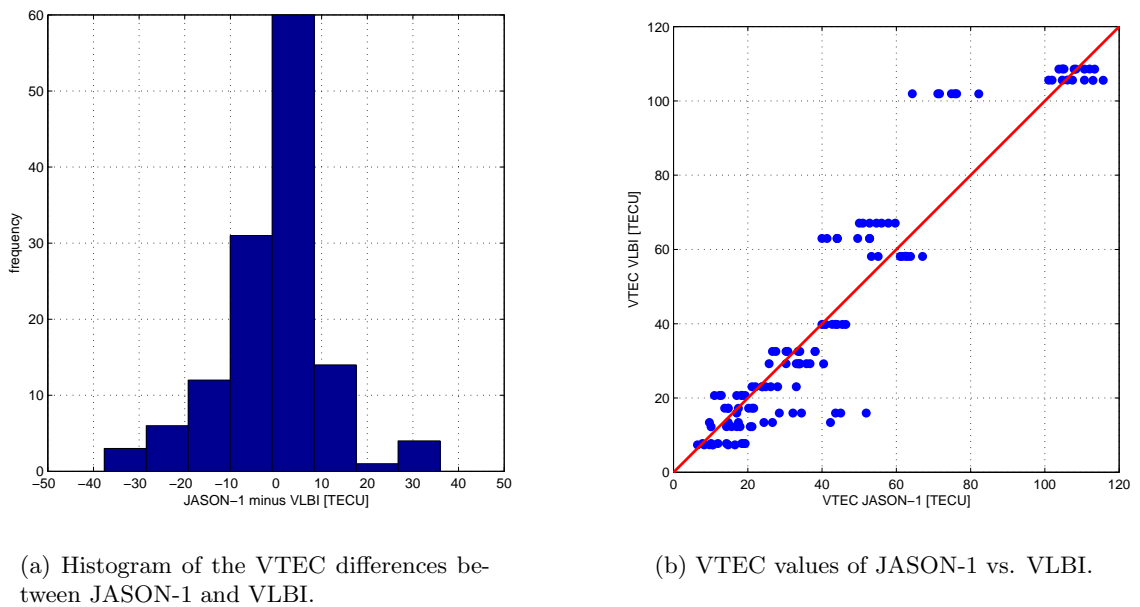
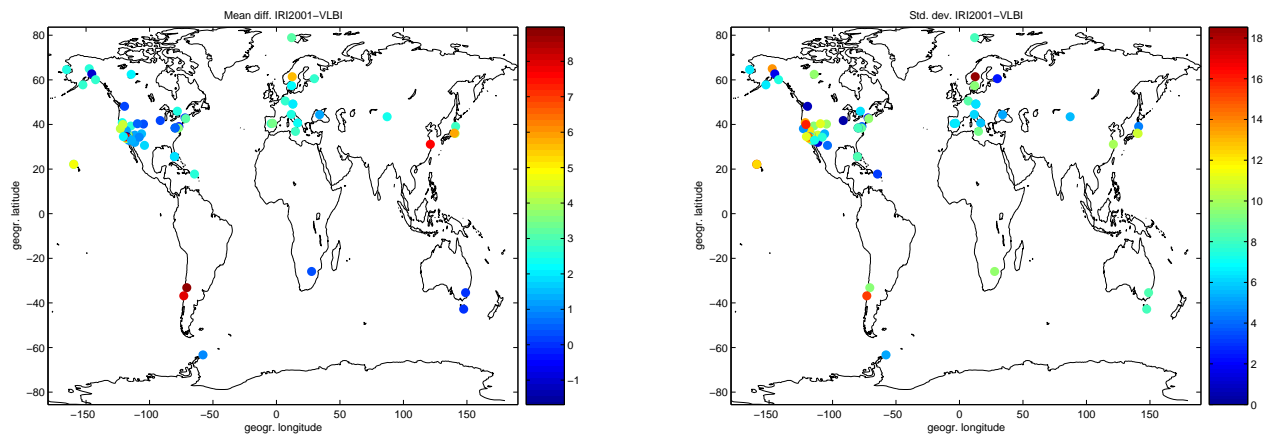


Figure 66: Comparison to JASON-1 VTEC for station KOKEE.

be used for testing the quality of theoretical models of the ionosphere. VLBI offers an opportunity to validate such models on long-term time scales, rather than any other space geodetic technique. For such comparisons, the most common of the theoretical models, IRI, in its actual version IRI2001, was chosen. This model has already been used to obtain time-series of hmF2 in order to set the height of the single layer mapping function properly. VTEC values for each VLBI site were calculated by integration over the theoretical electron density profiles for 30 minutes time-steps. These time-series were used to compare the IRI2001 TEC with the VLBI results. Figure 67 shows the spatial characteristics of the mean difference between IRI2001 and VLBI results in the left part (figure 67(a)). The standard deviation between the theoretical model and VLBI derived VTEC values is illustrated in the right plot (figure 67(b)). Mean differences are within the range of -2 to 8 TECU, whereas it is not possible to conclude on any systematics only from figure 67. The main reason is that the lengths of the time-series for each station are different and mismodeled long-term systematic effects will show



(a) Map of the mean differences between IRI2001 and VLBI.

(b) Map of the standard deviation IRI2001 and VLBI.

Figure 67: Spatial representation of mean differences and standard deviations between IRI2001 and VLBI.

up with different sign and amplitudes in the plots. Furthermore the VLBI station network is not well distributed around the globe, but contains much more stations in the northern hemisphere than in the southern one. Thus, a statement about the quality of global VTEC values from IRI2001 cannot be made. Figure 68 shows a histogram of the differences for all VLBI stations. The mean value is about 3.5 TECU (by means of “IRI2001 minus VLBI”) with a standard deviation of 13.6 TECU and the distribution of the discrepancies is skewed, which is thought to be related to the unbalanced station distribution. The small overall bias of 3.5 TECU however, will be even smaller, if it is taken into

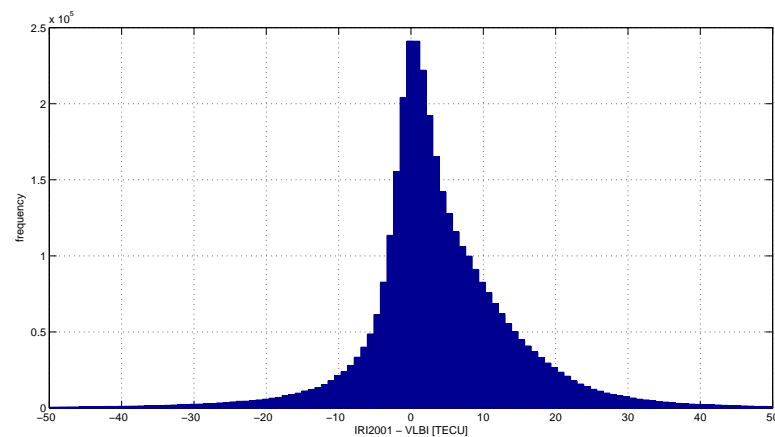


Figure 68: Histogram of the VTEC differences between IRI2001 and VLBI computed for all stations.

account that VLBI is underestimating the total electron content slightly. This means that IRI2001 is able to represent the the main trends of VTEC for globally distributed stations accurately. An example how the theoretical model differs from real measurements is given in figure 69. The upper plot shows the local noon VTEC values at station WETTZELL from the IRI2001 model and from VLBI. In the lower plot the relative differences to the theoretical model (in percent) are plotted against time. The black thick line corresponds to a running average using a window length of 50 days and reveals that some periodic variation seems to be unmodeled in IRI2001. After computing the Wavelet scalogram of the IRI2001 time-series (figure 70), it is clearly noticeable that the semi-annual amplitude is less than those of VLBI or GPS measurements (figure 53). Especially in the solar cycle around the year 1991 the strong semi-annual signal, as it can be seen in figure 53, is not covered by the theoretical model. An improvement of the model in that period domain is suggested to increase the precision in the next

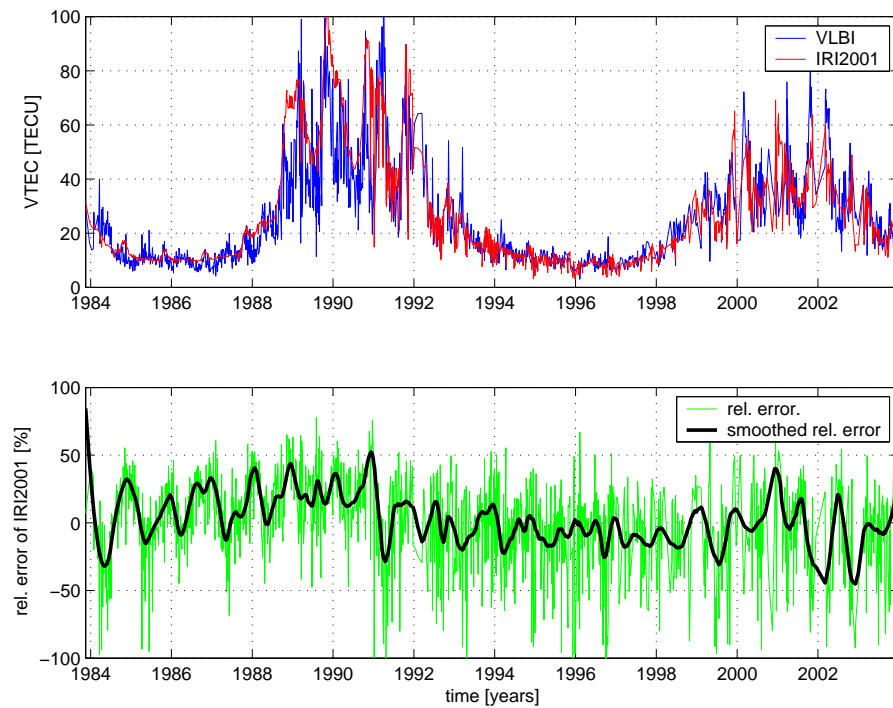


Figure 69: VLBI and IRI2001 VTEC values (at local noon) and the relative error for station WETTZELL.

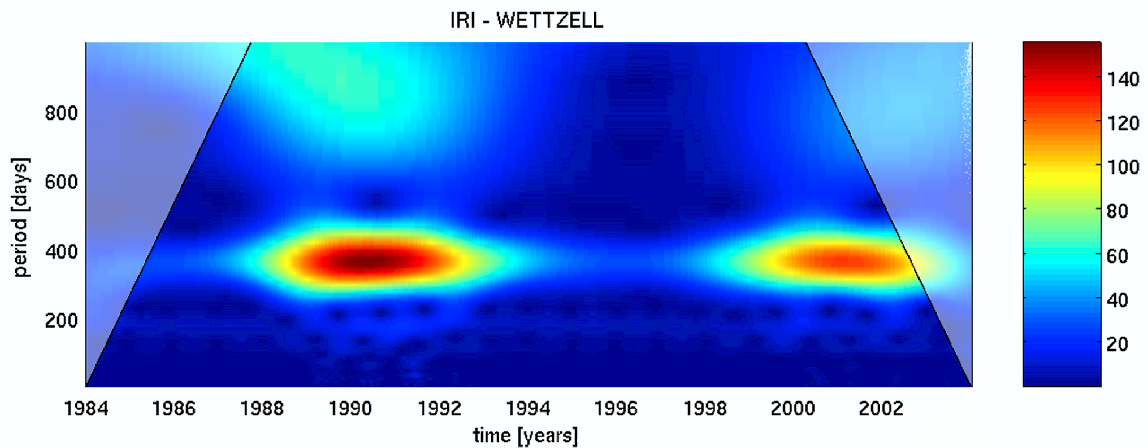


Figure 70: Wavelet spectrum of IRI2001 time-series for WETTZELL.

version of IRI. The diurnal and annual variations seem to be modeled well, instead. It is interesting to see that the overall relative error is not varying by the solar cycle, which can be explained by the fact that the sun spot number is used as an input parameter for the model. Generally, this means that the long-period trends are covered very well by IRI2001 and sub-annual (especially the semi-annual) periodic variations should be updated in a next version of IRI.

5.8 Correlations with solar and geomagnetic indices

Variations of total electron content are expected to be caused by Earth rotation and revolution (daily, semi-annual and annual period), by the geomagnetic field and its variations and by changes of solar characteristics (e.g. the solar cycle). When only 24-hour VLBI experiments are treated, one can compute time-series of daily main TEC and maximum TEC. Following the hypothesis that the total electron content of the ionosphere changes in the solar or geomagnetic environment with a certain time lag, correlation coefficients between two time-series, which are shifted against each other by a

given number of days (or smaller entities), can be computed. The time shift, corresponding to the peak of the cross-correlation function, is expected to represent the response time of the ionosphere to an outer physical excitation. For comparison reasons, the “reduced” daily mean and maximum VTEC values $VTEC_{red.mean}$ and $VTEC_{red.max}$ were computed, with the semi-annual and annual terms being removed. Figure 71 shows how the time-series look like for station WETTZELL, which will be used for the further investigations as it is the longest and most dense one. Time-series of sunspot numbers,

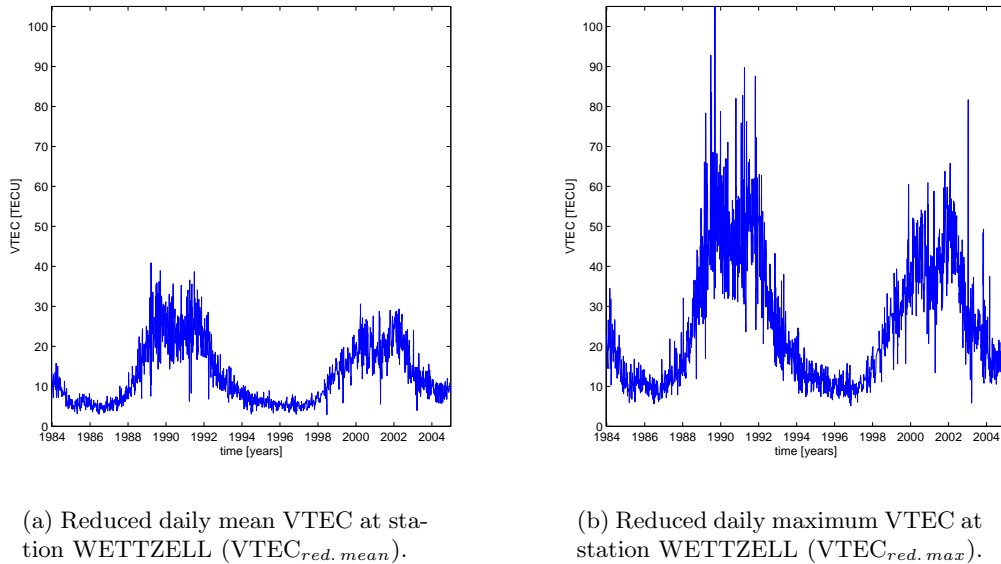


Figure 71: Reduced time-series of mean and maximum VTEC at station WETTZELL

solar flux, Kp and DST indices were obtained from the Space Physics Interactive Data Resource (SPIDR, 2005 [54]) and compared to $VTEC_{red.mean}$ and $VTEC_{red.max}$.

5.8.1 Correlations with sunspot numbers

The sunspot number and its definition was already discussed in section 2.2.1. For the following investigations daily values of the Wolf number R were compared to reduced VTEC time-series from VLBI. The correlation coefficient

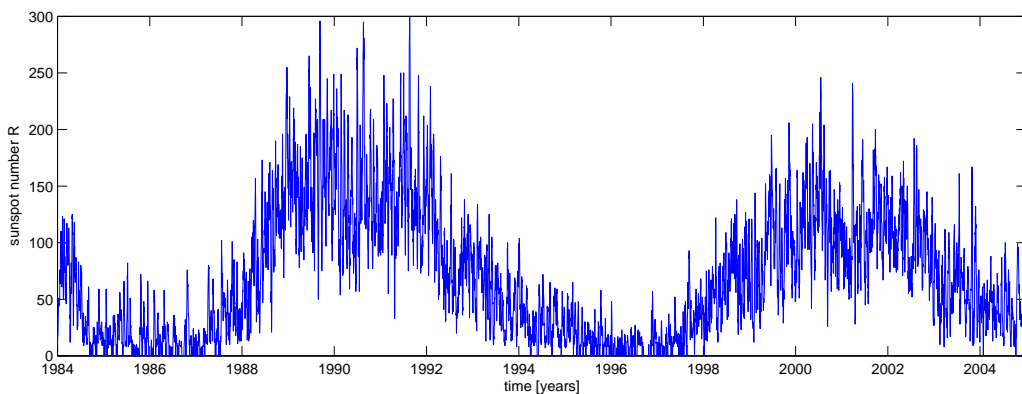


Figure 72: Plot of the daily sunspot numbers from beginning of 1984 until end of 2004.

$$r_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y} \quad (5.56)$$

was computed for time lags between -10 and 10 days, where positive lags mean that VLBI follows solar excitation. Figure 73 gives the results for that comparison. The left plot (figure 73(a)) gives the correlation coefficients between Wolf number and reduced daily mean VTEC. The highest correlation

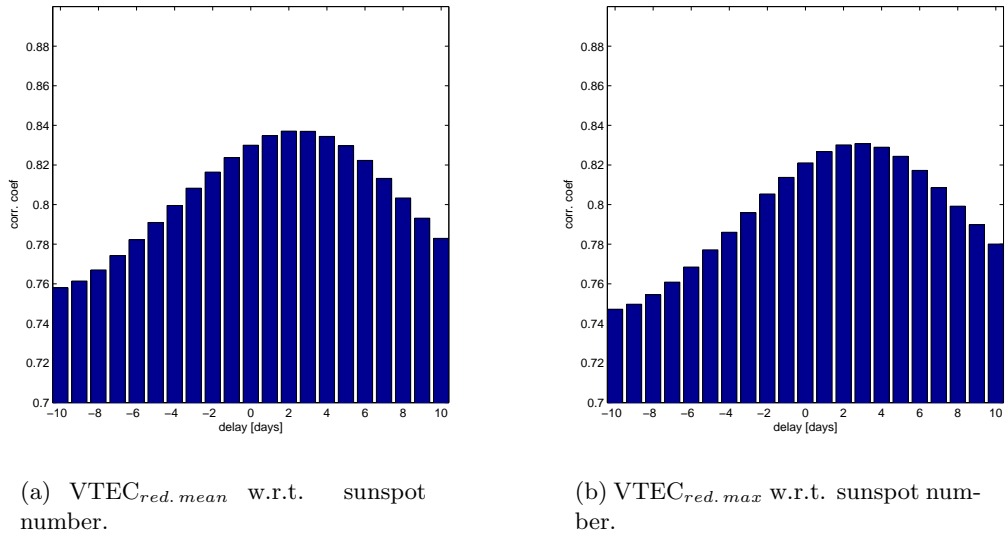


Figure 73: Correlation coefficients for reduced mean and maximum VTEC with respect to sunspot number, for lags between -10 and $+10$ days.

can be found for a time delay of about 2 days. A similar result is obtained when the $VTEC_{red.max}$ values are used for comparison (figure 73(b)). The maximum correlation coefficient is around 0.83 and shows a clear decrease when the lag is set to other values than two or three days.

5.8.2 Correlations with solar flux

Solar flux density at 2800 MHz has been recorded routinely by a radio telescope near Ottawa since February 14, 1947. Every day levels are determined at local noon (1700 GMT) and then corrected within a few percent for factors such as antenna gain, atmospheric absorption, bursts in progress, and background sky temperature. Beginning in June 1991, the solar flux density measurement source is Penticton, B.C., Canada. The data (figure 74) contain fluxes from the entire solar disk at a frequency

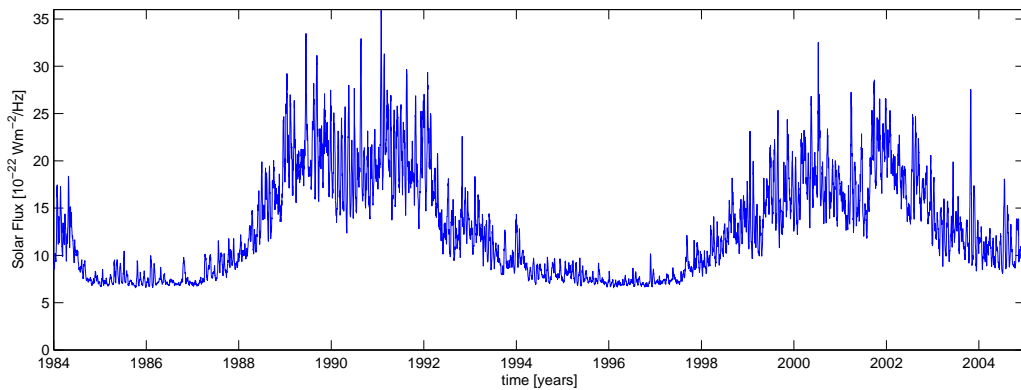


Figure 74: Solar flux at 2800 MHz.

of 2800 MHz in units of 10^{-22} Joule/second/square meter/Hertz. Three sets of fluxes - the observed, the adjusted, and the absolute - are summarized. Of the three, the observed numbers are the least refined, since they contain fluctuations as large as 7% that arise from the changing Sun-Earth distance. In contrast adjusted fluxes have this variation removed; the numbers in these tables equal the energy flux received by a detector located at the mean distance between Sun and Earth. Finally the absolute levels carry the error reduction one step further; here each adjusted value is multiplied by 0.90 to compensate for uncertainties in antenna gain and in waves reflected from the ground. Like before a time window of ± 10 days was set for correlation analysis. Figure 75 shows how the dataset correlates

with reduced VLBI time-series when a certain time-lag is introduced. Like in section 5.8.1, maximum

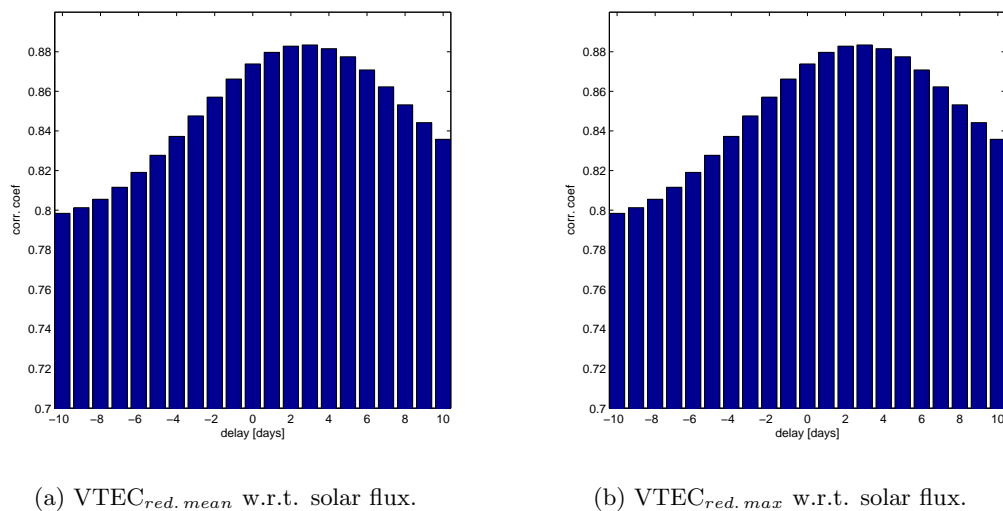


Figure 75: Correlation coefficients for reduced mean and maximum VTEC with respect to solar flux, for lags between -10 and $+10$ days.

correlation can be found after two days w.r.t. solar flux and a similar fall-off around that peak can be seen. But correlation significantly reaches higher values ($r_{xy} = 0.88$) than sunspot numbers. Schaer (1999, [61]) found similar results using GPS measurements. He explained higher correlations for solar flux measurements by the fact that sunspot numbers only provide an absolute measure in terms of solar activity. Generally it can be stated that a significant time-lag of two to three days can be verified, which is thought to be the reaction time of the ionosphere on solar excitations. Since the mean coronal mass ejection (CME) speed is about 500-600 km/s (Lara et al., 2003, [48]), one can compute the arrival time at the Earth, using an average distance to the Sun of $149 \cdot 10^6$ km, by 2.9-3.4 days. Therefore the VLBI results show a clear evidence that perturbations seen in VTEC are caused by CMEs.

5.8.3 Correlations with Kp and Dst

Kp indices isolate solar particle effects on the Earth's magnetic field. Over a 3-hour period they classify the range of variation of the more unsettled horizontal field component into disturbance levels. Each activity level relates almost logarithmically to its corresponding disturbance amplitude. Interpolating the VLBI data to the time intervals of the Kp index, removing periods longer than 7 days from the data and computing the correlation coefficient for time lags between ± 24 hours did not reveal any significant relationship ($r_{xy} < 0.1$). As a second geomagnetic index, the Dst (Disturbance Storm Time) equivalent equatorial magnetic disturbance indices were investigated on their impact on VLBI VTEC values. They are derived from hourly scalings of low-latitude horizontal magnetic variation and show the effect of the globally symmetrical westward flowing high altitude equatorial ring current, which causes the "main phase" depression worldwide in the H-component field during large magnetic storms. But as before, when comparing to the Kp index, no significant correlations for time-lags between ± 24 hours were detectable ($r_{xy} < 0.1$). In summary it can be ascertained that no correlations between long time-series of vertical total electron content measurement from VLBI and geomagnetic indices could be detected.

5.9 Station-dependent instrumental offsets

Besides parameters describing vertical total electron content and its time-variation, also station-dependent instrumental offsets $\tau_{inst,i}$ are obtained from the adjustment process. These instrumental delays are thought to be constant within one session and special parameterization is only necessary if jumps due to system changes (e.g. power outage) occur. Figure 76 shows how these parameters evolve

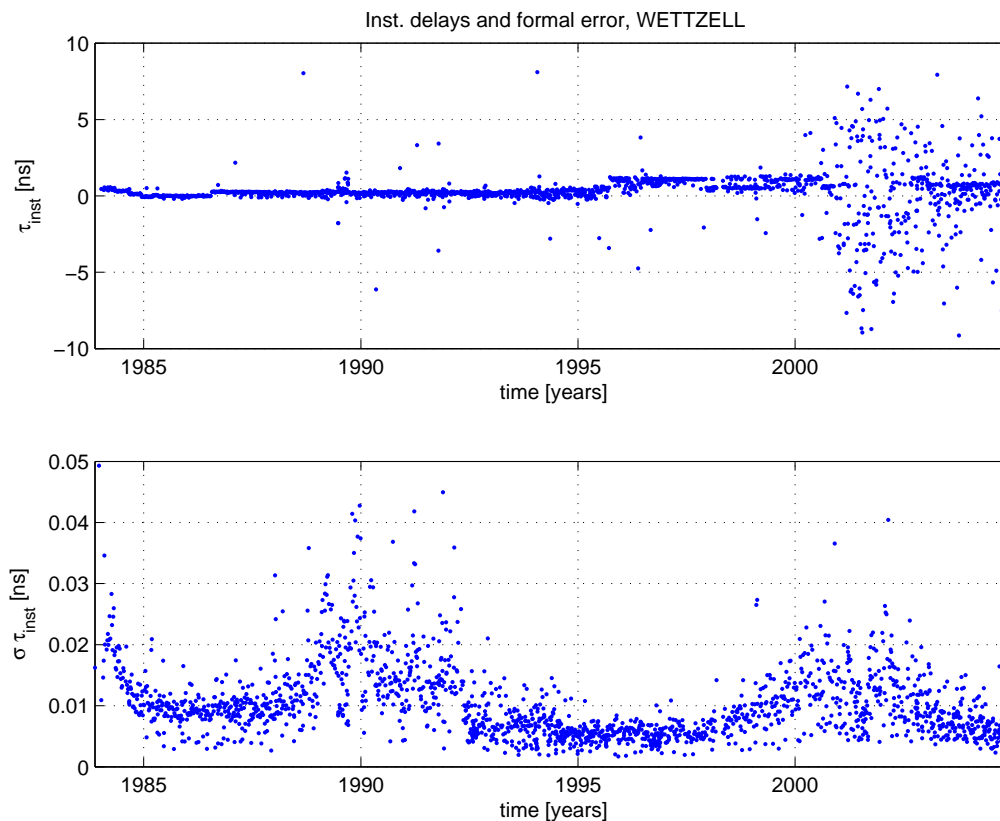


Figure 76: Instrumental offsets and their formal error at station WETTZELL.

with time for station WETTZELL, Germany. The upper plot shows the obtained values $\tau_{inst,i}$ in nanoseconds, the lower part gives the corresponding formal error of the estimates. It can be seen that up to the end of the year 2000 this instrumental bias is constant between the sessions and only suffers small changes, which can be explained by modifications of the receiving system at the observation stations. The scattering, which starts after that date, cannot be assigned to ionospheric influences on the VLBI measurements due to an increase of total electron content during solar maximum. If so, such an effect would be noticeable around the first solar cycle, too. But this cannot be found in the results. In fact there are several other origins, which could be considered for that effect. First of all a change of the VLBI system took place as the old Mark3(A) system was replaced by the new Mark 4 type and a new correlator architecture was introduced (Whitney et al., 2004 [80]). Several reported bugs⁴ in the initial phase of the new system could be the possible origin of the phenomena seen in figure 76. Approximately at the same time, several changes to the ambiguity fixing algorithm of CALC/SOLVE were made (Petrov L., personal communication, 2005). Also the way how effective ionospheric frequencies (section 4.2.1) were computed was changed around that time. Another station, WESTFORD, is plotted in figure 77. This station shows also scattering in the last part of the time-series, but starting already in the year 1999. Therefore some of the prior explanations do not hold for effects noticeable at this station. Additionally the time-series of station-dependent instrumental offsets at WESTFORD shows an interesting behavior in the first years of operation. The obtained values of $\tau_{inst,i}$ seem to be quantized at three discrete levels, which could be an effect of ambiguity fixing within post-correlation analysis steps. Generally it can be said that, despite of the fact that instrumental delays change their value from session to session, geodetic products are not effected in any manner, as these values are stable within sessions and $\tau_{inst,i}$ is absorbed into the clock-model of the analysis. The formal error of the obtained parameters (lower plots of figures 76 and 77) follows the total electron content strongly and therefore shows a similar pattern as it is found for daily mean values of VTEC (see figures 42 to 45). The instrumental delays are treated in a similar way as for

⁴<ftp://web.haystack.edu/pub/mark4/memos/memoindex.html>

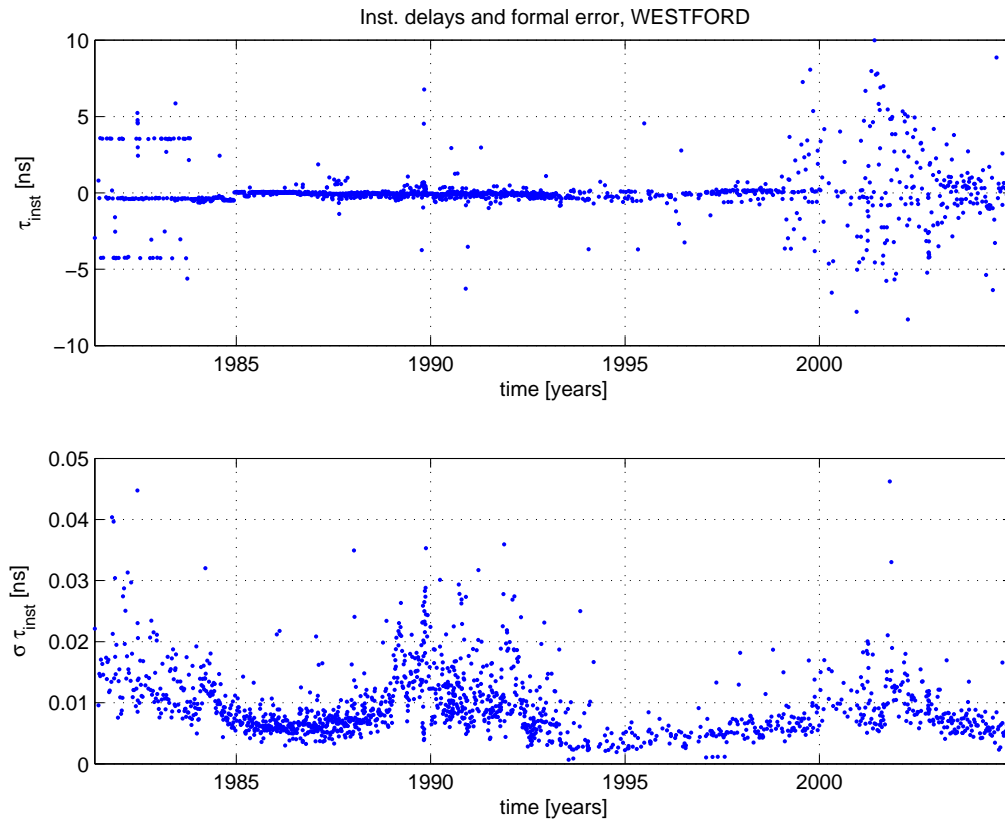


Figure 77: Instrumental offsets and their formal error at station WESTFORD.

receiver differential code biases (DCBs) in the GPS case. Comparing the formal errors of receiver DCBs⁵ and those of the VLBI solution, it can be seen that the GPS estimates have much higher (a factor of 2 – 10) standard deviations. This is interesting as the number of observations in the VLBI case is much smaller and, under the assumptions that both measurement types are of equal precision, the formal error is expected to be higher than for GPS.

⁵http://www.aiub.unibe.ch/ionosphere/p1p2_all.dcb

6 Determination of ionospheric parameters from intra-scan phases

After correlation of the VLBI signals a single-band delay search process provides fringe phases and amplitudes for parameter periods (PP) for each channel. In geodetic VLBI several channels are combined by bandwidth synthesis to obtain a precise estimate of delay and delay rate (section 3.3.3). The fringe phase $\phi_{i,j}$ of channel i for parameter period j can be split up into

- a residual delay $\Delta\tau_j$, i.e. the difference between the theoretical delay, computed from a polynomial of higher order, and the measured delay from single-band delay search,
- an ionospheric contribution $\tau_{ion,j}$,
- a phase offset $\hat{\phi}_i$, which is aligned within bandwidth synthesis by applying the phase-cal signals and
- a random error ϵ_{ij} caused by the electronics of the receiving system.

Assuming that all influences can be described by continuous processes, one can denote the fringe phase variations of channel i by

$$\phi_i(t) = 2\pi f_i \Delta\tau(t) - 2\pi f_i \tau_{ion}(t) + \hat{\phi}_i + \epsilon \quad (6.1)$$

where f_i is the RF frequency of the i -th channel and ϵ represents random noise. As described in the prior chapters, the ionospheric propagation influence can be approximated by the total electron content along the ray path. Thus (compare with equation (4.15)),

$$\tau_{ion}(t) = \frac{40.28}{cf^2} [STEC_A(t) - STEC_B(t)] = \frac{40.28}{cf_i^2} \Delta STEC(t) \quad (6.2)$$

the difference of the ionospheric conditions between station A and B determines the ionospheric delay. Therefore equation (6.1) can be re-written as

$$\phi_i(t) = 2\pi f_i \Delta\tau(t) - 2\pi \frac{40.28}{cf_i} \Delta STEC(t) + \hat{\phi}_i + \epsilon. \quad (6.3)$$

Expression (6.3) reveals that all time-variable influences on intra-scan phases are either proportional (residual delay) or inverse proportional (total electron content) to the frequency of the concerned channel. Pure phase variations, independent of channel frequency, are thought to be neglectable as they are mainly caused by frequency drifts of the oscillator, which is assumed to be stable within one scan. For longer scans changes in the room temperature might evoke a drift of the frequency standard, which leads to a phase variation within the channel. Using single channel data only it is impossible to separate ionospheric contributions from delay changes. But, if fringe phases can be computed for individual channels, it will become possible to distinguish the effects as they are scaled by the corresponding channel frequency.

6.1 Functional and stochastic model

Short period ionospheric variations are expected not to happen suddenly, but are thought to be phenomena which develop steadily or show continuous changes of total electron content. The corresponding excitations can be manifold and cover geomagnetic disturbances, solar variations, and gravity waves, which are caused by earthquakes, tsunamis (both Ducic et al., 2003, [19]) or rocket launches (Afraimovich et al., 2000, [1]). Therefore any descriptive model of $\Delta STEC$ should be a continuous, smooth and easy to implement function. Following Hu and Sheu (2000, [35]) a quadratic B-spline curve is taken for fitting not only variations of $\Delta STEC$, but also for the delay changes $\Delta\tau(t)$ using the (residual) fringe phases from all channels. The quadratic B-spline is treated like a Kernel function

$\Psi(s)$, defined by

$$\Psi(s) = \begin{cases} \frac{s^2}{2} & 0 \leq s < 1 \\ \frac{3}{4} - (s - \frac{3}{2})^2 & 1 \leq s < 2 \\ \frac{(s-3)^2}{2} & 2 \leq s < 3 \\ 0 & \text{otherwise} \end{cases} . \quad (6.4)$$

Thus, the model function can be described by

$$f(t) = \sum_{j=0}^N S_j \cdot \Psi(t - t_j) \quad (6.5)$$

where S_j is the unknown scale for the j -th Kernel function and t_j is the corresponding translation parameter. As equation (6.5) is linear in the parameter S_j , a least-squares adjustment following a Gauss-Markov model can be carried out to estimate the unknowns. As seen in expression (6.4), the nodes should not differ less than 1 in order to gain overlapping areas of $\Psi(t - t_j)$. This can be assured by rescaling the t axis first, so that N equally spaced intervals cover the whole time-span. Allowing the same number of Kernel functions for delay changes and for ionospheric variations, the functional model for the i -th channel is given by

$$\phi_i(t) = 2\pi f_i \sum_{j=0}^N A_i \cdot \Psi(t - t_j) - 2\pi \frac{40.28}{cf_i} \sum_{j=0}^N B_i \cdot \Psi(t - t_j) + \hat{\phi}_i \quad (6.6)$$

where A_i represents the unknown coefficients of the delay change model and B_i those for Δ STEC. The phase offset $\hat{\phi}_i$ can be determined together with the other unknowns in the adjustment process. Fringe phase information of several channels is necessary to separate the effects from each other and to obtain reliable and robust results. The standard deviation of phase measurements is inversely proportional to their signal to noise ratio (SNR) as denoted by equation (3.47). Therefore the weight matrix, used within the adjustment process, is given by

$$P_{nm} = \begin{cases} SNR_{nm}^2 & (n = m) \\ 0 & (n \neq n) \end{cases} . \quad (6.7)$$

Instead of the signal to noise ratio, also the correlation amplitude ρ_0 can be used to set up the stochastic model. Since the signal to noise ratio is given by $SNR = \rho_0 \sqrt{2BT}$ (see equation (3.46)) and the parameter period length T and the bandwidth B is the same within all channels, it is possible to use either signal to noise ratio or correlation amplitude to weight the data points.

6.2 An example for the detection of short-period TEC variations

In this subsection it will be shown how VLBI can detect small variations of total electron content and how good these findings agree with GPS measurements. As an example a scan of experiment SYW029 on Jan. 7th, 2004 on the baseline HARTRAO (South Africa) - HOBART26 (Australia) is taken.

6.2.1 VLBI data

Data were obtained in the KOMB format (Takahasi et al., 1991, [74]), which does not only contain the results of bandwidth synthesis processing (see appendix C) but also fringe phases and SNR values for each parameter period and channel. As phases are only measured between 0 and 2π , it is necessary to unwrap the fringe phases first, i.e. to carry out phase connection between the parameter periods. A TEC variation of about 0.8 TECU is equal to one cycle at S-Band but only one fourth of a cycle

in X-band. As the used KOMB data for this study has a parameter period length of 2 seconds, TEC variations of 0.2 TECU/s are assumed to be detectable when phase unwrapping is done properly. One limitation is that intra-scan phases show a large scattering, i.e. a low SNR, in case the observed sources have low flux densities. Geodetic observables can still be obtained from such sources, when the scan is long enough as SNR scales by the square root of scan length. But fringe phase data show a large scattering and phase unwrapping becomes impossible. The strong source 1954-388 (S-band flux 2.00 Jy, X-band flux 3.15 Jy⁶) was observed on Jan. 7th, 2004 from 8:48:00 UT until 8:54:50 on the baseline HARTRAO-HOBART26 (experiment SYW029). Figure 78 shows the fringe phases for each channel from that scan. Unwrapping the fringe phases of the eight X-band and six S-band

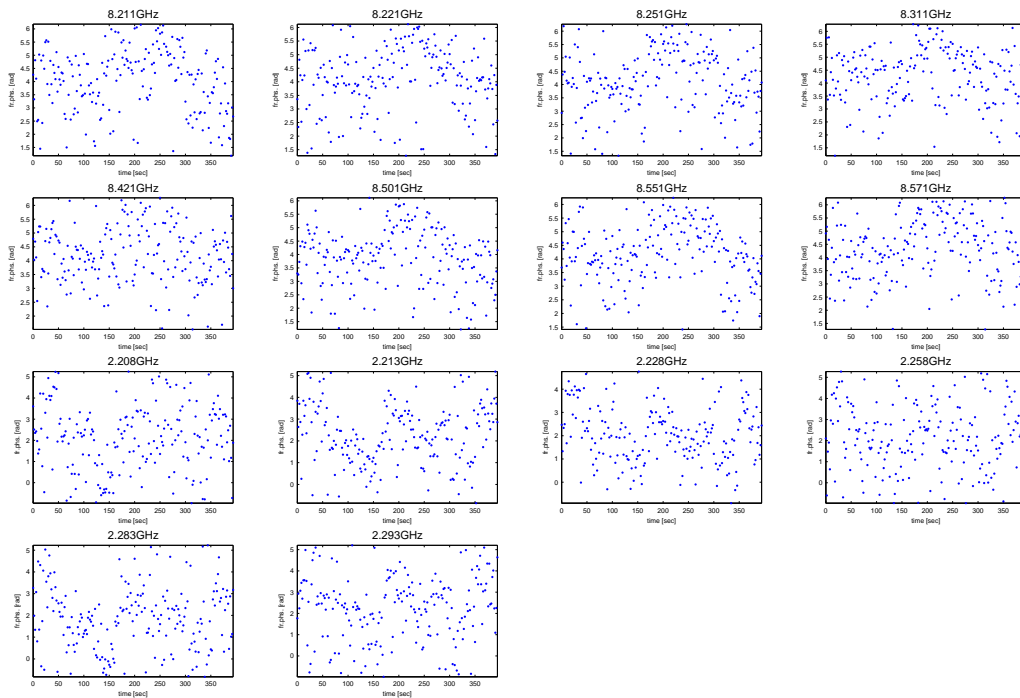


Figure 78: Fringe phases ($0 \dots 2\pi$) on baseline HARTRAO-HOBART26 from experiment SYW029, Jan. 7th, 2004 from 8:48:00 UT until 8:54:50 UT, having a parameter period length of 2 seconds.

channels provides the basis for separation of the different influences by a least-squares adjustment as described in the prior section. Although the variations can be assigned to different physical effects, the parameters are correlated among themselves as shown in figure 79. The first columns contain the parameters assigned to Δ STEC, followed by parameters depending on $\Delta\tau$ and the phase offsets in X- and S-band. It can be seen that ionospheric variations can be de-correlated from delay changes, but each of them is strongly correlated with the phase offsets. Whereas the Δ STEC is strongly correlated with S-band phase offsets the delay variations have a high correlation with X-band phase offsets. This can be explained by the physical nature of the effects. As ionospheric influences scale inversely proportionally to the frequency, lower frequencies are affected more than higher ones. Delay changes scale directly proportional to the frequency and therefore influence more X-band than S-band. Finally, it can be stated that X-band (and S-band) phase offsets are highly correlated with each other but show no correlation with S-band (X-band) phase offsets. Figure 80 displays the estimated variation of differential total electron content of the scan in TECU. The dashed lines represent the formal error of the adjusted curve at 1σ level. The fitted curve has a maximum of 0.20 TECU and a minimum of -0.45 TECU and reveals a periodic variation with a period of about three minutes. The estimated remaining delay variations are shown in figure 81. One can see that a maximum delay change of about 25 picoseconds occurs, which is equal to about 1.3 radians at X-band frequencies. Computing the correlation coefficient between Δ STEC variations and delay changes gives a value of -0.43 , which shows that both models describe the physical effects properly without influencing each other. In order

⁶The Australia Telescope National Facility (ATNF), Revised Calibrator Catalogue, <http://www.narrabri.atnf.csiro.au/calibrators/c007/atcat.html>

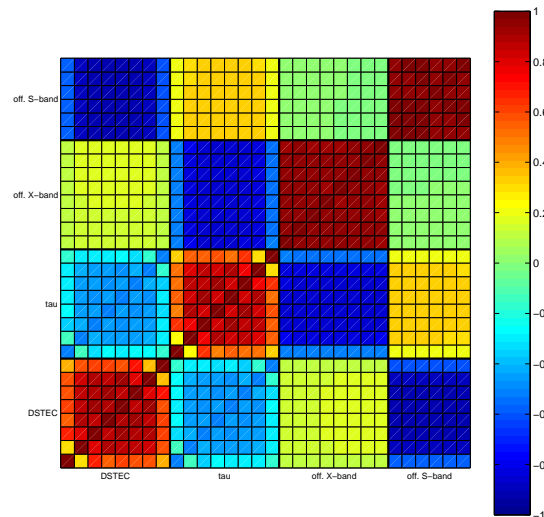


Figure 79: Correlation matrix of the parameters.

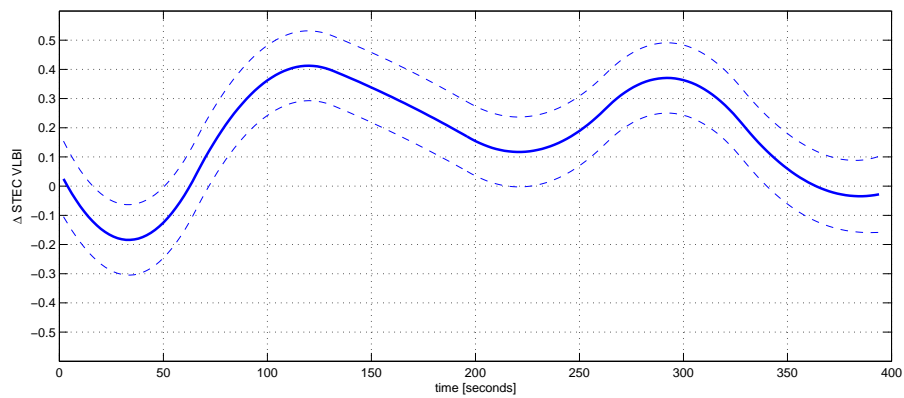


Figure 80: Intra-scan variation of ΔSTEC in TECU on Jan. 7th, 2004 from 8:48:00 UT until 8:54:50 on the baseline HARTRAO-HOBART26.

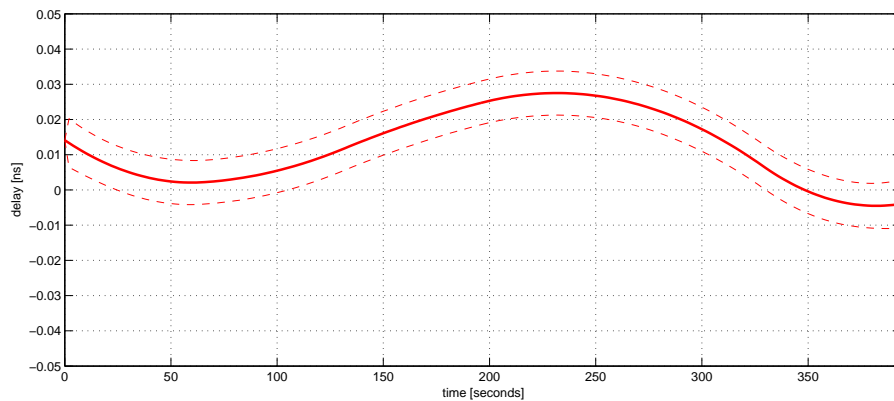


Figure 81: Intra-scan variation of τ in nanoseconds on Jan. 7th, 2004 from 8:48:00 UT until 8:54:50 on the baseline HARTRAO-HOBART26.

to prove that the effects are of real physical nature and not just artifacts caused by the receiving system, it will be necessary to find other techniques which could reveal the same effect. Therefore GPS data were downloaded and analyzed with respect to TEC variations.

6.2.2 GPS data

GPS technology provides a tool for ionospheric wave disturbance detection based on phase measurements of slant TEC (Afraimovich et al., 2003, [2]). The relation between STEC and phase measurements carried out on L1 and L2 can be found e.g. in Hoffmann-Wellenhof (2001, [34]).

$$STEC = \frac{1}{40.308} \frac{f_1^2 f_2^2}{f_1^2 - f_2^2} [(L_1 \lambda_1 - L_2 \lambda_2) + const. + nL] \quad (6.8)$$

The additional paths of the radio signal (in meter) caused by the ionosphere are expressed by $L_1 \lambda_1$ and $L_2 \lambda_2$, where L_1 and L_2 represent the number of phase rotations at frequencies f_1 and f_2 , λ_1 and λ_2 stand for the corresponding wavelengths in meter. The constant is caused by the unknown initial phase path and nL symbolizes errors of the phase paths. When the basic GPS frequency $f_0 = 10.23$ MHz is used f_1 and f_2 can be expressed by (Xu, 2003 [83])

$$f_1 = 154 \cdot f_0 \quad f_2 = 120 \cdot f_0. \quad (6.9)$$

The corresponding wavelengths can be computed, using the value of the speed of light, which relates frequency to wavelength by $c = \lambda f$. The data are usually stored in intervals of 30 seconds and provided by Receiver Independent Exchange Format (RINEX) files to a broad user community. In the last years a lot of stations have been equipped with 1 Hz receivers, which makes it principally possible to detect short period disturbances of the ionosphere. Special GPS receivers, dedicated to ionospheric research, have the capability to record data with up to several Hz (Shilo et al., 2000, [68]). Phase measurements can be made very accurately and errors of the determined TEC reach values of 0.01 TECU, when special receivers are used. But even for geodetic GPS receivers the error is expected not to exceed 0.1 TECU. In order to handle GPS observations based on the RINEX the open source GPSTK software package (Tolman et al., 2004, [78]) was used. Fortunately, the VLBI stations HARTRAO and HOBART26 have nearby GPS receivers which can be used for comparison reasons. RINEX data at these stations are recorded in time steps of 30 seconds, so that 13 data-points correspond in time with the VLBI scan. In order to compare GPS STEC values with VLBI measurements one has to find the satellites at each station, which are nearby the pointing direction of the radio telescope. Thus, sky-plots for stations HARTRAO and HOBART26 were generated. Figure 82 shows the tracked GPS satellites and the radio source at each site. At station HARTRAO the

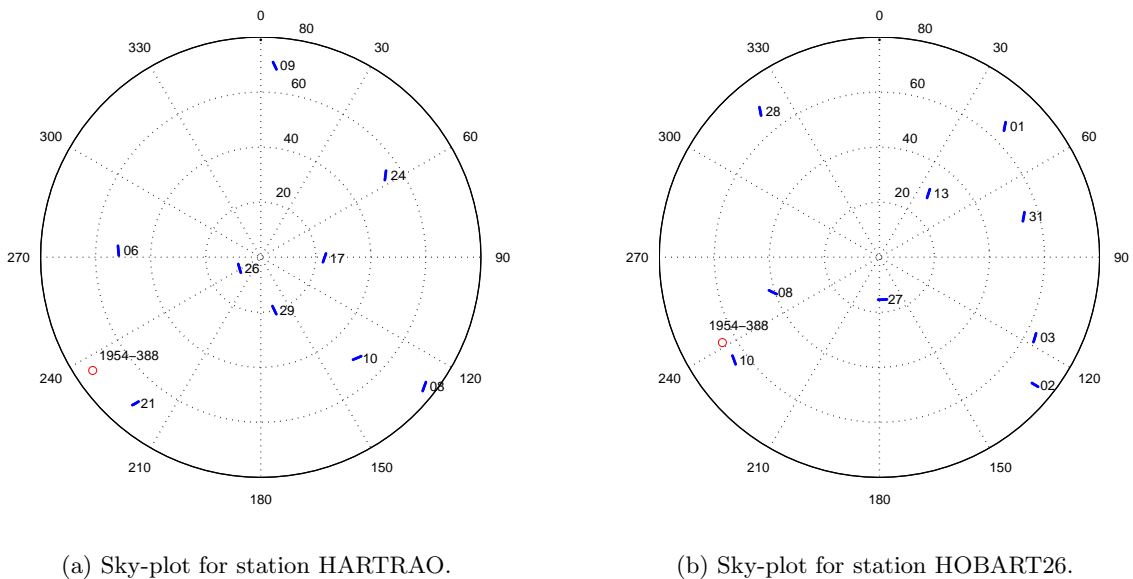


Figure 82: Sky-plot at stations HARTRAO and HOBART26 showing the position of the tracked GPS satellites during the VLBI scan.

space vehicle (SV) 21 GPS satellite is closest to the pointing direction of the antenna (figure 82(a))

and at HOBART26 the SV 10 satellite fulfills this spatial criteria (figure 82(b)). It is expected that the ionospheric variations detected by VLBI can be found in the GPS data, too, when the difference between the slant TEC values and the selected satellites is computed. The de-trended STEC values, computed according to equation (6.8) for each station are plotted in figure 83. One can see that STEC

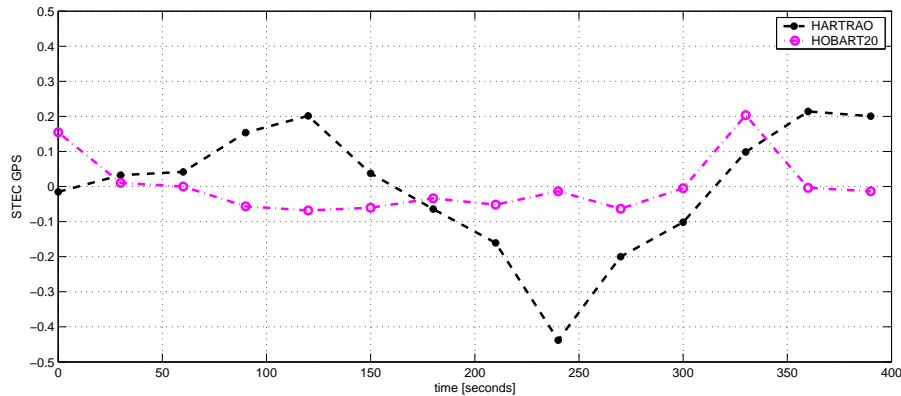


Figure 83: De-trended STEC to GPS satellite 21 from station HARTRAO and to GPS satellite 10 from station HOBART26.

at HARTRAO (black dashed line in figure 83) varies with the same range as detected by VLBI (figure 80) and that there are only very small TEC variations at station HOBART26 (red dash-dotted line in figure 83). Therefore the ionospheric variation measured by VLBI is mainly assigned to the South African station. Computing the difference between both GPS measurements should reproduce the VLBI measurement up to a certain accuracy. Figure 84 shows how the differenced STEC looks like. Unfortunately, high data-rate, i.e. 1 Hz sampling interval observations, are not available for station

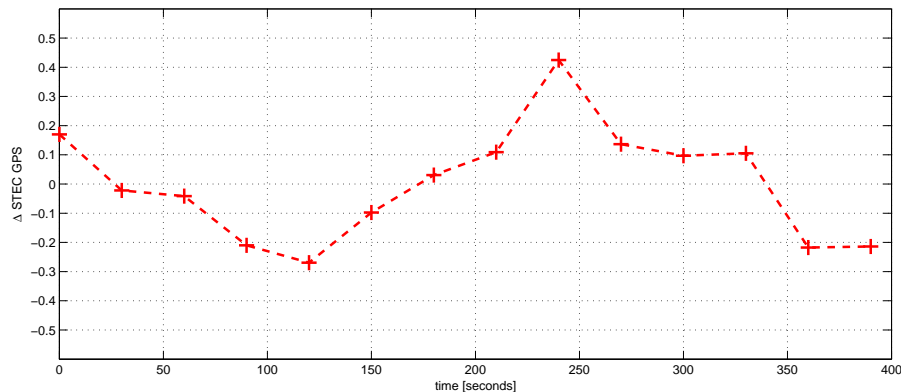


Figure 84: Δ STEC between HARTRAO and HOBART26 as measured from GPS.

HARTRAO and one has to keep in mind that the obtained curve has a precision of about ± 0.1 TECU. Comparing figure 84 to 80 leads to the following conclusions:

- Δ STEC computed from VLBI and GPS agree well in amplitude, having a maximum value of 0.4 TECU and a minimum value of -0.2 TECU.
- The results are time-shifted against each other. This can be explained by the fact that the GPS satellite at station HARTRAO is separated several degrees from the pointing direction of the radio telescope. One can imagine that an ionospheric disturbance traveling perpendicular to the pointing direction, i.e. from NW to SE, will affect the ionosphere around the VLBI ray path first.
- The B-spline function seems to be an appropriate representation of ionospheric variations and it is possible to de-correlate ionospheric variations from other effects reliably.

6.3 Possible fields of application and requisites

In the prior sub-section it has been shown that VLBI is able to detect even smallest variations of the ionosphere with a high time resolution as the sampling rate of fringe phases is equal to the parameter period length (here two seconds). As discussed in the prior paragraphs, strong sources are necessary to ensure high signal to noise ratio in order to enable phase connection and/or phase unwrapping. If this is assured, it should be principally possible to detect all TEC variations of the ionosphere, if the scan is long enough. Since VLBI is a differential technique, time-dependent variations of ΔSTEC cannot be clearly assigned to a single station without external information (like GPS in the example before) or other measurements. E.g. if two baselines connect one station, one can compute the ionospheric variations from both measurements and check whether the same signals can be found. VLBI has the big advantage that fringe phase information is very precise and that the formal error of each measurement can be derived easily from the corresponding correlation amplitude. Therefore one can think of the following application fields:

- Detection of traveling ionospheric disturbances (TIDs).
- Detection of plasma bubbles (Sitnov et al., 2005, [69]).
- Monitoring diurnal TEC variations with high precision by special experiments.
- Detection of ionospheric disturbances caused by earthquakes, tsunamis or rocket launches.
- Verifying the different reactions of the ionosphere to solar excitations and geomagnetic disturbances.

Besides ionospheric tasks the proposed method of this chapter provides also a basis for handling intra-scan delay variations and providing phase delay information for geodetic/astrometric purposes or for the navigation of spacecrafts.

7 Summary and outlook

7.1 Summary

In this thesis a thorough investigation of ionospheric influences on VLBI measurements was made. The origin and the magnitude of the influences were discussed and it was proved that the first order approximation taking only total electron content into account is sufficiently precise to describe ionospheric delays of geodetic VLBI measurements. The fact that such corrections are available for all geodetic dual frequency experiments leads to the concept of solving for ionospheric parameters, i.e. values of vertical total electron content. Since VLBI is only sensitive to differences of slant TEC it seems at first glance that it is impossible to solve for absolute values. But the usage of a station-dependent model, a mapping function and taking height variations of the maximum electron density in account allows computation of vertical TEC values at each station, as shown in this work. Globally distributed stations and an advanced weighting strategy make it possible to separate instrumental biases from VTEC values. Comparing the obtained values of vertical total electron content to GPS results gives an overall mean difference of only 1.0 TECU and a standard deviation of 7.6 TECU. It was discussed that the small bias between the two techniques is caused by the height of the single layer mapping function. Comparison to satellite altimetry data revealed approximately the same difference, when the unaccounted plasma above the satellite orbit is taken into account. The long term precision of a theoretical model of the ionosphere, IRI, was validated with the VLBI results and differences having periodic behavior were revealed. This knowledge can be used to update theoretical models, like IRI. As VLBI measurements cover the last two solar cycles, valuable information about VTEC trends at globally distributed stations can help to improve theoretical models or can be used for validation purposes. GPS measurements are already included in such models, and VLBI is a promising candidate as another space geodetic technique which can be incorporated into theoretical models of the ionosphere, too. Since the ionosphere is mainly driven by the activity of the Sun and by the geomagnetic field, correlations between indices, representing the physical properties of these excitations, and VTEC values from VLBI were computed. It was found out that the ionosphere takes about two days to react on variations of the Sun. Significant correlations to variations of geomagnetic conditions could not be verified as the time-series are not sampled properly for this purpose. A spectral analysis of the VTEC time-series showed that the same periods as expected from theory, can be found in the data, when the spectra are computed by a suitable algorithm, which is able to treat irregularly sampled data. The detected period of the solar cycle agrees very well to the period which was found in sunspot data of the last two solar cycles and gives a value of 10.5 years. By using Wavelet analysis, it was shown how the solar cycle modulates the amplitude of the annual and semi-annual signal. Finally short period variations of the ionosphere with periods of several 10s of seconds were investigated. Since fringe phase information can be used to separate ionospheric impact from delay changes, a new method was presented how intra-scan phase variations can be treated and how ionospheric information can be gained from such data. Significant changes of ΔSTEC were detected and verified by the help of GPS data. Thus, VLBI can contribute to ionospheric research when long period variations (more than one month), diurnal signals, or short period variations (period lengths of a few 10s of seconds) are investigated. As long as VLBI does not observe continuously it will not be possible to contribute to other period domains significantly. Therefore VLBI can be used as a tool to study the ionosphere, but its preferable use w.r.t ionospheric research will be to contribute to long-term studies or to validate theoretical or measurement-based local, regional and global ionospheric models rather than providing results for routine monitoring of the ionosphere.

7.2 Outlook

At present the combined IGS global ionospheric model is derived from the results of five different GPS analysis centers and is cross-validated against Jason and/or Envisat measurements on a routine basis, which shows an absolute bias of less than 1 TECU in most of the cases and a standard deviation of about ± 5 TECU. One can imagine that VLBI can be taken as a third independent technique for validation of IGS GIMs. As VLBI is the only technique which is covering more than two solar cycles, it is able to provide important input for theoretical models of the ionosphere and for modeling

long-term trends. Therefore it is intended to collaborate with specialists for theoretical models of the ionosphere in order to update their models with VLBI data. The final goal of ionospheric modeling using space geodesy should be the development of a global 4-dimensional model of the ionosphere, assimilating several independent space techniques, obtaining the most robust and reliable solution of electron densities values. How VLBI can contribute to such a model depends not only on data quality and the number of observations but also on the ability to process the observations in near real-time and to provide the measurements and/or ionospheric parameters as soon as possible to combination centers. Furthermore, fringe phase information seems to be a promising data source to model short period variations on intra-scan base. Geodetic VLBI experiments are most suitable for this purpose, since data is recorded at two distinct bands split up to several channels. In order to detect a significant variation of intra-scan phases it is necessary that fringe phases from correlator output can be connected within one scan without ambiguities. If strong sources are observed, this prerequisite can be fulfilled. At the moment a drawback is caused by the fact that geodetic experiments are scheduled in a way, that observations to strong sources have assigned short observation time, since sufficient SNR (for obtaining the geodetic observables) will be reached after a few 10s of seconds. Experiments dedicated to the detection of ionospheric scintillations, using geodetic VLBI technique, are a desirable goal for the near future. Additionally, the presented method does not only detect ionospheric scintillation effects, but might also have a huge impact on the analysis of geodetic VLBI sessions. A paradigm shift from bandwidth synthesis processing to parameter period data analysis would provide many new opportunities and is expected to improve the accuracy of the geodetic observables. Generally speaking, phase delay should be the upcoming observable for geodetic analysis in order to be competitive with other modern space geodetic techniques. First results of phase connection for navigation of space-crafts (Sekido et al., 2004, [66]) have already revealed the potential of this promising approach.

A Solving the quadratic programming problem

The quadratic programming problem (QP) in its general form is given by

$$\begin{aligned} \underset{\mathbf{d} \in \mathbb{R}^n}{\text{minimize}} \quad & q(\mathbf{d}) = \frac{1}{2} \mathbf{d}^T \mathbf{H} \mathbf{d} + \mathbf{c}^T \mathbf{d} \\ & \mathbf{A}_i \mathbf{d} = \mathbf{b}_i \quad i = 1, \dots, m_e \\ & \mathbf{A}_i \mathbf{d} \leq \mathbf{b}_i \quad i = m_e + 1, \dots, m \end{aligned} \quad (\text{A.1})$$

where \mathbf{A}_i refers to the i th row of the m -by- n matrix A . The described method is an active set strategy (also known as projection method) and following papers of Gill et al. (1984, [25], and 1991, [27]). Principally, the solution procedure involves two phases. In the first phase a feasible point is calculated (if one exists) and in the second phase an iterative sequence of feasible points is launched and it is expected that these converge to the solution. In this method an active set, $\bar{\mathbf{A}}_k$, is maintained which is an estimate of the active constraints (i.e. given in equation (A.1)) at the solution point. Virtually all QP algorithms are active set methods, which has to be emphasized since there exist many different methods that are very similar in their structure but are described in different terms. At each iteration k the matrix $\bar{\mathbf{A}}_k$ is updated, forming a search direction $\hat{\mathbf{d}}_k$, but the equality constraints always remain in the active set $\bar{\mathbf{A}}_k$. The search direction $\hat{\mathbf{d}}_k$ is calculated and minimizes the objective function while remaining on any active constraint boundaries. The subspace for $\hat{\mathbf{d}}_k$ is calculated from a basis \mathbf{Z}_k whose columns are orthogonal to the estimate of the active set $\bar{\mathbf{A}}_k$ (i.e. $\bar{\mathbf{A}}_k \mathbf{Z}_k = 0$). Therefore the search direction, which is a linear combination of the columns of \mathbf{Z}_k is guaranteed to remain on the boundaries of the active constraints. The matrix \mathbf{Z}_k is taken from the last $m - l$ columns of the QR decomposition of the matrix $\bar{\mathbf{A}}_k$, where l is the number of the active constraints and $l < m$. Therefore, \mathbf{Z}_k is given by

$$\mathbf{Z}_k = \mathbf{Q}_{1, \dots; n, l+1, \dots, m} \quad (\text{A.2})$$

where

$$\mathbf{Q}^T \bar{\mathbf{A}}_k^T = \begin{pmatrix} R \\ 0 \end{pmatrix}. \quad (\text{A.3})$$

After obtaining \mathbf{Z}_k one has to compute a new search direction $\hat{\mathbf{d}}_k$ which minimizes $q(\mathbf{d})$, where $\hat{\mathbf{d}}_k$ is in the null space of the active constraints. Therefore $\hat{\mathbf{d}}_k$ is a linear combination of the columns of \mathbf{Z}_k , which can be denoted by

$$\hat{\mathbf{d}}_k = \mathbf{Z}_k \mathbf{p} \quad (\text{A.4})$$

where \mathbf{p} is an arbitrary vector. When \mathbf{p} is substituted for $\hat{\mathbf{d}}_k$ into the quadratic function a new equation is obtained.

$$q(\mathbf{p}) = \frac{1}{2} \mathbf{p}^T \mathbf{Z}_k^T \mathbf{H} \mathbf{Z}_k \mathbf{p} + \mathbf{c}^T \mathbf{Z}_k \mathbf{p} \quad (\text{A.5})$$

Differentiating equation (A.5) with respect to \mathbf{p} gives the so-called projected gradient $\nabla q(\mathbf{p})$.

$$\nabla q(\mathbf{p}) = \mathbf{Z}_k^T \mathbf{H} \mathbf{Z}_k \mathbf{p} + \mathbf{Z}_k^T \mathbf{c} \quad (\text{A.6})$$

$\nabla q(\mathbf{p})$ is the gradient projected in the subspace defined by \mathbf{Z}_k and the expression $\mathbf{Z}_k^T \mathbf{H} \mathbf{Z}_k$ is called the projected Hessian. As the Hessian matrix \mathbf{H} is positive definite the minimum of the function $q(\mathbf{p})$ in the subspace defined by \mathbf{Z}_k occurs when $\nabla q(\mathbf{p}) = 0$, which is the solution of the system of linear equations.

$$\mathbf{Z}_k^T \mathbf{H} \mathbf{Z}_k \mathbf{p} = -\mathbf{Z}_k^T \mathbf{c} \quad (\text{A.7})$$

Thereafter a step of the form

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha \hat{\mathbf{d}}_k \quad \text{where} \quad \hat{\mathbf{d}}_k = \mathbf{Z}_k^T \mathbf{p} \quad (\text{A.8})$$

can be taken. Due to the quadratic nature of the objective function there are only two choices of step length α at each iteration step. An unity step along $\hat{\mathbf{d}}_k$ represents the exact step to the minimum, restricted to the null space of $\bar{\mathbf{A}}_k$. If such a step can be taken without violation of the constraints,

then the solution of the QP is found. Otherwise, the step along $\hat{\mathbf{d}}_k$ to the nearest constraint is less than unity and a new constraint is included in the active set at the next iteration. The distance to the constraint boundaries in any direction $\hat{\mathbf{d}}_k$ is given by

$$\alpha = \min_i \left\{ \frac{-(\mathbf{A}_i \mathbf{x}_k - \mathbf{b}_i)}{\mathbf{A}_i \hat{\mathbf{d}}_k} \right\} \quad (i = 1, \dots, m) \quad (\text{A.9})$$

which is defined for constraints not in the active set, and where the direction $\hat{\mathbf{d}}_k$ is pointing towards the constraint boundary, i.e.

$$\mathbf{A}_i \hat{\mathbf{d}}_k > 0 \quad (i = 1, \dots, m). \quad (\text{A.10})$$

When n independent constraints are included in the active set, without location of the minimum, Lagrange multipliers, λ_k , are calculated that satisfy the nonsingular set of linear equations

$$\bar{\mathbf{A}}_k^T \lambda_k = \mathbf{c}. \quad (\text{A.11})$$

If all elements of λ_k are positive, \mathbf{x}_k is the optimal solution of the quadratic programming problem. If any component of λ_k is negative, and the component does not correspond to an equality constraint, then the corresponding element is deleted from the active set and a new iteration is made.

B Estimation of Fourier components of unequally sampled data

The treatment of unequally sampled data in Fourier techniques is discussed in this section, following the papers of Foster(I) (1996, [22]) and Foster(II) (1996, [23]). Any time-series consists of N data values $x(t_\alpha)$ taken at a (discrete) set of N times t_α , $\alpha = 1, 2, \dots, N$. Therefore the observed data can be represented as a vector in a N -dimensional vector space which is called sampling space, by taking as the α^{th} component of the vector, the value of the data at time t_α . This can be written by

$$|x\rangle = (x(t_1), x(t_2), \dots, x(t_N)) . \quad (\text{B.1})$$

Any other function of time $f(t)$ defines a vector in sampling space as

$$|f\rangle = (f(t_1), f(t_2), \dots, f(t_N)) . \quad (\text{B.2})$$

The structure of this vector space can be imposed by defining the inner product of two vectors $|f\rangle$ and $|h\rangle$ as the (weighted) average of the product fh over the sampling t_α

$$\langle f|h\rangle = \frac{\sum_{\alpha=1}^N \tilde{w}_\alpha f(t_\alpha) h(t_\alpha)}{\sum_{\beta=1}^N \tilde{w}_\beta} \quad (\text{B.3})$$

where \tilde{w}_α is the statistical weight assigned to a data point at α . For further properties of this vector space it is referred to the paper. Using the constant vector $|\mathbf{1}\rangle = [\mathbf{1}, \mathbf{1}, \dots, \mathbf{1}]$ (of unit norm) one can define the sampling-dependent average

$$\langle f\rangle = \frac{\sum_\alpha \tilde{w}_\alpha f(t_\alpha)}{\sum_\beta \tilde{w}_\beta} = \langle \mathbf{1}|f\rangle . \quad (\text{B.4})$$

The variation of a vector $|f\rangle$ is the squared norm of that part of $|f\rangle$, which is orthogonal to the constant vector, i.e.

$$V_f = \langle f|f\rangle - \langle \mathbf{1}|f\rangle^2 . \quad (\text{B.5})$$

The average and the mean of the data can be denoted by

$$\bar{x} = \langle \mathbf{1}|x\rangle \quad V_x = \langle x|x\rangle - \langle \mathbf{1}|x\rangle^2 \quad (\text{B.6})$$

which leads to the estimated variance, finally.

$$s^2 = \frac{NV_x}{(N-1)} \quad (\text{B.7})$$

The observed data vector should be now projected onto the subspace of sampling space spanned by a set of r trial functions $\phi_a(t)$, $a = 1, 2, \dots, r$. These trial functions define a set of trial vectors

$$|\phi_a\rangle = (\phi_a(t_1), \phi_a(t_2), \dots, \phi_a(t_N)) \quad (\text{B.8})$$

and it is required that all trial vectors are linearly independent. The space spanned by these vectors will be called model space of dimension $r \leq N$. Any vector $|y\rangle$ in model space is a linear combination of trial vectors with constant coefficients y_a .

$$|y\rangle = \sum_a y_a |\phi_a\rangle \quad (\text{B.9})$$

The inner product of two models-space vectors $|y\rangle$, $|z\rangle$ can be easily gained by

$$\langle y|z\rangle = \sum_a \sum_b y_a z_b \langle \phi_a|\phi_b\rangle = \sum_a \sum_b S_{ab} y_a z_b . \quad (\text{B.10})$$

This defines the S -matrix

$$S_{ab} = \langle \phi_a | \phi_b \rangle \quad (\text{B.11})$$

which is a metric tensor for model space when the trial functions are used as a basis. Projecting the data vector onto model space results in a model vector $|y\rangle$, which lies in the subspace and a residual vector $|R\rangle$, which is orthogonal to the subspace.

$$|x\rangle = \sum_a y_a |\phi_a\rangle + |R\rangle \quad (\text{B.12})$$

$$\langle \phi_a | R \rangle = 0 \text{ for all } a \quad (\text{B.13})$$

Taking the inner product of (B.12) with the trial vectors and using (B.13) gives

$$\langle \phi_a | x \rangle = \sum_b S_{ab} y_b. \quad (\text{B.14})$$

Inverting the S -matrix yields the expansion coefficients

$$y_a = \sum_b S_{ab}^{-1} \langle \phi_b | x \rangle = \frac{\sum_b \sum_\alpha S_{ab}^{-1} \tilde{w}_\alpha \phi_b(t_\alpha) x(t_\alpha)}{\sum_\lambda \tilde{w}_\lambda}. \quad (\text{B.15})$$

If there is no time variation to detect, the signal is only a constant mean value.

$$|F\rangle = \mu |\mathbf{1}\rangle \quad (\text{B.16})$$

As the data vector is the sum of the signal vector on errors it is assumed that the errors have mean zero value.

$$\langle \epsilon(t_\alpha) \rangle = 0 \quad (\text{B.17})$$

Furthermore the weights correspond to the variances by

$$\tilde{w}_\alpha = \frac{1}{\sigma_\alpha^2}. \quad (\text{B.18})$$

and it is assumed that all the errors are independent, so for any two different errors $\epsilon(t_\alpha)$ and $\epsilon(t_\beta)$ $\alpha \neq \beta$ and any two exponents m and n

$$\langle (\epsilon(t_\alpha))^m (\epsilon(t_\beta))^n \rangle = \langle (\epsilon(t_\alpha))^m \rangle \langle (\epsilon(t_\beta))^n \rangle \quad (\text{B.19})$$

is obtained. This implies also that

$$\langle \epsilon(t_\alpha) \epsilon(t_\beta) \rangle = 0 \text{ for } \alpha \neq \beta. \quad (\text{B.20})$$

The assumptions (B.16) - (B.19) form the basic null hypothesis (BNH), which will be the basis for statistical evaluation of a variety of analysis methods. It can be concluded that when the BNH holds, the projection coefficients y_a follow a normal distribution (from central limit theorem). The expected value of a projection coefficient will be

$$\langle y_a \rangle = \mu \sum_b S_{ab}^{-1} \langle \phi_b | \mathbf{1} \rangle \quad (\text{B.21})$$

and the expectation value of the product of two projection coefficients is given by

$$\langle y_a y_b \rangle = \mu^2 \sum_c \sum_d S_{ac}^{-1} S_{bd}^{-1} \langle \phi_c | \mathbf{1} \rangle \langle \phi_d | \mathbf{1} \rangle + N^{-1} \sigma^2 S_{ab}^{-1}. \quad (\text{B.22})$$

Therefore the covariance between two coefficients can be derived

$$\langle y_a y_b \rangle - \langle y_a \rangle \langle y_b \rangle = N^{-1} \sigma^2 S_{ab}^{-1} \quad (\text{B.23})$$

and it can be seen that they are not uncorrelated. If trial functions are orthogonal then the S-matrix is diagonal and so its inverse. Then two different coefficients are uncorrelated. When the trial functions are orthonormal, i.e. $S_{ab} = S_{ab}^{-1} = \delta_{ab}$ all coefficients have the same variance of $N^{-1}\sigma^2$. If the subspace spanned by r trial vectors $|\phi_a\rangle$ should be orthogonal to the constant vector $|\mathbf{1}\rangle$ then all its generators must be orthogonal to $|\mathbf{1}\rangle$, so equation (B.21) simplifies to

$$\langle y_a \rangle = 0. \quad (\text{B.24})$$

If an orthonormal set $|\psi_a\rangle$ which spans the same subspace as the r trial vectors $|\phi_a\rangle$ is introduced with the corresponding projection coefficients z_a then equation (B.22) will become

$$\langle z_a z_b \rangle = N^{-1}\sigma^2\delta_{ab}. \quad (\text{B.25})$$

Since $|\phi_a\rangle$ and $|\psi_a\rangle$ span the same subspace, they define the same projection; the model vectors $|y\rangle$ and $|z\rangle$ are identical, and so are their variations V_y and V_z . Hence it is true for any basis (as long as the trial subspace is orthogonal to $|\mathbf{1}\rangle$), that the variation of the model function V_y is chi-square. Its expected value is

$$\langle V_y \rangle = rN^{-1}\sigma^2. \quad (\text{B.26})$$

This leads to a statistic

$$P = \frac{NV_y}{r\sigma^2} \quad (\text{B.27})$$

which follows the chi-square distribution with r degrees of freedom and expectation value 1. Principally σ is unknown, but it can be estimated by equation (B.7). This leads to the so-called universal power statistic (UPS).

$$P = \frac{NV_y}{rs^2} \quad (\text{B.28})$$

If the projection subspace is chosen to include the constant vector $|\mathbf{1}\rangle$ UPS is reduced in its degree of freedom and should be calculated as follows.

$$P = \frac{NV_y}{(r-1)s^2} \quad (\text{B.29})$$

As the residual vector is orthogonal to the model space these two projections are independent and the expected value of this variation is given by $N^{-1}\sigma^2$ times the number of degrees of freedom (without $|\mathbf{1}\rangle$). This makes it possible to substitute a value for σ and the following equation is obtained (trial subspace orthogonal to $|\mathbf{1}\rangle$)

$$Z = \frac{(N-r-1)V_y}{rV_R} \quad (\text{B.30})$$

or including $|\mathbf{1}\rangle$

$$Z = \frac{(N-r)V_y}{(r-1)V_R} \quad (\text{B.31})$$

where V_R as the total variation of the the residual vector. Z follows a F-statistic with $N-r$ and $r-1$ degrees of freedom and $\langle Z \rangle = 1$. For weighted data it is necessary to introduce the effective number of data points defined by

$$N_{eff} = \frac{(\sum \tilde{w}_\alpha)^2}{\sum \tilde{w}_\alpha^2}. \quad (\text{B.32})$$

If correlations (or covariances) between the data

$$\gamma_k = \langle \epsilon(t_\alpha)\epsilon(t_\beta) \rangle, \quad k = |\alpha - \beta| \quad (\text{B.33})$$

are considered one has to extend the theory. Equation (B.33) in connection with (B.16), (B.17) and (B.18) define the Correlated Random Variable Hypothesis (CRVH). This leads to a new expectation value of

$$\langle V_y \rangle = rN^{-1} + N^{-2} \sum_{k=1}^{N-1} \gamma_k \sum_{a=1}^r \sum_{b=1}^r S_{ab}^{-1} \cdot \sum_{\alpha=1}^{N-k} [\phi_a(t_\alpha)\phi_b(t_{\alpha+k}) + \phi_a(t_{\alpha+k})\phi_b(t_\alpha)]. \quad (\text{B.34})$$

The last one can be approximated by the spectral density at zero frequency and results in

$$\langle V_y \rangle \approx rN^{-1} \left[\sigma^2 + 2 \sum_{k=1}^{N-1} \gamma_k \right] = rN^{-1} \tau^2 \quad (\text{B.35})$$

where τ^2 is the modified variance. If it is obvious that the data shows some variation instead of BNH the Strong Signal hypothesis (SSH) has to be applied. Therefore one can write

$$|F\rangle = \mu|\mathbf{1}\rangle + |f\rangle \quad (\text{B.36})$$

where the signal variation vector $|f\rangle$ is constrained to satisfy

$$\langle \mathbf{1} | f \rangle = 0 \quad (\text{B.37})$$

and for the expectation value a contribution from the signal variation has to be taken into account

$$\langle V_x \rangle = \langle f | f \rangle + (1 - N^{-1})\sigma^2. \quad (\text{B.38})$$

If the trial subspace includes $|\mathbf{1}\rangle$ a new statistics is obtained.

$$s_{SSH}^2 = \frac{NV_R}{N-r} \quad (\text{B.39})$$

If $a = b$ is set in equation (B.23) the error in any given coefficient can be estimated by

$$\text{Var}(y_a) = \langle y_a^2 \rangle - \langle y_a \rangle^2 = N^{-1} \sigma^2 S_{aa}^{-1}. \quad (\text{B.40})$$

The expected value of the variation of the model vector is

$$\langle V_y \rangle = V_{y(F)} + rN^{-1} \sigma^2 \quad (\text{B.41})$$

where $V_{y(F)}$ is the variation of the model vector (for the signal alone, no noise). And finally

$$\text{Var}(V_y) = \langle V_y^2 \rangle - \langle V_y \rangle^2 = 4N^{-1} \sigma^2 V_{y(F)} + 2rN^{-2} \sigma^4. \quad (\text{B.42})$$

The pure discrete Fourier transform (PDFT) is defined only for a time-series with perfectly even time spacing.

$$t_\alpha = \alpha\tau \quad T = N\tau \quad (\text{B.43})$$

Only trial functions $\mathbf{1}(t)$ and $\sqrt{2}$ times $\sin(\omega_k t)$ and $\cos(\omega_k t)$ that satisfy the boundary condition

$$\omega_k T = 2\pi k \quad (\text{B.44})$$

for integer $1 \leq k \leq \frac{1}{2}N$ are taken. This leads to

$$\langle \cos(\omega_j t) | \cos(\omega_k t) \rangle = \frac{1}{2} \delta_{jk} \quad (\text{B.45})$$

$$\langle \sin(\omega_j t) | \sin(\omega_k t) \rangle = \frac{1}{2} \delta_{jk} \quad (\text{B.46})$$

$$\langle \sin(\omega_j t) | \cos(\omega_k t) \rangle = 0 \quad (\text{B.47})$$

It can be seen that ortho-normality is given and that the S-matrix is equal to the identity matrix. Therefore the projection coefficients can be computed very simply

$$\text{coeff}(\sqrt{2} \cos(\omega_k t)) = \langle \sqrt{2} \cos(\omega_k t) | x \rangle \quad (\text{B.48})$$

$$\text{coeff}(\sqrt{2} \sin(\omega_k t)) = \langle \sqrt{2} \sin(\omega_k t) | x \rangle \quad (\text{B.49})$$

$$\text{coeff}(\mathbf{1}) = \langle \mathbf{1} | x \rangle = \bar{x} \quad (\text{B.50})$$

The squared norm of the projection onto the subspace spanned by the trial function is given by

$$Q = \langle y|y \rangle = 2[\langle \cos(\omega_k t)|x \rangle^2 + \langle \sin(\omega_k t)|x \rangle^2] \quad (\text{B.51})$$

where $|y \rangle$ is the projected data vector for a single trial frequency ω_k , only. The (un-scaled) power is defined by

$$P = \frac{1}{2}Q = \frac{1}{2}\langle y|y \rangle \quad (\text{B.52})$$

or

$$P = \langle \cos(\omega_k t)|x \rangle^2 + \langle \sin(\omega_k t)|x \rangle^2 \quad (\text{B.53})$$

and under the assumption of BNH the power P is proportional to the UPS. It is a chi-square statistics with 2 degrees of freedom and expected value $N^{-1}\sigma^2$. For unevenly spaced data (or when the frequency does not satisfy the boundary condition) the discrete Fourier transform (DFT) is still defined by equation (B.53). It can be proved that this transform does not show orthogonality (esp. to $|\mathbf{1}\rangle$) and it is absolutely necessary to zero the data.

$$|z \rangle = |x \rangle - \bar{x}|\mathbf{1}\rangle \quad (\text{B.54})$$

The expected value of the DFT power can be derived by

$$\langle P \rangle = P_F + N^{-1}\sigma^2[1 - |W(\omega)|^2] \quad (\text{B.55})$$

where P_F is the power level due to signal with no noise and $W(\omega)$ is the complex window function.

$$W(\omega) = C(\omega) + jB(\omega) = \langle \mathbf{1} | \cos(\omega t) \rangle + j \langle \mathbf{1} | \sin(\omega t) \rangle \quad (\text{B.56})$$

If definition (B.52) is used instead of (B.53) a genuine projection onto the subspace spanned by the two trial functions $\sin(\omega t)$ and $\cos(\omega t)$. This is equivalent to Scargles modified periodogram (Scargle, 1982, [60]).

$$P_x(\omega) = \frac{1}{2} \left\{ \frac{[\sum x(t) \cos(\omega(t - \tau))]^2}{\sum \cos^2(\omega(t - \tau))} + \frac{[\sum x(t) \sin(\omega(t - \tau))]^2}{\sum \sin^2(\omega(t - \tau))} \right\} \quad (\text{B.57})$$

whereas the time delay τ has to be chosen as follows.

$$\tan(2\omega\tau) = \frac{\sum \sin(2\omega t)}{\sum \cos(2\omega t)} \quad (\text{B.58})$$

This is truly a projection, but still ignores the fact that the trial functions are not orthogonal to $|\mathbf{1}\rangle$. This problem was overcome by the Data-Compensated Discrete Fourier Transform (DCDFT). The logical extension of this approach to multi-periodic signals is to project the observed data onto trial vectors $|\mathbf{1}\rangle$ and $|\sin(\omega_A t)\rangle$ and $|\cos(\omega_A t)\rangle$ for a set of frequencies ω_A , $A = 1, 2, \dots, f$, which leads to the definition of power

$$P = \frac{NV_y}{2fs^2} \quad (\text{B.59})$$

Techniques using this or similar approaches can be seen as the basis for other concepts like CLEAN, which tries to isolate “real” periods from artificial ones by means of statistics.

C BWS plots for baseline HARTRAO - HOBART26, Jan. 7th, 2004

Figures 85 and 86 give the results from bandwidth synthesis for S- and X-band channels on the baseline HARTRAO-HOBART26 within experiment SYW029, carried out on Jan. 7th, 2004 from 8:48:00 UT until 8:54:50 UT. Information was extracted from KOMB databases⁷ provided by Geographical Survey Institute (GSI), Japan.

```

PLOT#2 by KOMB(KSP) OBS# = 7
P.292I-----I-----I4-----I-----I-----I-----6I-----444444-444-I-----
H I I I 2 I3 I I I I 6 I I I I I 007-0851 FRQG S
S 1 1I I I I 4 4I 5 I I P PI I 07JAN.04 1954-388
E I I I I 3 I 4 I I I I I HARTRAO -HOBART26=JH
I 1 I I I I I 5I 6 66 I I I I 59.5 ELEV(DEG) 26.5
I I I 2 2 I I I 5 I I I I I FILE K00009 KR# 1
I I I 2 I I I I 5 5 I 6 6 I P I I I BN(mHz/asec) .0
I 1 1 I I I 3 I 4 I I I 6 I I P I I BE(mHz/asec) .0
I I I 2 I I 3 I 4 4 I I I I P P I I EXP SYW029
I I I I 3 I 4 I I I I P I I OBS# 7
I 1 I I I 33 I 4 I I 5 5 I I I I START TIME 084813.00
I I I I I I I I I I 6 I P I I STOP TIME 085449.00
I I I I I I I I I I I EEEEEEEI INTG.TIME 396.000SEC
I I I I I I 4 I 5 I I I IEEEEE I SEARCH WINDOW WIDTHS
I I I 2 I I 3 I 4 I I I 6 I P E I SINGLB(=) 4.000000us
I I I 2 I 3 I I I I I I MULTIB(=) .200000us
I I I I I 4 I I I I P I I RATE (X) .500Hz
I I I I I I I I I I I FREQ(MHz) SB CAL
I I I I I I I I I I I 2207.99 U 27.49
I I I I I I I I I I I 2212.99 U 30.47
I 1 I I I I I I I I I I 2227.99 U 31.51
I I I I I I I I I I I 2257.99 U 26.50
I I I I I I I I I I I 2282.99 U 21.49
I I I I I I I I I I I 2292.99 U 29.47
I I I I I I I I I I I 11 I
I I I I I I I I I I I 1 1 I
I I I I I I I I I I I 1 1 1 I
I I I I I I I I I I I 155555555555I
I I I I I I I I I I I 51 666I
I I I U I I I I I I I 666666666222I
I I I I I I I I I I I 2 DDDDDI
I I I UI I I I I I I I 33DDDDFFFFFI KOMB for KSP Ver 0.0
I I I U I UU UU I I U I I DFFFF I AMP 2.898.01%
I I I U U I I I I U I U I T F SINGLBAND .003619us
I U U UIU I U U I U I U I U I T I MULTIBAND -.031711us
IU U U UIU I U U I U UIU I U I T T TI FR.RATE .003095Hz
I U U U U U IU U I U UU UI I U T T T I PEYD= .PEYD= .
U1U U 2 U I U U U IU U IU U UU I U UUIT T T I SUPR= .PPUP= .
I U I2 I I U U U U IU I I CCCCCCI XSLP= .YSLP= .
I U I I 4 I I I U I I CCCC I
I I I 23 I I 6 U PP CC I KOMVAL 1001
0 I-----I-----I5-----I6-----I-----I-----4--44I-----
0 12 24 36 48 60 72 84 96 SNR 22.0
E-03 FRINGE PHASE(-180 TO +180 deg) and AMPLITUDE as a function of TIME by FREQ. channel PCAL PHASES
FRQ# 1 FRQ# 2 FRQ# 3 FRQ# 4 FRQ# 5 FRQ# 6 FRQ# 7 FRQ# 8 FRQ# 9 FRQ#10 FRQ#11 FRQ#12 FRQ#13 FRQ#14 ALL CHANS
PHASE = 128.2 132.1 111.3 126.6 105.6 139.2 .0 .0 .0 .0 .0 .0 .0 .0 123.4 ERR= 0
FR. AMPL 2.3 3.5 4.1 2.5 2.5 2.8 .0 .0 .0 .0 .0 .0 .0 .0 2.9 GHA= 5.9H
PCALPH1 -45.8 -73.3 -87.3 178.8 -54.3 -69.6 .0 .0 .0 .0 .0 .0 .0 .0 EP1=-192.0
PCALPH2 -89.9 -157.7 -158.3 -86.9 77.0 -94.4 .0 .0 .0 .0 .0 .0 .0 .0 EP2=-6.0
SDBBOX# 35.7 34.9 34.8 35.2 34.6 35.6 .0 .0 .0 .0 .0 .0 .0 .0 33.1 BW=4.00MHz
SRCH RNGE PARAMS(RATE(us/s),FDEL(us),CDEL(us)) -.113225E-03 .113225E-03 -.100000E+00 .100000E+00 -.200000E+01 .200000E+01
GROUP DELAY usec .873104759980D+04 APRIORI DELAY usec .873107931078D+04 RESIDUAL DELAY usec -.317110E-01 +/- .22E-03
SINGLBAND DEL usec .873108292945D+04 APRIORI CLOCK usec .211999995372D+02 RESIDUAL DELAY usec .361867E-02 +/- .62E-02
PHASE DELAY usec .873107946602D+04 APRIORI CLOCK us/s .000000000000D+00 RESIDUAL DELAY usec .155241E-03 +/- .33E-05
PHAS DEL RATE us/s .147726905410D+01 APRIORI RATE us/s .147726769175D+01 RESIDUAL RATE us/s .136235E-05 +/- .29E-07
TOTAL PHASE deg 54.1 ( 344.7) APRIORI PHASE deg 290.7 RESIDUAL PHASE deg 123.4( 123.4)/+- 2.61
EPOCH HMMSS. 085125. ( 085131. ) REF FREQUENCY MHz 2207.99 AMP 2.898( 4.181)/+- .132 QF= 9 QB= 0%
RA 195759.819 DEC-384506.356 2000.0 APRIORI ACEL us/s/s -.634413982861D-04 REF.ST.CL.EPOCH ms -.300000E-03 TAPEQ=999999
RMS PH/SEG DG 36.0 RMS AMP/SEG % 68.3 RMS PH/FRQ DG 11.6 RMS AMP/FRQ % 22.0 TH.RMS 22.0 38.3 5.8 10.2 04254:2029:00
KOMB STATM=SLIP INCX
    
```

Figure 85: S-Band bandwidth synthesis result.

⁷more information at <http://www2.nict.go.jp/ka/radioastro/>

```

PLOT#2 by KOMB(KSP)                                OBS# = 7
-----I-----I-----I-----I-----I-----I-----I-----I-----I661--3--I 007-0851  FRQ# X
H I I I I I I I I I 61 F33 GI 07JAN.04 1954-388
S I I I I I I I I I 13F3 G I HARTRAO -HOBART26=JH
E I I I I I I I I I 3F GG I 59.5 ELEV(DEG) 26.5
I I I I I I I I I I I G DEI FILE K00009 KR# 1
I I I I I I I I I I F G EE I BN(mHz/asec) .0
I I I I I I I I I I I G E AAI BE(mHz/asec) .0
I I I I I I I I I I I G EE A I EXP SYWO29
I I I I I I I I I I IG D AA I OBS# 7
I I I I I I I I I I I DEA I START TIME 084813.00
I I I I I I I I I I G EA I STOP TIME 085449.00
I I I I I I I I I I IEA I INTG.TIME 396.000SEC
I I I I I I I I I I E I SEARCH WINDOW WIDTHS
I I I I I I I I I I IA I SINGLB(=) 4.000000us
I I I I I I I I I I A HI MULTIB(#).100000us
I I I I I I I I I I I HI RATE (X) .500Hz
I I I I I I I I I I I HH I FREQ(MHz) SB CAL
I I I I I I I I I I I H I 8210.99 U 15 21
I I I I I I I I I I I H 44I 8220.99 U 42 23
I I I I I I I I I I I H 444 I 8250.99 U 42 24
I I I I I I I I I I I H44 CI 8310.99 U 15 22
I I I I I I I I I I 444 C I 8420.99 U 49 23
I I I I I I I I I I IH CC I 8500.99 U 51 25
I I I I I I 4 I I I I I H C BI 8550.99 U 49 23
I I I I I I I I I I I C B I 8570.99 U 59 24
I I I I I I I I I I I CC BB I
I 11 I 2 I I I I I 5 I 6 I 8 I P I B I
I I 2 I 3 I I I I 7 I 8 I IC B I
I I I 3 I 4 I I I 77 I 8 I P C B 888I
I I 2 I I 4 I 5 I I 7 I 8 I P I B 88 I
I 1 I I 33 4 5 I I I I IB888 777I KOMB for KSP Ver 0.0
I I 2 I 3 I 4 I I 6 I 7 I P 88 577 I AMP 4.306X.01%
I U 1 I U I U 14 I 5 5 U U 6 I U U UI 8 I B5777 I SINGLBAND .015191us
I U U I 2U U IUU U I 4 UU 4155 5IU U UI UU U I8 I 77 I MULTIBAND -.041945us
UU U IUU U 2U 3UUU 3UU U I UUU U I UU U U7U7 U IU8 U UU 8TTPP TT TI I FR.RATE .009578Hz
I11 UI U UI 3 U I U U UUUU U I 6 U I 7 U U UIP T TPI 22 FI SUPR= .PPUP= .
I U U U U2 2 U I3 I U 4 I U I6 6 6I 7I U I I 222 F I XSLP= .YSLP= .
I 1 1 I 2 I 3 I I 5 I I 7 I 8 I P 22 FF66I
I 1I I I I I 6 I I I 6F6 3I KOMVAL 1002
0 I-----I-----I-----I-----I-----I-----I-----I-----I--6F--33-I SNR 37.6
0 10 20 30 40 50 60 70 80 90 100
E-03 FRINGE PHASE(-180 TO +180 deg) and AMPLITUDE as a function of TIME by FREQ. channel PCAL PHASES
FRQ# 1 FRQ# 2 FRQ# 3 FRQ# 4 FRQ# 5 FRQ# 6 FRQ# 7 FRQ# 8 FRQ# 9 FRQ#10 FRQ#11 FRQ#12 FRQ#13 FRQ#14 ALL CHANS
PHASE = -117.4 -110.2 -118.1 -97.9 -105.3 -128.4 -111.5 -102.2 .0 .0 .0 .0 .0 .0 -112.0 ERR= 0
FR. AMPL 3.7 4.2 4.6 4.3 4.1 5.3 5.0 3.7 .0 .0 .0 .0 .0 .0 4.3 GHA= 5.9H
PCALPH1 -174.5 -147.6 171.8 3.2 -100.3 -173.0 -102.2 -84.0 .0 .0 .0 .0 .0 .0 EP1=-192.0
PCALPH2 97.6 -61.7 -32.8 114.3 112.6 -178.8 135.0 17.4 .0 .0 .0 .0 .0 .0 EP2=-6.0
SDBBOX# 35.1 35.0 35.0 35.4 35.4 35.2 35.6 35.3 .0 .0 .0 .0 .0 .0 33.2 BW=4.00MHz
SRCH RNGE PARAMS(RATE(us/s),FDEL(us),CDEL(us)) -.304470E-04 .304470E-04 -.500000E-01 .500000E-01 -.200000E+01 .200000E+01
GROUP DELAY usec .873103736561D+04 APRIORI DELAY usec .873107931078D+04 RESIDUAL DELAY usec -.419452E-01 +/- .30E-04
SINGLBAND DEL usec .873109450191D+04 APRIORI CLOCK usec .211999995372D+02 RESIDUAL DELAY usec .151911E-01 +/- .38E-02
PHASE DELAY usec .873107927288D+04 APRIORI CLOCK us/s .000000000000D+00 RESIDUAL DELAY usec -.378958E-04 +/- .52E-06
PHAS DEL RATE us/s .147726881441D+01 APRIORI RATE us/s .147726769175D+01 RESIDUAL RATE us/s .112265E-05 +/- .45E-08
TOTAL PHASE deg 215.6 ( 272.6) APRIORI PHASE deg 327.6 RESIDUAL PHASE deg -112.0(-112.0)/+/- 1.53
EPOCH HMMSS. 085125. ( 085131. ) REF FREQUENCY MHz 8210.99 AMP 4.306( 5.275)/+/- .115 QF= 9 QB= 0%
RA 195759.819 DEC-384506.356 2000.0 APRIORI ACEL us/s/s -.634413982861D-04 REF.ST.CL.EPOCH ms -.300000E-03 TAPEQ=999999
RMS PH/SEG DG 31.9 RMS AMP/SEG % 36.0 RMS PH/FRQ DG 9.1 RMS AMP/FRQ % 12.3 TH.RMS 13.6 23.7 4.0 7.0 04254:2029:00
KOMB STATM=SLIP INCX
    
```

Figure 86: X-Band bandwidth synthesis result.

References

- [1] E.L. Afraimovich, E.A. Kosogorov, K.S. Palamarchouk, N.P. Perevalova and A.V. Plotnikov. The use of GPS arrays in detecting the ionospheric response during rocket launches. *Earth Planets Space*, Volume 52, pages 1061–1066, 2000.
- [2] E.L. Afraimovich, N.P. Perevalova and S.V. Voeikov. Traveling wave packets of total electron content disturbances as deduced from global GPS network data. *Journal of Atmospheric and Solar-terrestrial Physics*, Volume 65, pages 1245–1262, 2003.
- [3] Altimeter Database System (ADS), GeoForschungsZentrum Potsdam. http://adsc.gfz-potsdam.de/ads/adsCentral_index.html, 2005.
- [4] S. Baisch and G.H.R. Bokelmann. Spectral analysis with incomplete time series: an example from seismology. *Computers & Geosciences*, Volume 25, pages 739–750, 1999.
- [5] S. Bergstrand and R. Haas. Comparison of Ionospheric Activity Derived from GPS and Different VLBI Networks. In N. R. Vandenberg and K. D. Baver (editors), *International VLBI Service for Geodesy and Astrometry 2004 General Meeting Proceedings*, NASA/CP-2004-212255, pages 447–451, 2004.
- [6] D. Bilitza. International Reference Ionosphere 2000. *Radio Sci.*, Volume 36, Number 2, pages 261–275, 2001.
- [7] D. Bilitza, K. Rawer and S. Pallaschke. Study of ionospheric models for satellite orbit determination. *Radio Sci.*, Volume 23, Number 3, pages 223–232, 1988.
- [8] J. Böhm and H. Schuh. Vienna Mapping Functions. In *Proceedings of the 16th Working Meeting on European VLBI for Geodesy and Astrometry*, pages 131–143. Bundesamt für Kartographie und Geodäsie, Leipzig/Frankfurt am Main, 2003.
- [9] J. Böhm and H. Schuh. Vienna mapping functions in VLBI analyses. *Geophysical Research Letters*, Volume 31, 2004.
- [10] R.N. Bracewell. *The Fourier Transform and Its Applications*. McGraw-Hill, Inc., 1986.
- [11] K.G. Budden. *The propagation of radio waves: the theory of radio waves of low power in the ionosphere and magnetosphere*. Cambridge University Press, 1988.
- [12] B. K. Ching and Y. T. Chiu. A Phenomenological Model of Global Ionospheric Electron Density in the E-, F1-, and F2-Regions. *J. Atmos. Terr. Phys.*, Volume 35, pages 1615–1630, 1973.
- [13] Y. T. Chiu. An improved phenomenological model of ionospheric density. *J. Atmos. Terr. Phys.*, Volume 37, pages 1563–1570, 1975.
- [14] B.G. Clark. Interferometers and Coherence Theory. In J.A. Zensus, P.J. Diamond and P.J. Napier (editors), *Very Long Baseline Interferometry and the VLBA*, Volume 82 of *ASP Conference Series*, pages 3–16, 1995.
- [15] D.S. Coco, S.R. Dahlke and C. Coker. Effects of GPS system biases on differential group delay measurements. Technical Report ARL-88-32, Air Force Geophysics Laboratory, Hanscom AFB, Massachusetts, USA, 1998.
- [16] T.F. Coleman and Y. Li. A Reflective Newton Method for Minimizing a Quadratic Function Subject to Bounds on some of the Variables. *SIAM Journal on Optimization*, Volume 6, Number 4, pages 1040–1058, 1996.
- [17] R. E. Daniell, L. D. Brown, D. N. Anderson, M. W. Fox, P. H. Doherty, D. T. Decker, J. J. Sojka and R. W. Schunk. Parameterized ionospheric model: A global ionospheric parameterization based on first principles models. *Radio Sci.*, Volume 30, Number 5, pages 1499–1510, 1995.

- [18] W. Demtröder. *Experimentalphysik 2 - Elektrizität und Optik*. Springer Verlag, 1999.
- [19] V. Ducic, J. Artru and P. Lognonne. Ionospheric remote sensing of the Denali Earthquake Rayleigh surface waves. *Geophysical Research Letters*, Volume 30, Number 18, pages 1951–1958, 2003.
- [20] European VLBI Network. <http://www.evlbi.org>, 2005.
- [21] D. Fischer, A. Nothnagel, R. Kilger, S. Kurihara, W. Schlüter and K. Takashima. UT1 Intensive Series Using K4 Technology. In N. R. Vandenberg and K. D. Baver (editors), *International VLBI Service for Geodesy and Astrometry 2004 General Meeting Proceedings*, NASA/CP-2004-212255, pages 399–402, 2004.
- [22] G. Foster. Time series analysis by projection. I. Statistical properties of Fourier analysis. *The Astronomical Journal*, Volume 111, Number 1, pages 541–554, 1996.
- [23] G. Foster. Time series analysis by projection. II. Tensor methods for time series analysis. *The Astronomical Journal*, Volume 111, Number 1, pages 555–566, 1996.
- [24] G. Foster. Wavelets for period analysis of unevenly sampled time series. *Astronomical Journal*, Volume 112, Number 4, pages 1709–1929, 1996.
- [25] P.E. Gill, W. Murray, M.A. Saunders and M.H. Wright. Procedures for Optimization Problems with a Mixture of Bounds and General Linear Constraints. *ACM Trans. Math. Software*, Volume 10, Number 3, pages 282–298, 1984.
- [26] P.E. Gill, W. Murray and M.H. Wright. *Practical Optimization*. Academic Press, London, UK, 1981.
- [27] P.E. Gill, W. Murray and M.H. Wright. *Numerical Linear Algebra and Optimization*. Addison Wesley, 1991.
- [28] O. Hakegard. *A regional ionospheric model for real-time predictions of the total electron content in wide area differential satellite navigation systems*. Ph.D. thesis, Norwegian Institute of Technology, Norway, 1995.
- [29] J. K. Hargreaves. *The solar-terrestrial environment*. Cambridge University Press, 1992.
- [30] M. Hawarey, T. Hobiger and H. Schuh. Effects of the 2nd order ionospheric terms on VLBI measurements. *Geophysical Research Letters*, Volume 32, 2005.
- [31] M. Hernández-Pajares. Summary of IGS Ionosphere WG Activities (April 2004 - 2005). Technical report, http://gage152.upc.es/~ionex3/doc/Summary_of_IGS_Iono_WG_activities_2004-2005_v4.pdf, 2005.
- [32] T. Hobiger and T. Kondo. An FX software correlator based on Matlab. In *Proceedings of the 17th Working Meeting on European VLBI for Geodesy and Astrometry*, 2005.
- [33] T. Hobiger and H. Schuh. Modelling Vertical Total Electron Content from VLBI Observations. In N. R. Vandenberg and K. D. Baver (editors), *International VLBI Service for Geodesy and Astrometry 2004 General Meeting Proceedings*, NASA/CP-2004-212255, pages 306–310, 2004.
- [34] B. Hofmann-Wellenhof, H. Lichtenegger and J. Collins. *Global Positioning System - Theory and Practice*. Springer, 2001.
- [35] W.-C. Hu and H.-T. Sheu. Quadratic B-spline for Curve Fitting. *Proceedings of the National Science Council, Part A: Physical Science and Engineering, Repub. of China*, Volume 24, Number 5, pages 373–381, 2000.

- [36] International Geomagnetic Reference Field (IGRF). <http://swdcwww.kugi.kyoto-u.ac.jp/igrf/index.html>, 2005.
- [37] International GNSS Service (IGS). <http://igsceb.jpl.nasa.gov/>, 2005.
- [38] International VLBI Service for Geodesy and Astrometry. <http://ivscc.gsfc.nasa.gov>, 2005.
- [39] G. Kaiser. *A Friendly Guide to Wavelet*. Birkhäuser, 1994.
- [40] S. Kedar, G.A. Hajj, B.D. Wilson and M.B. Heflin. The effect of the second order GPS ionospheric correction on receiver positions. *Geophysical Research Letters*, Volume 30, Number 16, pages 1829–1836, 2003.
- [41] W. Keller. *Wavelets in Geodesy and Geodynamics*. de Gruyter, 2004.
- [42] K.-R. Koch. *Parameter Estimation and Hypothesis Testing in Linear Models*. Springer, Berlin, 1997.
- [43] T. Kondo. Application of VLBI data to measurements of ionospheric total electron content. In *Journal of the Communications Research Laboratory*, Volume 38, pages 613–622. Communications Research Institute, Ministry of Posts and Telecommunications, Tokyo, Japan, 1991.
- [44] T. Kondo. VLBI and ionosphere. In *Proceedings of the International Workshop for Reference Frame Establishment and Technical Development in Space Geodesy*, pages 53–60, 1993.
- [45] T. Kondo, M. Kimura, Y. Koyama and H. Osaki. Current Status of Software Correlators Developed at Kashima Space Research Center. In N. R. Vandenberg and K. D. Baver (editors), *International VLBI Service for Geodesy and Astrometry 2004 General Meeting Proceedings*, NASA/CP-2004-212255, pages 186–190, 2004.
- [46] Y. Koyama, T. Kondo, H. Osaki, A.R. Whitney and K.A. Dudevoir. Rapid Turnaround EOP Measurements by VLBI over the Internet. In F. Sanso (editor), *A Window on the Future of Geodesy, Proceedings of the IAG General Assembly, Sapporo, Japan, June 30 - July 11, 2003*, Volume 128, pages 119–124. IAG, Springer, 2005.
- [47] J. D. Kraus. *Radio Astronomy*. Cygnus-Quasar Books, 1986.
- [48] A. Lara, N. Gopalswamy, S. Nunes, G. Munoz and S. Yashiro. A statistical study of CMEs associated with metric type II bursts. *Geophysical Research Letters*, Volume 30, Number 12, pages 8016–8023, 2003.
- [49] R. Leitinger. Hohe Atmosphäre I. Lecture Notes, Institute for Geophysics, Astrophysics, and Meteorology/University of Graz, 2005.
- [50] J. Li. Shanghai Astronomical Observatory Analysis Center 2004 Annual Report. In D. Behrend and K. Baver (editors), *International VLBI Service for Geodesy and Astrometry 2004 Annual Report*, NASA/TP-2005-212772, pages 243–246, 2005.
- [51] National Aeronautics and Space Administration (NASA). <http://www.nasa.gov>, 2005.
- [52] A.E. Niell. Global mapping functions for the atmosphere delay at radio wavelengths. *Journal of Geophysical Research*, Volume 2, Number 101, pages 3227–3246, 1996.
- [53] National Geophysical Data Center (NGDC) National Oceanic & Atmospheric Administration (NOAA). <http://www.ngdc.noaa.gov/ngdc.html>, 2005.
- [54] Space Physics Interactive Data Resource National Oceanic & Atmospheric Administration (NOAA). <http://spidr.ngdc.noaa.gov/spidr/>, 2005.

- [55] J. Ping, K. Matsumoto, K. Heki, A. Saito, P. Callahan, L. Potts and C.K. Shum. Validation of JASON-1 Nadir Ionosphere TEC Using GEONET. *Marine Geodesy*, Volume 27, Number 3-4, pages 741–752, 2004.
- [56] J. A. Ratcliffe. *An introduction to the ionosphere and magnetosphere*. Cambridge University Press, 1972.
- [57] J.R. Ray and B.E. Corey. Current Precision of VLBI Multi-Band Delay Observables. In *Proceedings AGU Chapman Conference on Geodetic VLBI: Monitoring Global Change, Washington D.C., 22-26 April, 1991*.
- [58] P. H. Reiff. The Sun-Earth Connection. <http://space.rice.edu/IMAGE/livefrom/sunearth.html>, 1999.
- [59] D.H. Roberts, J. Lehar and J.W. Dreher. Time series analysis with CLEAN. I. Derivation of a spectrum. *Astronomical Journal*, Volume 93, pages 968–989, 1987.
- [60] J.D. Scargle. Studies in astronomical time series analysis. II. Statistical aspects of spectral analysis of unevenly spaced data. *The Astrophysical Journal*, Volume 262, pages 835–853, 1982.
- [61] S. Schaer. *Mapping and Predicting the Earth's Ionosphere Using the Global Positioning System*. Ph.D. thesis, Astronomical Institute, University of Berne, 1999.
- [62] S. Schaer. Modified Single-Layer Model Mapping Function. <http://www.aiub.unibe.ch/ionosphere/mslm.pdf>, 2005.
- [63] S. Schaer, W. Gurtner and J. Feltens. Ionex: The ionosphere map exchange format version 1. In *Proceedings of the 1998 IGS Analysis Centers Workshop*, pages 233–248. ESOC, Darmstadt, Germany, 1998.
- [64] M. Schmidt. *Grundprinzipien der Wavelet-Analyse und Anwendungen in der Geodäsie (in German)*. Shaker Verlag, 2001.
- [65] M. Sekido. *Pulsar Astrometry by VLBI*. Ph.D. thesis, Department of Astronomical Science, School of Mathematical and Physical Science, The Graduate University for Advanced Studies, Mitaka, Tokyo, Japan, 2001.
- [66] M. Sekido, R. Ichikawa, H. Osaki, T. Kondo, Y. Koyama, M. Yoshikawa, T. Kato, T. Ichikawa, Y. Murata, T. Ohnishi, W. Cannon, A. Novikoc, M. Berube and the NOZOMI VLBI observation group. Astrometric Observation of Spacecraft with Phase Delay. *IVS NICT Technical Development Center News No. 25, National Institute of Information and Communications Technology*, pages 40–44, 2004.
- [67] M. Sekido, T. Kondo and E. Kawai. Evaluation of GPS-based ionospheric TEC map by comparing with VLBI data. *Radio Science*, Volume 38, Number 4, pages 1069–1076, 2003.
- [68] N.M. Shilo, E.A. Essex and A.M. Breed. Ionospheric scintillation study of the southern high latitude ionosphere. In *WARS'00- Workshop on the Applications of Radio Science*, pages 154–159. National Committee for Radio Science, Australia, 2000.
- [69] M. I. Sitnov, P. N. Guzdar and M. Swisdak. On the formation of a plasma bubble. *Geophysical Research Letters*, Volume 32, 2005.
- [70] O.J. Sovers, J.L. Fanselow and C.S. Jacobs. Astrometry and geodesy with radio interferometry: experiments, models, results. *Rev. Mod. Phys.*, Volume 70, Number 4, pages 1393–1454, 1998.
- [71] J.O. Stenflo. Differential rotation of the sun's magnetic field pattern. *Astronomy and Astrophysics*, Volume 210, pages 403–409, 1989.

- [72] J. Strěštk. Spectrum of geomagnetic activity in the period range 5-60 days: possible lunar influences. *Annales Geophysicae*, Volume 16, pages 804–811, 1998.
- [73] F. Takahashi, T. Kondo, Y. Takahasi and Y. Koyama. *Wave summit Course: Very Long Baseline Interferometer*. Ohmsha Ltd., Tokyo, Japan, 2000.
- [74] Y. Takahashi, S. Hama and T. Kondo. K-3 software system for VLBI and new correlation processing software for K-4 recording system. In *Journal of the Communications Research Laboratory*, Volume 38, pages 481–502. Communications Research Institute, Ministry of Posts and Telecommunications, Tokyo, Japan, 1991.
- [75] A.R. Thompson, J.M. Moran and G.W. Swenson. *Interferometry and Synthesis in Radio Astronomy*. John Wiley & Sons, Inc., 2nd edition, 2001.
- [76] O. Titov, V. Tesmer and J. Boehm. Occam version 5.0 software user guide. Technical Report 7, AUSLIG, 2001.
- [77] S. Todorova, T. Hobiger, R. Weber and H. Schuh. Determination of the Ionosphere with Different Space Geodetic Techniques. In *International Symposium on Modern Technologies, Education and Professional Practice in Geodesy and Related Fields*. Sofia, Bulgaria, 2004.
- [78] B. Tolman, R.B. Harris, T. Gaussiran, D. Munton, J. Little, R. Mach, S. Nelsen, B. Renfro and D. Schlossberg. The GPS Toolkit – Open Source GPS Software. In *Proceedings of the 16th International Technical Meeting of the Satellite Division of the Institute of Navigation (ION GNSS 2004)*, 2004.
- [79] R.C. Walker. Astronomical VLBI: Comparison and Contrast with Geodetic/Astrometric VLBI. In N. R. Vandenberg and K. D. Baver (editors), *International VLBI Service for Geodesy and Astrometry 2000 General Meeting Proceedings*, NASA/CP-2000-209893, pages 42–51, 2000.
- [80] A. R. Whitney, R. Cappallo, W. Aldrich, B. Anderson, A. Bos, J. Casse, J. Goodman, S. Parsley, S. Pogrebenko, R. Schilizzi and D. Smythe. Mark 4 VLBI correlator: Architecture and algorithms. *Radio Science*, Volume 39, 2004.
- [81] A.R. Whitney. *Precision geodesy and astrometry via Very-Long-Baseline Interferometry*. Ph.D. thesis, Department of Electrical Engineering, MIT, 1974.
- [82] U. Wild. *Ionosphere and geodetic satellite systems: permanent GPS tracking data for modeling and monitoring*. Ph.D. thesis, Astronomical Institute, University of Bern, 1994. Geodaetisch - geophysikalische Arbeiten in der Schweiz, Bern, Switzerland, Vol. 48.
- [83] G. Xu. *GPS - Theory, Algorithms and Applications*. Springer, 2003.

Acknowledgments

I am indebted to a great number of people for help and advice throughout my time as a PhD student. Firstly, I would like to thank my supervisor, Prof. Harald Schuh, for providing guidance and support whenever required. He introduced me to the field of VLBI, gave me extra time and freedom to develop my ideas further and let me attend several scientific meetings to share my findings with a broader scientific community. I would like to acknowledge the enthusiastic supervision of Dr. Tetsuro Kondo, who is not only an expert in the field of ionospheric and magnetospheric physics, but also one of the leading developers of VLBI technology. He made my stay in Japan a very pleasant one and always had a spare minute to answer my questions or to help me with daily life problems. Dr. Johannes Böhm deserves credit for reading all sections of my work and giving me useful hints for my work. Several of the ideas presented in this work were discussed first in one of the numerous coffee breaks with colleagues from the Research Group of Advanced Geodesy. Dr. Robert Weber is acknowledged for introducing me to the secrets of GPS and for fruitful discussions about adjustment methods. Thanks Dr. Yasuhiro Koyama, Dr. Ryuichi Ichikawa, Dr. Mamoru Sekido and Dr. Hiroshi Takeuchi for imparting VLBI knowledge and for letting me participate in several office activities. Mrs. Christa Katzenberger (M.A.) is acknowledged for checking the final English version of the manuscript. I would also like to thank all the rest of the academic and support staff at the Institute for Geodesy and Geophysics and the Kashima Space Research Center, National Institute of Information and Communications Technology.

Finally, I am forever indebted to my parents and Ms. Rieko Omi for their understanding, endless patience and encouragement, when it was most required.

I am very grateful to the Austrian Science Fund (FWF), which funded the research project P16136-N06 "Investigation of the ionosphere by geodetic VLBI". Furthermore I want to thank the Japanese Society for the Promotion of Science, JSPS (project PE 04023) for supporting my research. The International VLBI Service for Geodesy and Astrometry (IVS), Kashima Space Research Center, Geographical Survey Institute (Japan) and the International GNSS Service (IGS) are acknowledged for providing data.

FWF Der Wissenschaftsfonds.



寺の門前に住む小僧は、いつの間にか聞き覚えでお経を読むようになる。
(A boy living near a temple gate can read sutras without formal learning.)

Curriculum Vitae

Name: Dipl.-Ing. Thomas Hobiger

Address: Engerthstrasse 161/603, 1020 Vienna, Austria

Date and Place of Birth: May 9th, 1978 in St. Pölten

Nationality: Austria

Employment Record

since 4/2005 Project scientist for the FWF (Austrian Science Fund) project P16136-N06 "Investigation of the ionosphere by geodetic VLBI"

7/2004 - 3/2005 Research fellow (JSPS) at Radio Astronomy Applications Group, Kashima Space Research Center, Japan

3/2003 - 6/2004 Project scientist for the FWF (Austrian Science Fund) project P16136-N06 "Investigation of the ionosphere by geodetic VLBI"

Fellowships - International Experiences

7/2004 - 3/2005 Japan Society for the Promotion of Science (JSPS) fellowship at Radio Astronomy Applications Group, Kashima Space Research Center, Japan

University and education

since 10/2002 Doctoral program, Vienna University of Technology

5/2002 Dipl.-Ing. degree (M.Sc.) with honors from the Vienna University of Technology Thesis title: Determination of Ionospheric Parameters from VLBI observations. Supervisor: Prof. H. Schuh (Vienna University of Technology)

10/1997 - 5/2002 Geodesy and Geophysics, Vienna University of Technology

10/1996 - 5/1997 Compulsory service in the Austrian Armed Forces

9/1988 - 6/1996 High School (Bundesrealgymnasium) in Krems