Numerical Weather Predictions for Network RTK

Anna B. O. Jensen

Publications Series 4, volume 10
National Survey and Cadastre - Denmark
2002

Abstract

When GPS satellite signals are transmitted through the atmosphere they are affected by the media, and in the neutral atmosphere the effect is a function of the meteorological conditions along the signal path. This effect is often referred to as the tropospheric delay.

In the GPS positioning process the tropospheric delay is normally handled by utilizing global tropospheric delay models. For high accuracy differential positioning based on the carrier phase observables, modelling of the error is however not sufficiently accurate, and other techniques must be implemented, especially when working with longer baselines.

In this thesis numerical weather predictions (NWP) are introduced in GPS data processings. NWPs are predictions of the three dimensional meteorological conditions for a given area and point in time, and can as such be used for predicting the tropospheric delay for a satellite signal by integration along the signal path through the information given with the NWP. Since NWPs are predictions they can be used for both post processed and for real time GPS positioning.

The theory and the procedure for determining zenith delays based on NWPs is described, and a number of tests are carried out in order to verify the performance of this novel technique. An initial verification of zenith delays based on a regional NWP show that these do have a better accuracy than the best global tropospheric delay models.

The method, where NWP zenith delays are used instead of global tropospheric delay models, has been tested for static and kinematic post processed GPS positioning, and also for real time kinematic (RTK) positioning based on a network of reference stations. The results show that the use of zenith delays determined from a NWP does have a promising potential for high accuracy GPS positioning, and especially the ambiguity resolution process is improved; the ambiguities are solved faster and more reliably when using the NWP approach.

For signals received at lower elevation angles a ray tracer can be implemented, instead of using standard mapping functions for mapping the NWP zenith delays down to the low elevation angles. This aspect is discussed, and finally a few test results show that the use of a ray tracer might be beneficial for satellite signals received at elevation angles below 10°.

CONTENTS

INTRODUCTION

When using the Global Positioning System (GPS) for differential carrier phase based positioning, the largest error sources are the errors in the satellite positions, and the effects on the signal propagation caused by the ionosphere and the neutral atmosphere.

These errors are spatially correlated, and they are almost negligible for short baselines with less than about 10 km between the reference and the roving GPS receivers. For longer baselines the errors must be considered in the positioning process in order to obtain position accuracies at the cm level, as required for instance for land and hydrographic surveying, construction surveys, or airborne gravity surveys.

The orbit error can be mitigated using predicted or precise orbit information available via the internet. The first order effects of the ionosphere can be eliminated using linear combinations of data collected on both the L1 and L2 GPS frequencies, and thereby the effects of troposphere, the lowest part of the neutral atmosphere, is left as the largest remaining error source.

Traditionally the tropospheric error has be handled by removing the majority of the error using a global tropospheric model followed by double differencing where most of the remaining differential error is removed. Mendes (1999) showed that the best global delay models have an accuracy of about 3 cm in zenith, and Raquet (1998) showed that the residual tropospheric error in zenith after both modelling and double differencing has a size of a few cm depending on baseline length. The size of the residual error is increasing for signals received at lower elevation angles, and since the residual tropospheric errors propagate directly into the position of the roving GPS receiver it can have a large impact on the accuracy of the positions determined.

For high accuracy post processed GPS positioning this residual effect can be estimated as a part of the positioning adjustment process as described for instance by Beutler et al. (2000). This does, however, require that all other error sources are modelled sufficiently well, and this is normally not possible for commercial static or kinematic positioning, neither for "real-time kinematic" positioning (RTK). The need for another method for dealing with the residual tropospheric error for these types of GPS positioning is therefore important.

Network RTK, where the residual tropospheric effect is estimated based on GPS data from a network of reference stations as described for instance by Raquet (1998), has proven to be the solution for RTK positioning. However, the network RTK techniques only work well when the distances between the reference stations are about

50 km or less, because the increasing residual tropospheric errors for longer distances are corrupting the generation of RTK corrections. An improved technique for dealing with the troposphere in connection with network RTK is therefore also important, and it is the main area of interest for this thesis.

For static and kinematic positioning the size of the residual tropospheric error can be minimized by introducing numerical weather predictions (NWPs) in stead of the global tropospheric models. This subject has been treated by a few groups lately. Schueler et al. (2000b) suggest the use of NWPs for WADGPS, and Pany et al. (2001b) discuss a number of methods for using NWPs for static positioning. Tsujii et al. (2001) improved kinematic ambiguity resolution by introducing NWPs, and finally Behrend et al. (2001) obtained improvements in the position domain when using NWPs for single point post processed GPS positioning.

These promising results lead to the idea of introducing NWPs for network RTK. Normally the network based RTK corrections are generated in a processing center, and logistically introduction of NWPs will be straight forward, since the weather data can easily be acquired by the processing center.

The introduction of NWPs in the GPS data processing is however a challenge and it is therefore treated in detail in this thesis, where also a number of tests have been carried out with the purpose of investigating whether any improvement is found in the network RTK positioning process by introducing NWPs.

Since an improved handling of the tropospheric error is also interesting for ordinary static and kinematic GPS positioning, the method has also been tested for these types of positioning.

Structure of the thesis

Initially the Global Positioning System is described with emphasis on its use for carrier phase based positioning. RTK and network RTK is treated in detail in the end of Chapter 1. Chapter 2 is a general introduction to the ionosphere and the neutral atmosphere, and the effects these media exert on GPS signal propagation. Also the measures used for mitigating the effects in the positioning process are discussed. Chapter 3 deals with numerical weather predictions, and the procedure for using NWPs as a basis for determining zenith delays is described. In Chapter 4 the NWP derived zenith delays are tested for static and kinematic GPS positioning as well as for network RTK, and in Chapter 5 a few tests are carried out using a ray tracer for determining slant tropospheric delays through the NWP. The work is put into perspective in Chapter 6 where implementational issues are discussed, followed by a discussion on the possibilities for using NWPs for GPS positioning in the future with the modernized GPS and the new European Galileo system. Finally conclusions and recommendations follow in Chapter 7.

The last part of the thesis contains three appendices. Appendix A is a description of indices for ionospheric activity, and examples of signatures of both high and normal ionospheric conditions are given. Appendix B describes the GPS data used for the tests in Chapter 4, and finally Appendix C deals with some detailed issues related to the determination of tropospheric delays from NWPs.

As network RTK is a relatively new positioning technique, which is expected to play a major role for both positioning and navigation in the future, more emphasis is placed on the network RTK positioning technique than on the traditional static and kinematic techniques. The thesis has been written with the academic GPS community in mind, and the introduction to GPS in Chapter 1 is therefore relatively brief compared to the comprehensive introduction to the atmosphere in Chapter 2.

ACKNOWLEDGMENTS

The research and results described in this thesis are the result of interactions with several persons and institutions, and I would like to express my gratitude:

- To my supervisors C. C. Tscherning, the University of Copenhagen, and J. F. Madsen, KMS - the National Survey and Cadastre, for guidance and encouragement throughout the three years.

- To M. E. Cannon, the University of Calgary, for encouraging me to pursue a Ph.D. and for generous advice during the studies.

- To P. Høgh and M. B. Sørensen at the Danish Meteorological Institute for making the DMI-HIRLAM data available.

- To B. Jonsson, the National Land Survey of Sweden, for making the SWEPOS data available.

- To S. Syndergaard, the Institute of Atmospheric Physics, at the University of Arizona, for making the ROSAP ray tracing code available, and for advice on its use.

- To G. Lachapelle, the University of Calgary, for making the NetCor software available.

- To H. Vedel, the Danish Meteorological Institute, for advice and suggestions with respect to the estimation of zenith delays from NWPs.

- To J. Johansson, B. Stoew, and L. Gradinarsky, Onsala Space Observatory, for an interesting and inspiring visit to the observatory.

- To B. Townsend, Forgis Technologies, for advice and discussions related to network RTK.

- To everybody in the Department of Geomatics Engineering at the University of Calgary for creating a dynamic and inspiring atmosphere that provided an invaluable inspiration for me. My Ph.D. would not have been the same without the visits to the U of C.

- To L. P. Fortes, the University of Calgary, for lots of advice and help, and to P. Alves, also from the U of C, for helping with the NetCor software.
- To G. Xu, GeoForschungsZentrum, Potsdam, for suggestions and ideas that lead to the troposphere being the main area of interest for the thesis.
- To P. Skjellerup, Rambøll, and T. B. Larsen, KMS, for very constructive comments on the draft version of the thesis.
- To all my dear colleagues in the Departments of Geodesy and Geodynamics at KMS for support and interest in my work. Especially, B. Madsen and M. Weber are thanked for helping with the Bernese software, T. Knudsen is thanked for advice on C and Fortran coding problems, A. Olesen is thanked for helping with the geoid data, and J. Høyer is thanked for helping with LATEX. Also thanks to K. Keller for lots of advice and discussions, and for patiently having shared an office with me for several years.
- Finally, to my family for their continuous support and encouragement.

Funding for the Ph.D. was provided by KMS, the National Survey and Cadastre - Denmark, and by the Danish Research Agency.

NOTATION

b	vertical coordinate parameter
B	double difference matrix
c	speed of light
C	covariance matrix
d_{ion}	ionospheric signal delay
dr	inaccuracy in satellite position
dt, dT	satellite and receiver clocks drift
d_{trop}	tropospheric signal delay
e	partial pressure of water vapor
g	gravity
f	frequency
h	ortometric height
IF	ionosphere free linear combination
k_1, k_2, k_3	refractivity constants
l	measurement minus range observation vector
M_d, M_w	molar mass for dry air and water vapor
n	refractive index
n_g, n_Φ	group and phase refractive index
N	ambiguity (initial number of cycles), in Chapters 1 and 4
N	refractivity
N_d, N_w	dry and wet refractivity
N_e	electron density
N_h	hydrostatic refractivity
N_{L1}, N_{L2}	L1 and L2 phase ambiguity
N_{WL}	wide lane phase ambiguity
P	atmospheric pressure
P_d	partial pressure of dry air
q	specific humidity
r	geometric range
R_d, R_w	gas constants for dry air and water vapor

v	speed of signal in a media different than vacuum
W	surface geopotential
WL	wide lane linear combination
T	temperature
X_{sv}, Y_{sv}, Z_{sv}	position of satellite in cartesian coordinates
$X_{rec}, Y_{rec}, Z_{rec}$	position of receiver in cartesian coordinates
Z_d, Z_w	compressibility factors for dry air and water vapor
zhd	tropospheric zenith hydrostatic delay
ztd	tropospheric zenith total delay
zwd	tropospheric zenith wet delay
α_T, α_q	lapse rate for temperature and specific humidity
δh	change in height
δl	vector with NetAdjust corrections
δT	change in temperature
δq	change in specific humidity
ϵ	ratio between R_d and R_w
λ	wave length
ν	multipath and receiver noise
ϕ	latitude
Φ	carrier phase observation
ρ	density
σ	sigma coordinates used for vertical reference in NWP
σ_{L1}, σ_{L2}	standard deviation of L1 and L2 phase observations
σ_{WL}	standard deviation of wide lane
τ	signal travel time in time units
$\triangle$	single difference
$\nabla\triangle$	double difference

1

Relative GPS Positioning

1.1 Introduction to GPS positioning

NAVSTAR, the Global Positioning System, is a system that can be used continuously for positioning, navigation, and timing purposes all over the world. GPS was developed and is maintained by the US Department of Defense, and was designed as a purely military system. Experiments quickly showed, however, that there was a large potential for civil applications of the system, and it is today being used extensively for non-military purposes. GPS will be going through a modernization phase in the coming years, but the current configuration is designed to consist of 24 satellites, orbiting the Earth at an altitude of approximately 20 000 km. The satellites are distributed into six orbital planes having an inclination angle of approximately 55°.

Satellite signals

The satellites are transmitting electromagnetic waves at two frequencies:

- L1: 1575.42 MHz
- L2: 1227.60 MHz

Three different codes are modulated onto the carrier waves. The P- and C/A-codes (for Precision and Coarse / Acquisition, respectively) are pseudo random sequences of 0's and 1's that are unique for each satellite. The codes do not contain any information, but the binary pattern, that must be known by the GPS receiver, makes it possible for the receiver to determine the signal transit time. The transit time can be converted to ranging information which is used in the positioning process. The P-code is encrypted and is only directly available to military users. The P-code is modulated onto both L1 and L2, whereas the C/A-code is only modulated onto L1.

Each code is combined with the "navigation message" before being modulated onto the carrier waves. The navigation message contains information such as the status of the satellites, the clock drift, and the predicted orbits of the satellites. This is the so-called broadcast ephemeris.

The control system

Monitoring of the system is carried out by the GPS Control Centers. Based on continuous logging of satellite data, the orbits of the satellites are predicted, and the satellite clock drift is monitored and modelled. This information is uploaded to the satellites once per day (Misra and Enge, 2001), and is then transmitted to the users as a part of the navigation message.

Absolute positioning

When using GPS for positioning the basic principle is to determine distances between the GPS receiver and the GPS satellites. Once the distances are determined, the position of the receiver can be calculated from the "known" positions of the satellites as given by the broadcast ephemeris.

This is illustrated with Equation (1.1) where *dist* is the distance between the receiver and satellite, X_{sv}, Y_{sv}, Z_{sv} is the position of the satellite, $X_{rec}, Y_{rec}, Z_{rec}$ is the position of the receiver, dT is the receiver clock drift, which is explained below, and c is the speed of light:

$$dist = \sqrt{(X_{sv} - X_{rec})^2 + (Y_{sv} - Y_{rec})^2 + (Z_{sv} - Z_{rec})^2} + c \cdot dT \tag{1.1}$$

With four equations, i.e. distances to four different satellites, a set of equations can be established for every time epoch and the four unknowns (receiver coordinates and clock drift) can be estimated. In practice up to 12 satellites are observed at the same time whereby it is possible to perform a least squares adjustment after linearization of the observation equations. This is described for instance by Strang and Borre (1997) and Blewitt (1997).

According to the specifications from DoD (2001), GPS can be used for absolute positioning with accuracies within 13 meters in the horizontal and 22 meters in the vertical (95% probability). Noise from the GPS receiver itself and from the environment surrounding the GPS antenna can degrade this positioning performance.

Differential positioning

To obtain positions with better accuracies, a relative or differential technique must be used where at least two GPS receivers are employed. One receiver is located at a reference point with known coordinates and the other receiver, the rover, is located at a point (or moving between points) with unknown coordinates. The position of the rover is then determined relative to the reference station.

This can be carried out in either real-time, utilizing a data link, or in a post processing mode where data is logged by the receivers and processed on a computer afterwards.

GPS positioning can be based on code observations only, where the receiver–satellite distances are determined from the C/A-code, or it can be based on a combination of code and carrier–phase observations. Further GPS positioning can be based on L1 observations only, or on both L1 and L2 observations. This thesis deals with high accuracy positioning that requires both code and phase data from dual frequency GPS receivers.

1.1.1 Carrier–phase observations and error sources

In carrier–phase based GPS positioning, the receiver–satellite distances are determined by the carrier wave. In order to use the carrier it must be recovered from all the noise received with the GPS signals, and also the modulated codes must be removed before the carrier can be tracked.

The carrier–phase on L1 is tracked using the C/A code and a code-correlation technique, where the received C/A code is correlated with a receiver generated replica of the code. When the highest correlation occur, timing information can be extracted from the C/A code, an approximate absolute position can be determined, and the pure carrier wave can be recovered by removing the C/A code and the navigation message.

On L2 the P-code is encrypted so code-correlation is not possible for civil GPS receivers. Instead a codeless techniques can be used where L2 is recovered by cross correlation with the L1 carrier. The alternative is a semi-codeless technique that is based on a priori knowledge about the encrypted P-code e.g. that the encryption is identical on L1 and L2. The disadvantage with both these techniques is that the tracking noise is increased, and carrier–phase observations on L2 are thus affected by a higher noise level than L1 observations.

The actual carrier–phase measurement is the difference between the received and recovered carrier wave, and a receiver generated copy of the signal. The distance to the satellite can be determined as the number of full cycles between the satellite and the receiver, plus the measured fractional part of a cycle.

When the receiver has locked on to a satellite signal it starts counting the full number of cycles received from the satellite. The carrier–phase observation equation below shows the relation between the measured phase (fractional part and counted number of full cycles), Φ, and the wanted geometric range to the satellite.

$$\Phi = r + dr - d_{ion} + d_{trop} + c(dt - dT) + \lambda N + \nu \tag{1.2}$$

where: r is the geometric range (receiver–satellite) in meters, dr is the inaccuracy in satellite position in meters, d_{ion} is the ionospheric signal delay in meters, d_{trop} is the tropospheric signal delay in meters, c is speed of light given in meter/sec, dt and dT are the errors in satellite and receiver clocks respectively, both given in seconds, λ is the wavelength (approximately 0.19 meters for L1, and 0.24 meters for L2), N is the unitless ambiguity (the initial number of cycles), and ν is multipath and receiver noise also given in meters (Lachapelle, 1993).

The ambiguity is the initial number of cycles, i.e. the number of cycles between the satellite and receiver at the initial measurement epoch before the receiver has started counting the number of cycles. The ambiguity can not, however, be converted to the receiver–satellite distance by just multiplying with the wavelength. This is because the ambiguity also contains a receiver dependent oscillator offset (Blewitt, 1997), and the N in the above equation is therefore not necessarily a positive integer value, as it intuitively would look like. The challenge in connection with carrier–phase based positioning is to solve this ambiguous number of cycles, so the distance to the satellites can be determined.

Equation (1.2) indicates that the geometric distance, the ambiguities, and all the errors in the system add up to the observed phase, considering the sign of the various factors. If the influence of the errors can be minimized it will be easier to determine the ambiguity and thereby also the distance to the satellite.

The error sources are briefly described below:

Satellite position The satellite positions given by the broadcast ephemeris are predicted with a 3D accuracy of about 3 meters (Misra and Enge, 2001). For applications with high accuracy demands, precise post–processed orbits, based on data from numerous permanent GPS stations, are available about two weeks after the data was collected. Precise orbits can be obtained for instance from the International GPS Service (IGS), and these orbits have an estimated accuracy better than 5 cm (Weber et al., 2002).

Ionosphere When the satellite signals are transmitted through the ionosphere their speed of propagation is affected by electrons along the signal path. This causes a time delay that is normally about 1–15 meters in the zenith, when converted to range units (Misra and Enge, 2001). The delay can, however, be much larger during periods with high ionospheric activity. The ionosphere is dispersive for radio waves so the refractive index is a function of frequency, and thus different for observations on L1 and L2 (Langley, 1996). The ionospheric effect is described in detail in Chapter 2.

Troposphere When the satellite signals are transmitted through the troposphere they are refracted, and the refraction is a function of the meteorological conditions. The refraction causes a signal delay that has a size of 2–3 meters in the zenith depending on meteorological conditions and the altitude of the GPS receiver. The tropospheric delay is treated in detail in Chapter 2.

Clock errors The atomic rubidium and cesium clocks in the GPS satellites, and the clocks in the GPS receivers, are referenced to GPS time. No clock is perfect, however, and they all drift from the reference time. GPS receivers often have quartz crystal oscillators where the short term drift (over one second) is approximately $10^{-9} - 10^{-11}$ seconds per second and the long term drift (over a day) is 2–3 orders of magnitude larger (Misra and Enge, 2001). The short term drift for satellite clocks is about the same size, but the long term drift is smaller, about $10^{-12} - 10^{-13}$ seconds per second (Misra and Enge, 2001). The clock drifts must be taken into consideration in the positioning process since even a small clock error can cause a significant range error, due to the short transmission time (about 0.07 seconds) of the signals.

The satellite clock drift is modelled by polynomials, and polynomial coefficients are given in the navigation message. The receiver clock drift and any unmodelled part of the satellite clock drift, can be estimated through the positioning process.

Multipath arises when the spread spectrum GPS signal follows multiple paths to the receiver, for instance if parts of the signal are reflected by objects in the environment of the GPS antenna.

Signals affected by multipath are delayed and the signal strength is lower than for non-reflected signals. Most modern GPS antennas and receivers are built to mitigate the influence of multipath by tracking changes in the signal strength. The phase measurement error due to multipath can be no more than a quarter wavelength (Misra and Enge, 2001), which is about 5-6 cm.

Receiver noise is caused by the activities within the GPS receiver and antenna, and is different for all receivers and antennas. The error due to receiver noise will increase relatively with a lower signal strength, and is thus larger for signals received at low elevation angles (Misra and Enge, 2001). The size of the carrier–phase receiver noise can be investigated for any GPS receiver by performing a zero-baseline test. This was for example done by Raquet (1998) who used Ashtech Z-XII receivers and found L1 phase noise of about 0.3 mm for high elevation satellites, and of about 0.8 mm for satellite signals received at a 10° elevation angle. The noise was slightly higher for the L2 data.

Finally the “anthroposphere”, i.e. the GPS users who handle the GPS equipment, should also be considered a major error source. Typical errors in this category are erroneous recording of antenna heights, or the use of wrong coordinates for reference stations. These errors are external to the GPS system and the equipment, but nevertheless very common.

1.1.2 Double differences

The largest errors in carrier–phase based differential positioning are the inaccuracy in the satellite position and the effects of the atmosphere. These errors are spatially correlated, so when the distance between the two receivers increases, the influence of the errors is also increased. If the two receivers are located close to each other (closer than about 15 km (Goad, 1996)) the influence of the atmospheric error will be almost the same for the two receivers since the signals are transmitted through basically the same parts of the atmosphere. The double differencing technique, which is described in the following, takes advantage of this when used in the positioning process.

For differential positioning, observations from at least two receivers and four satellites at the same epoch in time must be available. The observation equation (1.2) can be generated for each receiver–satellite combination, and the observation equations can then be differenced. The first step is to generate a single difference equation, i.e. the difference between observation equations for two receivers, A and B, observing the same satellite, I:

$$\begin{aligned} \triangle\Phi^I_{AB} &= \Phi^I_A - \Phi^I_B \\ &= \triangle r^I_{AB} + \triangle dr^I_{AB} - \triangle d^I_{ion_{AB}} + \triangle d^I_{trop_{AB}} + c(dT_B - dT_A) + \lambda\triangle N^I_{AB} + \triangle\nu^I_{AB} \end{aligned} \tag{1.3}$$

The triangle indicates single difference, the subscript denotes receivers, and the super script denotes satellites. With single differencing the satellite clock error cancels out, because it will have the same effect for signals received in both receivers. The influence of the spatially correlated errors is also reduced as mentioned above.

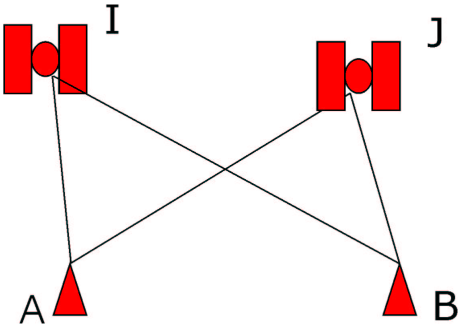

Figure 1.1: Double difference with satellite I and J, and receivers A and B

The next step is to generate the difference between two single differences, i.e. a double difference, where observations from a second satellite, J are introduced (see Figure 1.1):

$$\nabla\triangle\Phi^{IJ}_{AB} = \triangle\Phi^I_{AB} - \triangle\Phi^J_{AB} \tag{1.4}$$

$$
\begin{aligned}
&= (\triangle r^I_{AB} + \triangle dr^I_{AB} - \triangle d^I_{ion_{AB}} + \triangle d^I_{trop_{AB}} + c(dT_B - dT_A) + \lambda \triangle N^I_{AB} + \triangle \nu^I_{AB}) \\
&\quad -(\triangle r^J_{AB} + \triangle dr^J_{AB} - \triangle d^J_{ion_{AB}} + \triangle d^J_{trop_{AB}} + c(dT_B - dT_A) + \lambda \triangle N^J_{AB} + \triangle \nu^J_{AB}) \\
&= \nabla\triangle r^{IJ}_{AB} + \nabla\triangle dr^{IJ}_{AB} - \nabla\triangle d^{IJ}_{ion_{AB}} + \nabla\triangle d^{IJ}_{trop_{AB}} + \lambda\nabla\triangle N^{IJ}_{AB} + \nabla\triangle \nu^{IJ}_{AB}
\end{aligned}
$$

Now also the receiver clock errors cancel out, and the size of dr, d_{ion} and d_{trop} is further reduced. For baselines shorter than about 15 km these errors generally cancel out.

Omitting the receiver and satellite identifiers, the final double difference expression becomes:

$$\nabla\triangle\Phi = \nabla\triangle r + \nabla\triangle dr - \nabla\triangle d_{ion} + \nabla\triangle d_{trop} + \lambda\nabla\triangle N + \nabla\triangle\nu \tag{1.5}$$

The double differenced ambiguity is an integer number since the initial receiver phase offset has been eliminated through the double differencing process. In order to obtain position accuracies of a few cm or mm, the correct integer number for the double differenced ambiguity must be determined. This can be a complicated task, but several ambiguity resolution techniques do exist, see for instance Blewitt (1989) for very long baselines, Frei and Beutler (1990) for the FARA method - Fast Ambiguity Resolution Approach, Teunissen and Tiberius (1994) for the LAMBDA method - Least Squares Ambiguity Decorrelation Adjustment, or see Han and Rizos (1997) for a discussion of different techniques.

Hatch and Euler (1994) made an evaluation of different techniques for "on the fly" (OTF) ambiguity resolution, where the ambiguities are solved while the rover is moving.

Validation of the ambiguities after resolution is also of importance and the subject has been treated by e.g. Joosten and Tiberius (2000) where "the ambiguity success rate" is described as a measure of the probability of a correct solution for the ambiguities.

When the correct integer numbers for the ambiguities have been determined, the ambiguities are said to be fixed. If the ambiguity resolution process is not successful the ambiguities might be fixed to wrong integer numbers, which in general is difficult to discover once it happens in the positioning process (Jagieniak et al., 2000). The ambiguities can also be retained as real numbers, whereby they are said to be floating since the constraint introduced by fixing them to integer numbers, is removed.

A position determined on float ambiguities is very unreliable since all unmodelled errors in the positioning process will propagate into the floating ambiguities. However, if a satellite is observed without interruptions for longer time intervals, 30–60 minutes, if the noise level in general is low, and if no sudden changes in the receiver–satellite geometry occur, then the floating ambiguities will converge towards the correct integer number, and the accuracy of the resulting position can be as good as a few cm. This is for instance shown by Jensen and Cannon (2000).

1.1.3 Linear combinations of the observables

Based on double differences, a number of linear combinations of the phase observables can be generated. These can be helpful in the ambiguity resolution process for baselines longer than 15 km, where the influence of the spatially correlated errors is significant.

The widelane, WL, is the difference between the L1 and L2 observables:

$$\Phi_{WL} = \Phi_{L1} - \Phi_{L2} \quad \Rightarrow \quad \nabla\triangle N_{WL} = \nabla\triangle N_{L1} - \nabla\triangle N_{L2} \tag{1.6}$$

The advantage of the widelane observable is the longer wavelength (0.86 meters) that makes it easier to solve the ambiguities, $\nabla\triangle N_{WL}$.

Another combination of the L1 and L2 observables is the ionosphere–free (IF) combination:

$$\Phi_{IF} = \Phi_{L1} - \frac{f_2}{f_1}\Phi_{L2} \quad \Rightarrow \quad \nabla\triangle N_{IF} = \nabla\triangle N_{L1} - \frac{f_2}{f_1}\nabla\triangle N_{L2} \tag{1.7}$$

The advantage of the ionosphere–free combination is that the first order ionospheric effects are removed. The disadvantage is that the double difference ambiguity is not an integer, so it is more sensitive to noise, and it is difficult to test whether it is solved correctly. Odijk et al. (2002) show, however, that it is possible to rewrite Equation (1.7) to obtain an expression with an integer property of the ionosphere–free double differenced ambiguities. The IF combination is further discussed in Chapter 2.

When using the Φ_{IF} most of the ionospheric errors are removed, but higher order effects of the ionosphere are still present. For baselines longer than 50–100 km these higher order effects must be taken into consideration in the data processing (Brunner and Gu, 1991). During high ionospheric activity the higher order effects will also affect the shorter baselines. This is further discussed in Chapter 2

Φ_{WL} and Φ_{IF} can be useful in the ambiguity resolution, where $\nabla\triangle N_{L1}$ and $\nabla\triangle N_{L2}$ can be resolved easier after $\nabla\triangle N_{WL}$ and $\nabla\triangle N_{IF}$ have been determined. An example of this is given by Raquet (1998).

Other linear combinations of the phase observables do exist, see for instance Seeber (1993). A major disadvantage of these linear combinations is that the noise, originating from the L1 and L2 observations, is amplified by the linear combinations. For instance the standard deviation for the wide lane, σ_{WL}, is given as (Lachapelle, 1993):

$$\sigma_{WL} = \sqrt{\sigma_{L1}^2 + \sigma_{L2}^2} \tag{1.8}$$

σ_{L1} and σ_{L2} are the standard deviations for the phase observations of L1 and L2 respectively.

1.2 Reference Systems

Several reference systems are being used for GPS positioning, and it is important to be consistent in the selection of a reference system. When using differential techniques, the position of the rover will be given in a reference system that is a combination of the systems used for the satellite orbits and for the reference station coordinates. If the orbits are given in one reference system and the reference station coordinates in another, a geometrical misalignment will be introduced in the data processing.

The following sections describe two of the reference systems commonly used for GPS positioning.

ITRF

The International Earth Rotation Service (IERS) is in charge of providing realizations of worldwide celestial and terrestrial reference systems, for instance the International

Terrestrial Reference System (ITRS), which is considered to be the best reference system defined to date. The ITRS is realized in the form of reference frames (ITRF), and new reference frames are defined when it is considered necessary because of the dynamic nature of the Earth. The latest realization is the ITRF2000, and it is described on http://lareg.ensg.ign.fr/ITRF/ITRF2000/.

When GPS is used for geodetic and geodynamic purposes, the ITRF is most commonly used as the reference frame. By using precise orbits given in ITRF, and reference station coordinates also given in the ITRF, at the same epoch in time, geometrical misalignments can be eliminated in the positioning. The position of the rover will be free of any errors originating from reference frame mix-ups, and sub-cm accuracy can be obtained, if the other error sources are handled with care.

WGS84

The World Geodetic System 1984 is a conventional terrestrial reference system defined to be used in connection with GPS. The definition and realization of WGS84 is described by NIMA (2000).

WGS84 was originally realized, i.e. coordinates for a number of physical reference stations were determined, using the TRANSIT system in 1987. The newest realization of WGS84, called WGS84(G1150), is from 2001. It is based on data from a number of permanent GPS stations, including some IGS stations. In the data processing the IGS stations were kept fixed at their ITRF2000 coordinates, whereby WGS84 became close to coincident with the ITRF2000 (Merrigan et al., 2002).

WGS84 was defined as the reference system to be used in connection with GPS, and the satellite coordinates determined from the broadcast ephemeris are given in WGS84. When performing absolute GPS single point positioning using broadcast ephemeris the position will therefore also be given in WGS84.

1.3 Real-time Kinematic (RTK)

Real-time kinematic (RTK) is a differential GPS positioning technique where positions are determined with a few cm's accuracy. The technique is now widely used for a range of applications like land surveying, marine surveying, precision farming, and precise navigation etc.

To obtain positions in real-time, the reference receiver and the roving receiver are connected by a data link, so observations can be transmitted from the reference to the rover. The observations are commonly transmitted using either (or both) a proprietary format or the RTCM[1] format which is the de facto standard being used world wide for both code and carrier–phase based differential real-time positioning.

For RTK positioning the ambiguities for the baseline are solved "on the fly" (OTF) while the rover is moving to the first survey location. The ambiguity resolution is referred to as the "initialization" in RTK terminology.

For OTF ambiguity resolution, the widelane observable is often used to quickly resolve the ambiguities, but since the Φ_{WL} is affected by a higher noise level than the L1 and L2 observables, most RTK systems switch to the L1 observable after initialization in order to obtain a higher positioning accuracy (Cannon, 1997). Estimation

---

[1] The Radio Technical Commission for Marine Services (RTCM) is a non-profit corporation dealing with various aspects related to marine radio communication, radio navigation etc. Special Committee no. 104 deals with standards for radio navigation.

of the L1 and L2 ambiguities, once the widelane ambiguities are determined, is described by Misra and Enge (2001). The success of OTF ambiguity resolution is highly dependent on the satellite geometry, and the ambiguities are solved faster when more satellites are in view.

RTK is based on double differencing, and it is anticipated that the spatially correlated errors cancel out during the double differencing process. This assumption generally works well when the distance between the reference and the rover is less than 15 km. But as the baseline length increases, the residual errors will also increase, and the accuracy of the position is degraded as problems with the ambiguity resolution arise.

For distances of about 20–30 km more observations are necessary to reliably solve the ambiguities (Cannon, 1997). If too many data epochs are required the system can no longer be said to work in real-time.

A few RTK techniques do work with an acceptable performance for baselines up to 40 km. Lutz and Gounon (2001) describe a system based on floating ambiguities. First the float wide lane ambiguities are estimated, and then the float L1 ambiguities are estimated using the ionosphere–free linear combination (Equation 1.7). Position accuracies with a 3D standard deviation of 4 cm, have been obtained for a 36 km baseline. The initialization time for the tests had a mean of 77 seconds and a standard deviation of 113 seconds.

Performance of commercial RTK systems is often evaluated in terms of initialization time and reliability of the positions, i.e. how often the accuracy of the position is within a given limit, for instance 1–5 cm. In order to provide a reliable RTK service within a large area, several reference stations are necessary, since distances longer than 10–15 km to the nearest reference station should be avoided. This would make a state- or nationwide RTK system very expensive in terms of hardware. During the last 5–6 years a number of “network RTK” algorithms and systems have therefore been developed.

1.4 Network RTK

In a network RTK system, GPS observations from multiple reference stations are gathered and processed in a common adjustment process. The output is a set of corrections that can be used to correct for the spatially correlated errors affecting the code and phase observations in a roving GPS receiver located within the area encompassed by the reference stations.

The corrections are distributed to the rover, and applied to the raw code and phase observations, before the position is determined. The influence of the satellite position error and the atmospheric errors will be reduced, and the result is improved positioning performance.

Several network RTK methods have been developed during the last years, see for instance Wübbena et al. (1996), Han and Rizos (1996), Wanninger (1997), Raquet (1998), and Vollath et al. (2000). A discussion of these and other methods is given by Fotopoulos and Cannon (2001).

Generally the advantages of using a Network RTK as opposed to traditional RTK can be listed as follows:

Smaller effect of spatially correlated errors: The effect of the spatially correlated errors is reduced in the positioning process. Observations from the same satellites are available for the processing center from a number of reference stations distributed throughout an area. This gives a better possibility for modelling the spatial and temporal variation of the errors over the area.

Improved ambiguity resolution: The ambiguity resolution process is improved, because the effect of the residual errors is reduced, so the correct integer value for the ambiguities are found more easy. The entire ambiguity resolution process is thus faster and more reliable.

Improved positioning performance: With a faster and more reliable ambiguity resolution, the position of the rover is also determined faster and more reliably. Further if the user is performing RTK surveys that are stretching over longer distances, a "jump" in the positions determined will not be experienced when changing reference stations en route. Jumps in the position accuracy are otherwise often seen for traditional RTK, when changing from one reference station to another.

Increased distance to reference station: It is possible to increase the distance between rover and reference station keeping the position accuracy at a consistent level. When the spatially correlated errors are modelled better, the ambiguity resolution can be performed over longer distances, whereby cm-level accuracies can be obtained over distances longer than 20–30 km.

System surveillance: It is possible to perform integrity monitoring, quality control, and system surveillance. This is of major importance for operation and maintenance of a network RTK system. If for instance one reference station is affected by power failure, an emergency procedure can be initiated. Meanwhile, the user will experience only a slight degradation in performance since the rest of the reference stations will continuously provide corrections.

A few commercial network RTK systems are available from Trimble/TerraSat, and GEO++, both located in Germany, and from The University of Calgary's University Technologies International/Roberton Enterprises in Canada. The systems have been implemented in several places throughout the world, and they generally work well with distances of about 50 km between the reference stations.

The RTCM format is also being used for network RTK. It is however not ideal, and the Special Committee No. 104 is working on a new format that will be more convenient for network RTK (Euler et al., 2001).

1.4.1 Virtual reference stations

The Virtual Reference Station (VRS) concept is described in detail by Marel (1998) in connection with post-processing of GPS observations. He describes how observations from a network of reference stations in Holland is used to generate observations from a virtual, non-existent, reference station. The VRS-data is then provided to GPS users in need of data from a reference station close to the area where they have been collecting data with a roving GPS receiver.

In order to process the data using standard GPS processing software, the VRS-data is generated to resemble data from a real reference station as closely as possible. The atmospheric errors, estimated from the network of reference stations, or from a regional model, are therefore included in the data, but site specific errors such as multipath are not present.

When virtual reference stations are used for network RTK, a preliminary position for the rover is transmitted to the processing center where the VRS-data is generated in real-time based on data from the reference network. The VRS-data is determined for a location close to the roving receiver, and it is then transmitted to the rover, and is used in the positioning process, just as data from a real reference station is used for traditional RTK.

The use of virtual reference stations for network RTK is described by Wanninger (1997), and the concept is being used by the three commercially available network RTK packages.

1.4.2 Coordinate conversions

When working with network RTK the selection of the reference system is an issue of importance. All reference stations must be coordinated with high accuracy, preferably in the same reference system as used for the satellite positions, because any inaccuracy in the reference station coordinates will propagate into the RTK corrections and thereby into the position of the rover. If a network RTK system is operational over several years, geodynamic activities might induce the need to recompute the coordinates for the reference stations, in order to keep the positions as accurate as possible.

If the reference stations are located in an area that is affected by known local or regional surface displacements, it is important to be aware of this. In the Northern parts of Sweden the Fennoscandian post glacial rebound is causing a land uplift of more than 1 cm per year. The rebound in the Southern parts of Sweden is only about 1 mm per year (Johansson and Scherneck, 1997). If a national network RTK system is established in Sweden this effect will soon be causing geometrical misalignments in the network adjustment processes.

Dynamical coordinates for the reference stations will affect the coordinates of the user positions, and the majority of RTK users will not be able to handle dynamical coordinates, causing different coordinates for a site that is revisited e.g. every year. In order to keep the user coordinates in a stable reference frame, a coordinate conversion has to be carried out somewhere in the system.

Normally this coordinate conversion is carried out by the user or in the software of the roving GPS receiver. Another solution is to use a virtual reference station for the coordinate conversion. If the VRS is located very close to the rover, the position of the VRS can be given in the GPS concordant reference system used for surveying and mapping, for instance the European EUREF89 (ETRF), without significantly degrading the RTK positioning performance. In this way the coordinate conversion is encapsulated into the VRS-data, and the user will not notice that the coordinates of the reference stations are dynamic. This procedure is suggested and described by Townsend and Jensen (2001).

1.4.3 NetAdjust

One example of a network RTK algorithm is the NetAdjust method. This method is described in detail, because it is being used for testing purposes in Chapter 4

The NetAdjust method was developed by Raquet (1998) as a part of his Ph.D. thesis. The method can be summarized by the following two equations that show how to generate corrections for the observations in the reference stations, δl_n and in the rover, δl_r:

$$\delta l_n = C_{\delta ln} B_n^T (B_n C_{\delta ln} B_n^T)^{-1} (B_n l_n - \lambda \nabla \triangle N_n) \tag{1.9}$$

$$\delta l_r = C_{\delta lr,\delta ln} B_n^T (B_n C_{\delta ln} B_n^T)^{-1} (B_n l_n - \lambda \nabla \triangle N_n) \tag{1.10}$$

where δl_n are estimated corrections for the observations collected in reference stations, δl_r are estimated corrections for observations collected in the rover, $C_{\delta ln}$ is the covariance matrix for the observations in the network, $C_{\delta lr,\delta ln}$ is the cross covariance matrix between network and rover observations, B_n is the double difference matrix (consisting of +1, -1, 0 to indicate which receivers and satellites that are combined to form the double differences), l_n are the measurement minus range observations from the network, λ is the wavelength, and $\nabla \triangle N_n$ are the double difference ambiguities. Subscript n denotes network, and subscript r denotes rover.

Note that the observables in these equations are measurement minus range observations. This means that the geometric distances between the satellites and the known reference station positions are subtracted from the raw code and phase observations. The elements in the covariance and cross covariance matrices are determined by a covariance function described by Raquet (1998) and Raquet et al. (1998a).

The last part in the two Equations (1.9) and (1.10):

$$B_n l_n - \lambda \nabla \triangle N_n \tag{1.11}$$

are the double difference misclosures. The misclosures are the difference between the double differenced measurement minus range observations ($B_n l_n$), and the double differenced ambiguities ($\nabla \triangle N_n$ multiplied by wavelength to obtain distance units instead of cycles).

Since the observations at this point have been corrected for the major part of the atmospheric effects by ionospheric and tropospheric modelling, the misclosure is a measure of the residual errors left in the processing after modelling and after double differencing. The misclosures thus indicate the level of unmodelled errors in the observations, and this is exactly the errors that the RTK corrections determined with the NetAdjust method should be accounting for in the roving receiver.

The misclosures give the size of the residual spatially correlated errors (orbit and atmospheric errors) for the reference stations. With the covariance function described by Raquet (1998), and thereby with $C_{\delta ln}$ and $C_{\delta lr,\delta ln}$, these errors are mapped over the area covered by the reference stations. So when the first part of Equations (1.9) and (1.10) is multiplied with the misclosures, an estimate of the residual errors for the site of the rover is determined.

The locations of the rover and the reference stations are encapsulated into the covariance function, which is a function of not only the stations coordinates, but also of the variances of both the spatially correlated errors and the site specific errors (multipath and receiver noise). These error variances can be determined based on a set of GPS data from the reference stations, and for the roving RTK receiver the error variances are set to nominal values based on a priori knowledge of, for instance, the receiver and antenna types (Raquet, 1998).

The elements in B, l, and N in Equations (1.9) and (1.10) are dependent only on the geometry of the reference network, and on observations from the reference stations. This means that corrections for the rover can be determined without the observations from the rover. The method does, however, require a preliminary position for the rover. In real-time this can be handled either by letting the rover transmit a code based DGPS position to the nearest reference station, or by generating corrections for a grid of points, and then let the rover perform an interpolation in the grid to obtain corrections for its own position, or simply use the corrections given for the nearest grid point.

The corrections generated by the NetAdjust method are applied to the raw code and phase data from the reference and roving receivers, and an improved position can then be determined using a traditional RTK positioning algorithm.

The performance of the NetAdjust method has been shown in a number of papers, for instance by Raquet et al. (1998b) and Townsend et al. (1999) where the method was tested using data from a Norweigian reference network. Fortes et al. (1999) tested the method was using data from Canadian reference stations, and Fortes et al. (2000) used data from reference stations in Brazil.

NetAdjust and floating ambiguities

One of the basic assumptions for NetAdjust is that the double difference ambiguities in the ambiguity matrix, $\nabla\triangle N_n$, have been fixed to the correct integer values. It can, however, be difficult to fix the ambiguities to the correct integer values in real-time if there is not enough data available, if the baseline between the two receivers is too long, if the residual errors are too large, etc.

The stability of the NetAdjust method in a situation where the ambiguities are not fixed but are left floating at real values has been investigated by Jensen (1999). The tests, that were based on 24 hours of data, showed that RTK positions can be determined with a 3D RMS of 4 cm for a 68 km baseline when using the NetAdjust method and a network of seven reference stations in Sweden. This level of accuracy was obtained when the ambiguities for the baselines in the reference network were either float or fixed, and when these were combined with either float or fixed ambiguities for the reference–rover baseline. See also (Jensen and Cannon, 2000).

2

The Atmosphere

When satellite signals are transmitted through the various layers of the atmosphere they are affected by the media, and the main disturbances of the signals happen in the ionosphere and the neutral atmosphere. These two parts of the atmosphere, and their influence on satellite signals, are described in the following chapter. The chapter is not a complete description of the activities in the regions as only the factors with importance for GPS users are described.

All latitudes mentioned in relation to the ionosphere are magnetic latitudes. A procedure for conversion between geographic and magnetic latitudes is given by Skone (1998).

2.1 The Ionosphere

The ionosphere is the ionized part of the atmosphere. It stretches from approximately 60 to 1000 km above the Earth's surface, and it contains a significant number of ionized particles, i.e. free electrons and positively charged ions. The media as a whole is neutral, since there is an equal amount of particles with negative and positive charge. The electrons are the particles that exert the main influence on the propagation of electromagnetic waves through the ionosphere (Hargreaves, 1992).

Electromagnetic waves can only be transmitted through the ionosphere if the frequency is higher than 30 MHz (Misra and Enge, 2001). At lower frequencies the signals are reflected by the ionosphere, and this was utilized extensively for radio communication and navigation before satellite communication. Radio signals were transmitted over large distances, e.g. between continents, by “zigzagging” between the ionosphere and the surface of the Earth.

In early studies of the ionosphere, electromagnetic waves at different frequencies were reflected at different heights of the ionosphere, and that led to a division of the ionosphere into four regions, the D, E, F1, and F2 regions, defined by the electron density (Campbell, 1997).

The D region is the part that is located below 90 km altitude. The electron density is $10^8 - 10^{10}$ electrons/m^3 Normally the D region has very little effect on GPS signals (Klobuchar, 1996).

The E region is located between 105 and 160 km above the Earth's surface, and the electron density is more than 10^{11} electrons/m^3. Under normal conditions this region does not have much effect on GPS signals (Johansson, 1997). The region is, however, characterized by the presence of irregularities in election density at high latitudes in the auroral region, and these irregularities can affect the signal propagation significantly during times with much auroral activity (Skone, 1998).

The F regions are the regions with the highest electron densities, and further the F1 region has the highest electron production rate (Johansson, 1997). The F1 region is located between 160 to 180 km altitude, and the F2 region is located between 180 and 1000 km. For the F1 region the electron density is $10^{11} - 10^{12}$ electrons/m^3 , and for the F2 region the density is 10^{12} electrons/m^3. The maximum intensity of electrons is normally located around 250–350 km above the Earth's surface (Skone, 1998), and the largest effect on the GPS signals thus happen in the F2 region.

The electron densities and altitude limits are from Hargreaves (1992).

The main effects on a GPS signal propagated through the ionosphere are the group delay, the phase advance, and scintillation, and all three effects are related to the electron density in the ionosphere. The effects are described in Section 2.2, following a description of the effects that are causing the variation in the electron density.

Total Electron Content

The Total Electron Content (TEC) is the number of electrons along a signal path, ds:

$$TEC = \int N_e \, ds \tag{2.1}$$

N_e is the electron density given in number of electrons/m^3 (Johansson, 1997). TEC is often given in TEC units (TECU), where 1 TECU corresponds to 10^{16} electrons per m^2. 1 TECU causes a GPS signal delay in the zenith direction of 0.16 meters on L1 and 0.27 meters on the L2 frequency (Skone, 1998).

2.1.1 Solar activity and sunspots

Generally, ultraviolet radiation from the sun, along with radiation from other cosmic sources, drive the ionization processes in the ionosphere (Johansson, 1997).

Disturbances in the ionosphere are to a large extent caused by the solar wind which is a flow of electrons and ionized particles (mainly H^+ and He^{++}) travelling from the sun into interplanetary space at a supersonic speed (Skone, 1994). The solar wind is a plasma, an ionized gas, and therefore a good conductor (Budden, 1985)

The magnetic field associated with the solar wind is called the interplanetary magnetic field (IMF), and it will, during disturbed periods, cause disturbances in the geomagnetic field, as described in Section 2.1.3. The magnetic field strength of the IMF is often given in the solar–magnetospheric reference system, where the X axis is oriented towards the sun, and the Z axis is oriented towards North (Campbell, 1997). The Z component of the IMF plays an important role in the description of the interactions with the geomagnetic field.

The amount of emission from the sun is correlated with the number of sunspots. Sunspots are local magnetic fields that are formed when magnetic field lines of the sun are twisted (Langley, 2000). The temperature in a sunspot is lower than in the surroundings, which makes the spot look like a dark spot on the sun when it is being observed from the Earth. The sunspot number, and also the solar activity, varies according to a cycle that has an average length of 11 years, see Figure 2.1.

During periods when the sun is more active, the sun also emits radiation caused by solar flares. These are sudden eruptions from the sun, often occurring close to

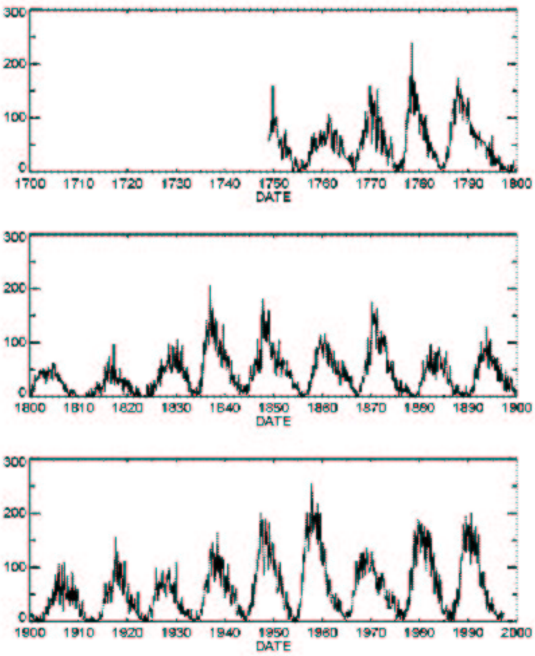

Figure 2.1: Sunspot numbers from 1750 to 2000. Sunspots have been counted since Galileo Galilei began in 1611. Figure from http://helios.gsfc.nasa.gov/.

sunspots, causing an intense emission of ultraviolet radiation. This radiation causes a short increased ionization in the D-region of the ionosphere with a temporal duration of about one hour (Budden, 1985).

Coronal mass ejections are another type of emissions from the sun which occur more frequently during high solar activity. A coronal mass ejection occurs when a large amount of charged particles, mainly protons and electrons, are ejected from the sun. The particles travel with the solar wind, and reach the Earth after 1–5 days where they will cause an increased ionization (Campbell, 1997).

2.1.2 Variation of TEC

The number of electrons along a signal path depends on the ionization and recombination (or de-ionization) of particles as well as on the transport processes in the ionosphere.

The intensity of radiation from the sun affecting the Earth's atmosphere varies with the diurnal rotation of the Earth, so the strongest ionization is occurring on the day side of the Earth, while the night side is less affected by the radiation.

Diurnal variation of TEC

The recombination process is dependent on the collision rate between the particles, and is thereby a function of density, pressure, and temperature. The combination of ionization and recombination processes causes a diurnal cycle of TEC that has a maximum at 14:00, and a daily minimum at sunrise, local time. In the equatorial region there is further a smaller TEC maximum at 22:00 local time which is caused by the so called "fountain effect" described below.

Geographic variation of TEC

The geographical variation of TEC can be described by dividing the ionosphere into three latitudinal regions (Wanninger, 1993):

The Equatorial region is located between approximately 30° South and North of geomagnetic Equator, and is where the highest TEC values, the strongest gradients, and the largest disturbances occur.

The Midlatitude region is where the fewest disturbances occur, and in this region the diurnal variation of TEC is fairly easily predicted.

The Auroral and Polar region is located at geomagnetic latitudes North and South of approximately 65°. These are also regions with major disturbances that are correlated with geomagnetic activity. The auroral zone is treated in detail in Section 2.1.4.

The TEC has a global geographical maximum in two regions located about 10°–20° latitude on either side of the Equator. The high TEC in this region is often referred to as the Equatorial anomaly (Fedrizzi et al., 2002). Intuitively it would make more sense if the TEC maximum was located exactly above the Equator where the radiation from the sun is strongest. The geographical offset is, however, caused by the "fountain effect", whereby electrons are interacting with the Earth's magnetic field by causing movements of the ionization (Langley, 2000), with which the electrons are "lifted" to higher latitudes. Ionospheric errors up to about 100 meters, on GPS signals at low elevation angles have been observed in this region (Wanninger, 1993).

Global TEC-maps are provided by several organizations, for instance by CODE - Center for Orbit Determination in Europe, located at the University of Bern in Switzerland (http://www.cx.unibe.ch/aiub/ionosphere.html). Figure 2.2 is an example of a global TEC map where both the equatorial anomaly and the diurnal variation of TEC is seen.

Annual variation of TEC

The TEC also varies annually, with a maximum at the Equinoxes and a minimum during the summer.

The summer low occur because the recombination rates depend on the temperature, so a higher temperature increases the amount of recombination taking place in the ionosphere, whereby TEC is reduced.

Finally TEC varies along with the 11 year solar cycle, since the ionization rate is dependent on solar radiation and particles emitted from the sun.

According to Klobuchar (1996), the highest TEC value ever observed was about 1000 TECU. This must be seen in relation to an average or "normal" TEC value, but such a value is very difficult to determine because of the high variability. Langley (2000) makes an attempt with a typical value of about 100 TECU.

2.1.3 The auroral zone

The auroral zone is an oval shaped band located at approximately 65°–75° North and South in the ionosphere. These geographical limits are given for a night time ionosphere. The width of the zone is normally $5 - 7°$ in latitude, and the borders

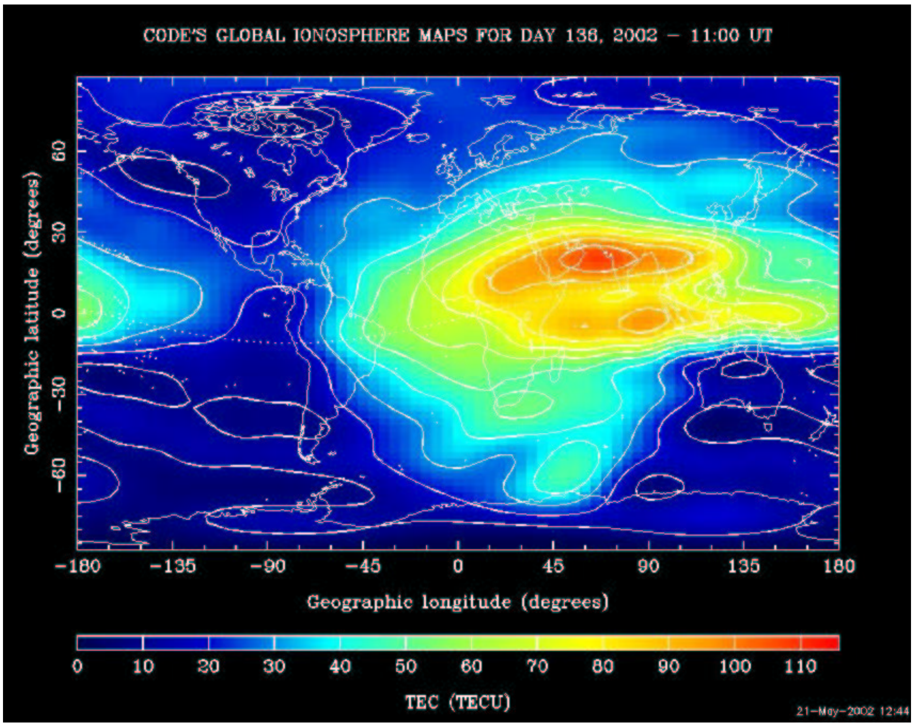

Figure 2.2: Global Ionosphere Map for May 21. 2002 at 11:00 UTC. From: CODE, the University of Bern.

of the zone vary with time of day and solar activity. The zone is generally smaller and closer to the pole on the day side of the Earth. The oval is characterized by an enhanced conductivity and electron precipitation (Skone, 1998). An example of the location of the auroral zone is given in Figure A.15 in Appendix A.

Ionospheric activities in the auroral zone are related to the interaction between the current systems in the auroral E region and energized particles from the solar wind and from solar flares that enter the geomagnetic field (Skone, 1998). When the particles enter the magnetosphere (see Section 2.1.4), they are trapped along the geomagnetic field lines, and move in the direction of the field lines, whereby they are transferred to high latitudes, and can reach down to low altitudes in the auroral ionosphere (Budden, 1985).

When the activity is increased, the visible auroras can be caused by collisions between energetic electrons and neutral molecules - oxygen, nitrogen, and hydrogen (Skone, 1998). See Figure 2.3.

This increased activity in the auroral zone is not taking place at lower latitudes, and the difference in activity causes sharp TEC gradients in the area called the “main trough” that is located just South of the auroral oval. The effect caused by the gradients has been called the “Canadian border effect” because of the geographic location of the area in relation to the USA (Skone, 1998). An example of the trough is given in Figure 2.4.

The gradients in the trough can cause both scintillation and an increased signal delay on GPS satellite signals, two effects that are described in Section 2.2.

During geomagnetic disturbances, the features of the auroral zone move Southward, and can cause disturbances at latitudes where the aurora is not normally seen.

Figure 2.3: Visible aurora. Photo from Spitsbergen on Svalbard. Image from http://www.bergheims.com/hornsund/.

With a moderate auroral activity (Kp=5)[1] the lower boundary of the auroral zone is located at about 60° geomagnetic latitude, which, in Scandinavia, is close to Oslo. With a major auroral activity (Kp=7) the lower boundary is located at about 55° whereby the entire Scandinavian region is enclosed in the auroral zone (Skone et al., 2001). One example of this was the ionospheric and geomagnetic storm on April 6–7, 2000. This incident, and the related effect on GPS data, is described in Appendix A.

Auroral activity has generally not been modelled in global ionospheric models for satellite navigation systems, because the auroral areas have not been interesting for most GPS users. But presently there is more focus on the subject because the American Wide Area Augmentation System (WAAS) for flight navigation will be covering Alaska and parts of Canada, and will thereby operate in the auroral region.

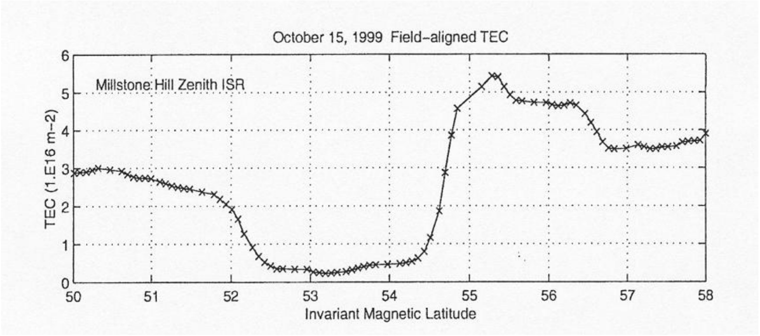

Figure 2.4: An example of the trough. Electron content in TECU versus geomagnetic latitude in degrees. The plot is based on radar data from Milestone Hill, MIT, located at the geographic position 42.2 N, 288.5 E. Figure from Foster (2000).

---

[1] The Kp index and other indices for ionospheric activity are described in Appendix A.

2.1.4 The geomagnetic field

Close to the Earth, variations in the interplanetary magnetic field (IMF) arriving with the solar wind affect the electric current systems, and thereby the geomagnetic activity, so that a high level of solar activity also causes a high level of geomagnetic activity. Signatures of ionospheric activities can therefore be observed using ground based geomagnetic observations (Skone, 1998). In order to interpret the geomagnetic observations, the interaction between the IMF and the geomagnetic field must be understood, and it is described below.

The Earth's magnetic field can be described as a dipole[2] with magnetic field lines surrounding the Earth from the geomagnetic South pole to the geomagnetic North pole (Campbell, 1997). The strength of the magnetic field decrease as the distance from the Earth increase. The region between the Earth and the solar wind, which is the region where the geomagnetic field has the most important effect, is called the magnetosphere (Allnutt, 1989). See Figure 2.5.

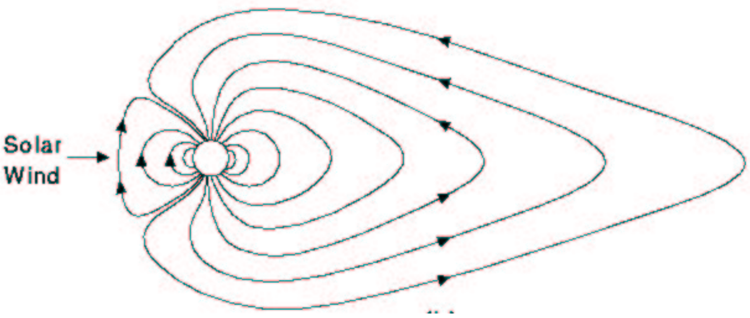

Figure 2.5: Simplified drawing showing a closed magnetosphere. From the web site of C.T. Russel. http://www-ssc.igpp.ucla.edu/˜ctrussel/images/.

When a magnetic field is present within an electric conductor, as in the solar wind, it takes time to change the magnetic field. The field is said to be "frozen in", and for the same reason a magnetic field cannot quickly penetrate into a highly conductive body (Budden, 1985). This is how the geomagnetic field prevents particles from the solar wind from entering the Earth's atmosphere. The geomagnetic field cannot penetrate into the flow of particles, and the particles are thus trapped in the solar wind, and are diverted around the Earth.

The solar wind does, however, distort the geomagnetic field, so on the side of the Earth that faces the sun, i.e. on the day side, the field is compressed. On the night side of the Earth the magnetic field is decompressed, and it reaches out to a distance of about 100 times the Earth radius (Budden, 1985). The decompressed part of the magnetosphere is called the magnetotail. See Figures 2.5 and 2.6.

Disturbance of the geomagnetic field

Normally the IMF does not cause a major disturbance of the geomagnetic field. But sometimes the Z-component of the IMF magnetic field becomes negative, and if the field strength at the same time is suddenly increased, interactions occur. When the Z component of the IMF is negative the IMF is said to be Southward, and when the

---

[2] In reality the geomagnetic field is more complex than a dipole, and the description of the geomagnetic field and the interaction with the solar wind in this chapter is simplified.

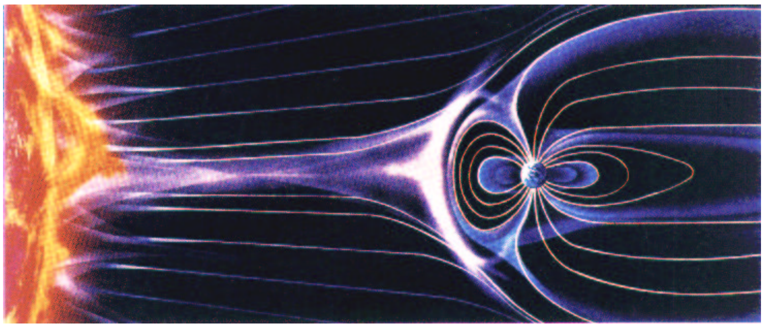

Figure 2.6: Drawing showing the sun, the solar wind, and the geomagnetic field. Courtesy of Windows to the Universe, http://www.windows.ucar.edu.

IMF meets with the geomagnetic field lines that are oriented towards the North, the field lines interconnect and the geomagnetic field is distorted (Campbell, 1997). See Figure 2.7. This opens the magnetosphere, and particles from the solar wind can enter the geomagnetic field, mainly through the night side of the magnetosphere and in the polar regions. These particles will then cause increased ionization.

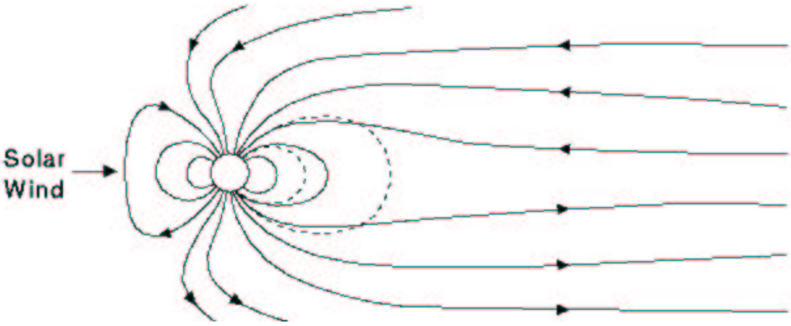

Figure 2.7: Simplified drawing showing an open magnetosphere. From the web site of C.T. Russel. http://www-ssc.igpp.ucla.edu/˜ctrussel/images/.

When the IMF field lines are oriented in the same direction as the geomagnetic field lines, i.e. against the North, which is normally the case, the geomagnetic field is undisturbed and the magnetosphere is closed over the Equator as shown in Figure 2.5.

A Southward IMF can be registered by geomagnetic observations on the Earth surface, as an indication of a geomagnetic or ionospheric storm. In Appendix A some examples of IMF signatures and the corresponding effects of the geomagnetic field are given.

Disturbances of the ionosphere and geomagnetic field can also be caused by meteors and meteorites for instance. Furthermore, eclipses of the sun can cause changes in the ionization rate in the upper part of the atmosphere (Budden, 1985). For example, GPS data collected during the solar eclipse on August 11, 1999 showed a lower noise level, and thereby a higher accuracy of the GPS derived position, than data collected on the following days (Fortes et al., 2000).

2.1.5 Geomagnetic and ionospheric storms

Magnetic storms are global events characterized by magnetic signatures at mid and low latitude geomagnetic observatories (Skone, 1998). The storm activities occur when charged particles from the sun interact with the neutral atmosphere of the Earth and produce extra ions and electrons. The electric fields thereby generated in the neutral atmosphere cause disturbances in the ionosphere, and these can affect the GPS signals (Langley, 1996).

An ionospheric storm is much like a magnetic storm, but it is characterized by changes in TEC instead of changes in magnetic field strength (Skone, 1998).

Magnetospheric sub–storms are local events happening either over the Equator or in the auroral or polar region (Skone, 1998).

Generally, geomagnetic storm activity follows the solar activity cycle, but the storm activities tend to reach a maximum twice during each sunspot cycle. There will thus be increased storm activity just before and just after the maximum sun spot number for each solar cycle. Also there is an annual cycle with more storm activity at Equinoxes (Kunches and Klobuchar, 2000).

Several indices have been developed to characterize ionospheric and geomagnetic activity, and some of these indices are described in Appendix A.

2.2 Effects of the Ionosphere on GPS Signals

If the GPS satellite signals were transmitted through a vacuum, the signal travel time multiplied by the speed of light would equal the geometric distance between the satellite and the receiver. The Earth atmosphere is however not a vacuum, and the signal propagation is affected by particles that cause refraction, i.e. a change of the speed and direction of the signal.(Langley, 1996).

The refractive index, n, is given as the relation between the speed of propagation of a signal in vacuum, c, and the speed of the signal in the current medium, v:

$$n = \frac{c}{v} \tag{2.2}$$

The refractive index changes through the ionosphere as the composition of the media changes. The changing refractive index causes a signal bending, which makes the signal path longer than the geometrical distance. According to Fermats principle, the travel time is, however, shorter than if the signal travelled along the geometrical path, since the signal will follow the path with the shortest travel time. For signals received in the zenith direction the signal bending is negligible, and it will thus be ignored in the rest of this section.

The signal travel time in time units, τ, can be determined as:

$$\tau = \int \frac{1}{v} ds \tag{2.3}$$

where the integration is carried out along the signal path, ds.

By multiplying by c the length of the signal path is determined, and by ignoring the signal bending, the excess signal path, d, is given as the actual signal path subtracted by the geometric distance between satellite and receiver (Misra and Enge, 2001):

$$d = \int \frac{c}{v}\, ds - \int ds \quad = \quad \int n\, ds - \int ds \quad = \quad \int (n-1)\, ds \tag{2.4}$$

where the integration again is carried out along the signal path from receiver to satellite. In the rest of this thesis, the excess path length will be referred to as the signal delay, d, in order to comply with most of the GPS literature.

2.2.1 Phase advance and group delay

If the medium is dispersive, as the ionosphere is for GPS signals, the refractive index is frequency dependent, and it is often described based on the complex Appleton-Hartree formulation of the refractive index given, for instance, by Falthammar (1994).

By introducing several assumptions, namely that the medium contains an equal amount of ions and electrons, that the magnetic field is uniform, that the ions have no effect on the wave propagation, that no electron collisions are taking place, and finally that the effect of the longitudinal component of the Earth's magnetic field can be ignored, the following relatively simple expression for the phase refractive index of a GPS signal through the ionosphere can be obtained:

$$n_\Phi = 1 - \frac{40.3 \cdot N_e}{f^2} \tag{2.5}$$

N_e is the electron density, and f is the frequency of the signal.

A GPS signal, that is a modulated carrier wave, can be considered a superposition of a group of waves of different frequencies. Since the refractive index is a function of frequency, the modulation of the signal will experience another refraction, and will thus propagate at a different speed through the ionosphere than the pure carrier. The group refractive index can be defined as (Langley, 1996):

$$n_g = \frac{c}{v_g} = n + f\,\frac{dn}{df} \tag{2.6}$$

By inserting the expression for the phase refractive index from Equation (2.5) the ionospheric group refractive index can be rewritten as:

$$n_g = 1 + \frac{40.3 \cdot N_e}{f^2} \tag{2.7}$$

Note that the part of the ionospheric refractive index that is deviating from unity has opposite signs for the carrier and the modulated signal. This is causing what is referred to as the phase advance and the group delay, and the interpretation of this is nicely illustrated by Misra and Enge (2001).

The signal delay caused by the ionosphere can now be determined by Equation (2.4), and by the expressions for n_g or n_Φ, Equations (2.5) and (2.7):

$$d_{ion} = \frac{40.3}{f^2} \int N_e ds \tag{2.8}$$

Remembering that:

$$TEC = \int N_e ds \tag{2.9}$$

the size of the ionospheric signal delay for a GPS satellite signal, based on all the assumptions given above, can be expressed as:

$$d_{ion} = \frac{40.3 \cdot TEC}{f^2}. \tag{2.10}$$

The size of the ionospheric signal delay is smallest in the zenith direction, and the zenith delay normally varies, along with the variation of TEC, within 1–3 meters at night and 5–15 meters in the afternoon, at mid-latitudes (Misra and Enge, 2001). These "normal" ranges are larger in the equatorial region.

Examples of the effect for lower elevation angles is shown by Klobuchar et al. (1995) and the trend is that the ionospheric zenith delay is increased with a factor of about 3 for a 10° elevation mask on the L1 frequency. This value also varies slightly with the variation of TEC.

2.2.2 Scintillation

Scintillation is random and very rapid fluctuations of an electromagnetic wave travelling through the ionosphere. The fluctuations are caused by small irregularities in the electron density that essentially is causing the wave to interfere with itself. Scintillation causes a change of the amplitude or a phase shift of the wave (Allnutt, 1989), and scintillation is thus an effect on the signals, not a phenomenon in the ionosphere.

Scintillation occurs in the polar cap, the auroral and the equatorial regions, and it is strongest in the equatorial region. In the polar region the irregularities in electron density are caused by particles from the solar wind entering the ionosphere through the opening in the magnetic field lines over the poles. In the auroral region the irregularities are caused by energetic particles that have entered the magnetosphere and are transferred to the auroral regions along the magnetic field lines. In both regions these particles tend to form "lumps" along the magnetic field lines, mainly in the E and F regions of the ionosphere, and scintillation is caused by small-scale irregularities inside the larger lumps. In these regions scintillation seems to be associated with storm events (S. Skone, University of Calgary, personal communication).

In the equatorial region, or more precisely in the region of the equatorial anomaly, scintillation is caused by the movement of electrons with the fountain effect. The movements cause irregularities in the electron density with either an increased or reduced TEC within relatively small areas (S. Skone, University of Calgary, personal communication). In this region scintillation mainly occur at night, from sunset to midnight, and the activity follows the 11-year solar cycle (Wanninger, 1993).

The irregularities in electron density can cause variations in the refractive index, and at the receiver end of the signal path this shows up as amplitude scintillation, i.e. either a signal fading or a signal enhancement. In both cases the signal strength can jump out of the threshold of the receivers' tracking loop bandwidth causing loss of lock and thereby cycle slips. Rapid variations in the signal to noise ratio, for a received satellite signal, can indicate scintillation (Wanninger, 1993).

Phase scintillation is also caused by sudden changes in the refraction in the ionosphere. With phase scintillation the phase of the signal is changed, and this again can cause the GPS receiver to lose lock if the phase change exceeds the tracking capability of the receiver (Wanninger, 1993).

Langley (2000) gives some numerical examples of this. For a receiver tracking the L1 frequency a change in TEC of 0.19 TECU, corresponding to 0.2% of the typical

TEC value during a time interval of 1 second, can be enough to cause problems with the signal tracking if the receiver tracking bandwidth[3] is 1 Hz.

Nichols et al. (2000) and Skone et al. (2001) both tested the effect of auroral scintillation on dual frequency GPS receivers. Their conclusions are similar in that loss of lock frequently occurs for the L2 frequency during high levels of activity. Nichols et al. (2000) state that loss of lock will affect a receiver tracking the L2 with a semi-codeless technique less than a receiver using a codeless technique.

Skone et al. (2001) confirmed this by performing tests using both types of receivers. The semi-codeless receivers tested did not lose lock as often, instead they suffered from multiple cycle slips at time intervals when the codeless receiver lost lock.

These results occur because when using a codeless technique the tracking noise is higher than when using a semi-codeless technique, and when the noise of the incoming signal is further increased because of scintillation, recovery of the L2 is made more difficult.

Both papers point out that tracking of the L1 frequency was not significantly affected by scintillation. The better results on L1 occur because the L1 tracking is based on code correlation using the C/A code as described in Section 1.1.1.

2.3 Handling the Ionospheric Effect in GPS Positioning

The ionospheric effect on a GPS signal is the largest error source in the positioning process, so it is important to handle the ionospheric effect in order to obtain accurate positions. An ionospheric model, the so-called Klobuchar-model, is broadcast with the navigation message, and it can be used for absolute positioning, for code-based DGPS, and also for single frequency carrier–phase based positioning over short baselines. The model is described by Misra and Enge (2001).

For high accuracy positioning this model is not sufficient and other techniques must be introduced.

The ionosphere–free linear combination of the L1 and L2 observables

The majority of the ionospheric effect can be eliminated by double differencing, and because of the spatial correlation of the ionospheric effect the residual effect is dependent on baseline length. The residual ionospheric effect after double differencing can be about 4 cm (RMS) for a 67 km baseline or about 10 cm (RMS) for a 192 km baseline during afternoon hours at a period with low ionospheric activity (Raquet, 1998).

For high accuracy positioning, double differencing might therefore not be sufficient, so the ionosphere–free linear combination is introduced in the processing to reduce the remaining part of the ionospheric effect.

Several ionosphere–free combinations are seen in GPS literature. One example is given in Equation (1.7), and another example is the following which is given by, for

---

[3] For a geodetic GPS receiver used in static mode for tracking the carrier–phase, a tracking bandwidth of about 2 Hz is sufficient (Langley, 1996), and therefore normally used. With a smaller tracking bandwidth the noise level is reduced, but the signal is also lost more easily if the phase of the signal is shifted by interference. Many modern GPS receivers therefore use variable tracking bandwidths, where the bandwidth is wide until signal lock is obtained, and then narrows down. Receivers made for dynamic environments further monitor the dynamics and keep a highly variable bandwidth also after signal lock has been obtained.

instance, Beutler et al. (2000):

$$\Phi_{IF} = \frac{f_{L1}^2}{f_{L1}^2 - f_{L2}^2}\Phi_{L1} - \frac{f_{L2}^2}{f_{L1}^2 - f_{L2}^2}\Phi_{L2} \tag{2.11}$$

Both expressions remove the first order effect of the ionosphere, $d_{ion} = 40.3 \cdot TEC/f^2$ when the IF combination of the L1 and L2 observation equations are formed. But considering the assumptions necessary to derive the expression for the first order delay, given in Section 2.2.1, higher order effects will be present. For higher elevation angles these effects are negligible, but for elevation angles below 15° the higher order effects become significant (Johansson, 1997).

Brunner and Gu (1991) developed an improved model, also based on the Appleton-Hartree formulation, where some higher order terms are included. Their results show a considerable improvement in estimation of the ionospheric range error for test scenarios with high TEC values. They do conclude that the IF combination given above is good within about 3 cm for elevation angles above 15° when the electron content is less than about 140 TECU. With the improved correction model the corresponding residual range errors were at the mm-level also when tested with a high electron content of 455 TECU.

The estimated accuracy of the IF combination of 3 cm given by Brunner and Gu (1991) is based on undifferenced observations. If the IF combination is used in a double difference mode, as given above, the residual ionospheric effect will be almost negligible for medium sized baselines. The model suggested by Brunner and Gu is therefore only relevant for long baselines or during very high ionospheric activity.

Before leaving the ionosphere the reader is referred to Appendix A, where several indices for describing the ionospheric activity are described. The appendix also contains references to internet sites where forecasts of ionospheric activity are available. GPS survey practice should include consultation of these forecasts when planning survey activities, also if dual frequency receivers are used. The influence of the higher order effects cannot be deducted directly from the forecasts, but the influence of these effects are more severe during high ionospheric activity. Scintillation effects cannot be directly deducted from the forecasts either, but when a geomagnetic storm has been predicted the risk of scintillation effects is increased and should be considered.

2.4 The Neutral Atmosphere

The neutral atmosphere is electrically neutral and nondispersive to electromagnetic waves transmitted at radio frequencies below 30 GHz (Johansson, 1997). The refractive index in the neutral atmosphere is very close to unity, about 1.0003 at sea level and decreasing with higher altitude (Misra and Enge, 2001), so the signal bending is small. The speed of the signal in the neutral atmosphere is slower than in a vacuum, and this causes a signal delay which affects GPS positioning. This delay and the signal bending is the focus of the following sections.

Figure 2.8 shows the characteristics of pressure and temperature within the neutral atmosphere. The parameters are plotted as a function of orthometric height, and are based on the US standard atmosphere.

The neutral atmosphere covers two regions of the atmosphere:

The Troposphere is the lowest part of the atmosphere extending from the surface of the Earth to an altitude of about 10 km. In the troposphere, the temperature decreases with altitude. This is also the part of the atmosphere where the "weather" occurs, i.e. where most of the clouds are located and where precipitation occurs (Johansson, 1997).

The troposphere is an active region and is characterized by strong vertical mixing where a particle can travel through the entire troposphere within a few minutes e.g. in connection with a strong thunderstorm (Wallace and Hobbs, 1977).

The tropopause marks the top of the troposphere and is characterized by a temperature lapse rate of zero. The height of the tropopause varies with latitude and is located at an altitude of about 16–18 km at the Equator and at about 7–10 km at the Poles (Mendes, 1999). The height of the tropospause is more unstable at the Poles where it is affected by seasonal variations (Spilker, 1996).

The Statosphere is the upper part of the neutral atmosphere and it is characterized by a fairly constant temperature in the lowest part, and by a temperature increasing with height in the upper part (Johansson, 1997).

The lowest part of the stratosphere is also characterized by a sharp decrease in water vapor, and by an increase in the amount of ozone (Wallace and Hobbs, 1977). The natural presence of ozone is the reason for the increasing temperature in the upper part of the region, since ozone absorbs ultraviolet radiation from the sun (Johansson, 1997). The stratosphere is a quiet region, and the vertical mixing is small (Wallace and Hobbs, 1977).

The upper boundary of the statosphere is the statospause, where the temperature lapse rate again is zero. The stratospause is located at an altitude of about 50 km (Johansson, 1997).

These two regions affect radio waves in almost the same manner and are thus often treated as one region - the neutral atmosphere.

2.5 Effects of the Neutral Atmosphere on GPS Signals

As the signals travel through the neutral atmosphere they are affected by refraction, causing the receiver–satellite distance determined from the satellite signals, to be longer than the geometrical path the signal would have followed if it travelled through a vacuum. In GPS terminology, this effect is often referred to as the "tropospheric delay", since the largest part of the effect occurs in the troposphere.

GPS signals will also be slightly affected by attenuation caused by water vapor and rain, but these effects are negligible for GPS frequencies (Spilker, 1996). Further, scintillation caused by atmospheric turbulence can be significant at lower altitudes and for signals received at low elevation angles. But the effects are generally small for GPS frequencies (Spilker, 1996), and are therefore not considered any further in this thesis.

The signal delay experienced by a satellite signal traversing the neutral atmosphere can be determined by integrating the refractive index along the signal path, just as

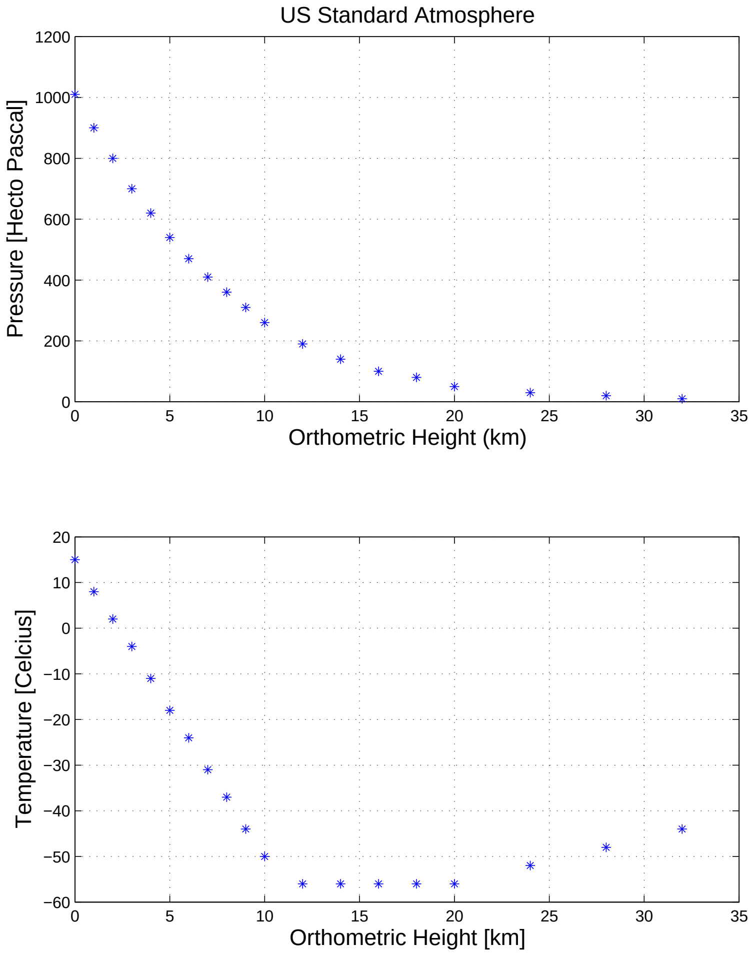

Figure 2.8: Pressure and temperature plotted as a function of orthometric height. Values for the US standard atmosphere are from Holton (1992).

for the signal delay in the ionosphere (Equation 2.4). Introducing the refractivity, $N = (n-1) \cdot 10^6$, as a new variable, the signal delay is given as:

$$delay = 10^{-6} \int N ds \tag{2.12}$$

The integration is carried out along the signal path, ds, from the GPS antenna to the top of the neutral atmosphere.

The refractivity can be determined using the equation below, as given for instance by (Langley, 1996).

$$N = k_1 \left(\frac{P_d}{T} \right) Z_d^{-1} + \left(k_2 \frac{e}{T} + k_3 \frac{e}{T^2} \right) Z_w^{-1} \tag{2.13}$$

k_1, k_2, k_3 are constants, P_d is the partial pressure of dry air in mbar, T is temperature in Kelvin, and e is the partial pressure of water vapor also in mbar. Z_d and Z_w are the compressibility factors for dry air and water vapor respectively, and they account for the deviation of the gasses from an ideal gas.

The relation between the pressure of dry air and the water vapor pressure is given by Dalton's law of partial pressures (Wallace and Hobbs, 1977):

$$P = P_d + e \tag{2.14}$$

where P is the total atmospheric pressure.

Several sets of values for the empirical k-constants have been determined and there has been some dispute over which values are the best (Mendes, 1999). A new set of k-values based on a re-analysis of older experimentally determined constants was provided by Bevis et al. (1994). These values will be used for testing purposes in Chapter 4 and are given in Table 2.1.

k_1 $[K/mbar]$	k_2 $[K/mbar]$	k_3 $[K^2/mbar]$
77.60 ± 0.05	70.4 ± 2.2	373900 ± 12000

Table 2.1: Refractivity constants from Bevis et al. (1994).

The compressibility factors represent the non-ideal behavior of the gasses in the neutral atmosphere. This is described by the equation of state:

$$P_i = \rho_i \; R_i \; T \; Z_i \tag{2.15}$$

P_i is the partial pressure for constituent i, ρ_i is the mass density, R_i is the specific gas constant for the constituent, and T is the absolute temperature (Davis et al., 1985).

For an ideal gas the compressibility factor equals 1 (one), and for the neutral atmosphere the factor only differs a few parts per thousand (Davis et al., 1985). The compressibility factors are functions of temperature and pressure, and the expressions for the wet and dry compressibility factors are given for instance by Spilker (1996).

Returning to Equation (2.13),the first term in the formula is called the dry refractivity, since it depends only on the dry constituents. The term in the last bracket is correspondingly called the wet refractivity.

Davis et al. (1985) introduced another expression for the refractivity:

$$N = k_1 R_d \rho + \left(k_2' \frac{e}{T} + k_3 \frac{e}{T^2} \right) Z_w^{-1} \tag{2.16}$$

where the new constant $k_2' = k_2 - k_1 M_w / M_d$. M_w and M_d are the molar mass of water vapor and dry air, respectively.

Here the first term, called the hydrostatic refractivity, is not only dependent on the dry constituents since the total density is included. The last term is still based on the wet constituents and is thus still called the wet refractivity, although the term non-hydrostatic refractivity is also found in the literature.

Equation (2.16) is useful for determining the hydrostatic signal delay in the zenith direction, since it is not affected by the water vapor content (Langley, 1996).

The total tropospheric delay, d_{trop}, can now be determined as:

$$d_{trop} = 10^{-6} \int N_h \; ds + 10^{-6} \int N_w \; ds \quad = 10^{-6} \int N_d \; ds + 10^{-6} \int N_w \; ds \tag{2.17}$$

where hydrostatic is denoted by the subscript h, the wet part is denoted by subscript w, and subscript d denotes dry. Either of the expressions can be used depending on whether the dry or the hydrostatic approach is preferred.

The dry or hydrostatic part of the delay is approximately 90% of the total tropospheric delay, and has, when determined for a signal received in the zenith, a size of about 2.3 meters when the GPS receiver is located at sea level. The dry delay is easily predictable and is thus quite well modelled (Spilker, 1996).

Contributions to the wet delay are only present in the lowest part of the neutral atmosphere, i.e. in the troposphere. The wet part of the delay accounts for the remaining 10% of the total delay, and it is highly variable. The size of the wet delay is only a few mm in dry arctic areas or in deserts, but it can be up to 40 cm in tropical regions (Brunner and Welsch, 1993). The behavior and distribution of water vapor in the atmosphere is difficult to predict, and therefore it is also difficult to model the wet part of the delay. The spatial and temporal variability of water vapor and the wet tropospheric delay is further discussed in Chapter 3.

2.6 Modelling the Tropospheric Delay

In GPS positioning, the major part of the tropospheric delay is handled by means of global tropospheric delay models based on climate data. The models are functional solutions for the integral given in Equation (2.17), and they are often dependent on surface meteorological observations as input parameters. Also, the latitude and height of the GPS station are necessary parameters for determining the tropospheric delay.

Several models exist and three different models are briefly described here.

The Hopfield model

The Hopfield model (Hopfield, 1969) was originally developed for Doppler positioning, but is also used for GPS positioning. The tropospheric delay in the zenith direction, ztd, is given as:

$$ztd = \frac{1}{5}10^{-6}\Big(77.6 \cdot \frac{P_s}{T_s}(h_d^e - h_{ant})\Big) + \Big(3.73 \cdot 10^5 \frac{e_s}{T_s^2}(h_w^e - h_{ant})\Big) \qquad (2.18)$$

Where P_s is the surface pressure of dry air in mbar, T_s is surface temperature in Kelvin, e_s is the surface partial pressure of water vapor in mbar, and H_{ant} is the geometric height of the GPS antenna in km. h_d^e and h_e^e are the dry and wet equivalent heights, i.e. the geometric heights where the dry and wet refractivity is zero. Hopfield (1969) determines h_w^e to 12 km and $h_d^e = 43.130km - 5.206\ sin^2(\phi_{ant})$, where ϕ_{ant} is the latitude of the GPS antenna.

Hopfield later published improved expressions for both the dry and wet equivalent heights, so several Hopfield models exist, see for example Hopfield (1971).

The work by Hopfield forms the basis for many tropospheric models, and the basic Hopfield model has been used extensively, with various modifications, for GPS positioning.

The Ifadis model

Ifadis (1986) developed several models for the zenith wet delay, zwd, applicable for various climate regions or seasons of the year. He also developed a global model given below:

$$zwd = 0.00554 - 0.880 \cdot 10^{-4}(P_s - 1000.0) + 0.272 \cdot 10^{-4} e_s + 2.771\left(\frac{e_s}{T_s}\right) \tag{2.19}$$

As for the Hopfield model P_s is in mbar, T_s is in Kelvin, and e_s is in mbar.

The Saastamoinen model

Saastamoinen (1973) developed both a precise model and a more simple standard model. The standard model, which is given below, is widely used for GPS positioning:

$$d_{trop} = 0.002277 D \cdot sec(90 - elv)\left(P_s + \left(\frac{1255}{T_s} + 0.05\right)e_s - B \cdot tan^2(90 - elv)\right) + \delta \tag{2.20}$$

where elv is the elevation angle of the satellite signal in degrees, and again P_s is in mbar, T_s in Kelvin, and e_s in mbar. B and δ are correction quantities determined from tabular values given in the original papers and also in e.g. Spilker (1996). $D = 1 + 0.0026 cos^2(\phi) + 0.00028 h_{ant}$ where ϕ is the latitude and H is the height of the GPS station.

This model is not only used for determining zenith delays, as the previous models, since it is also a function of the elevation angle of the received satellite signal. Other functions for mapping the zenith delay to lower elevation angles are found in Section 2.6.1.

Evaluation of the models

Mendes (1999) performed a comprehensive analysis and evaluation of several tropospheric delay models. The evaluation was based on signal delays derived by radiosonde observations collected over one year in 50 sites distributed globally.

When analyzing the differences between zenith hydrostatic delays determined from models and from radiosondes, the four hydrostatic delay models tested all gave a total error of about 5 mm or less. The Saastamoinen model performed the best with a total error of 0.2 mm.

When the size of the zenith hydrostatic delay is 2.3 meters, as it normally is at sea level (see for instance Figure 4.3), the modelling accuracy is consequently about 0.009 %.

When analyzing wet delay models the results varied more. The total error for the best models was about 3 cm, and the error for the worst of the 12 models tested, was about 4 cm. The Saastamoinen and Ifadis models for wet delay performed the best (Mendes, 1999).

Considering the size of the zenith wet delay, which is about 5–20 cm at sea level in Denmark (see Figure 4.4), the modelling accuracy of 3 cm corresponds to 15 - 60 % of the delay. This clearly illustrates the difficulties in modelling the wet delay as compared to the hydrostatic delay.

The basic expressions for refractivity and therefore the global tropospheric delay models do not consider the effect on the signal delay caused by nongaseous constituents in the atmosphere. A discussion on the relatively small influence of e.g. hydrometeors, volcanich ash, rain, and snow on GPS signal delays is given by Solheim et al. (1999). According to Johansson (1997) the influence on path delays of clouds and rain is almost always less than 1 cm.

2.6.1 Mapping functions

The total delay for a satellite signal received at any elevation angle can be estimated by first determining the delay for a signal received in the zenith direction and then multiplying the zenith delay with a unitless mapping factor. The factor is determined by a mapping function, $m(elv)$, and often both a hydrostatic and a wet mapping function is used, so that the total delay for a signal received at any elevation, elv, is determined as:

$$delay(elv) = zhd \cdot m_h(elv) + zwd \cdot m_w(elv) \tag{2.21}$$

Where zhd is the zenith hydrostatic delay, zwd is the zenith wet delay, and the subscripts h and w denote hydrostatic and wet mapping functions respectively. The mapping factor is about 4 for a 15° elevation angle, about 6 for a 10° elevation, and about 10 for a 5° elevation angle.

The most simple form of a mapping function is:

$$m(elv) = \frac{1}{sin(elv)} \tag{2.22}$$

This function can be used for signals received at high elevation angles (Langley, 1996), but for signals received at lower elevation angles, with more signal bending caused by the water vapor in the troposphere, this simple function is insufficient.

Several mapping functions have been developed during the last two decades. The mapping functions developed by Ifadis (1986), Herring (1992), Niell (1996) and by Goad and Goodman and described by Leick (1990), will all be encountered later in this thesis. The functions are fairly voluminous and the reader is referred to the original publications, or to Mendes (1999), for the equations.

The mapping functions differ, as the zenith delay models do, in that they require input of different meteorological parameters. For example, the Goad and Goodman mapping function requires both surface pressure, temperature, water vapor pressure, and station height, whereas the mapping function developed by Niell does not require any meteorological input, but only the station latitude and height.

Mendes (1999) analyzed 12 different mapping functions, using the same radiosonde data as for the zenith delay tests mentioned above. He concluded that the NMF mapping function (Niell, 1996) is the overall best mapping function when no meteorological data is available. If meteorological observations are available, the Ifadis (Ifadis, 1986) and the MTT (Herring, 1992) mapping functions are recommended.

Mapping functions are based on an assumption of a horizontally homogeneous or isotropic atmosphere. However, the satellite signals are not encountering exactly the same conditions when received at different azimuth angles. This is in particular the case for satellites received at low elevation angles where a larger part of the signal path goes through the more variable lower parts of the atmosphere.

Introducing horizontal gradients in the processing of GPS data can provide a solution to this problem. Emardson and Jarlemark (1999) tested the influence of introducing horizontal gradients in a single point positioning simulation. When comparing the results for the vertical coordinate component with and without horizontal gradient estimation, an improvement in the RMS error of up to 7 mm was found.

This subject is treated in more detail in Chapter 5 where the performance of various mapping functions is tested for various azimuth angles.

2.7 Estimation of Tropospheric Delay from GPS Data

The amount of water vapor in the atmosphere can be measured with a water vapor radiometer. The instrument is accurate, but it is unreliable during precipitation and also fairly expensive. Thus other methods for extracting information about the water vapor content in the atmosphere has been proposed, and one of them is to make use of data from permanent GPS reference stations. This was suggested by Bevis et al. (1992) and has since become a considerable research area within meteorology, where better estimates of the water vapor distribution in the atmosphere, e.g. from GPS data, can improve weather predictions (Gutman and Benjamin, 2001).

Consider the carrier–phase observation equation again:

$$\Phi = r + dr - d_{ion} + d_{trop} + c(dt - dT) + \lambda N + \nu \tag{2.23}$$

The tropospheric delay can be estimated from this equation if the ambiguities are successfully resolved, and if the other error sources are well modelled. The station coordinates must be known, and precise satellite orbits are necessary. The influence of the ionospheric delay must be minimized, e.g. by using the ionosphere–free linear combination, and the clock errors can be handled either by double differencing or by explicit modelling of the clock drift (Bevis et al., 1992). The noise, ϵ, is reduced by carefully selecting the GPS site and by using high quality equipment. Multipath is further reduced by using antennas equipped with choke rings.

When using the Bernese software, for example, for determining the tropospheric delay from GPS data, the majority of the delay is estimated using a global delay model, and the residual part is estimated as a parameter, valid for a time interval of for instance two hours, through the double difference based adjustment process (Beutler et al., 2000).

Another approach, which is used in the GIPSY software, is to use a stochastic model, treat the residual delay as a time-varying parameter, and use Kalman filtering for the adjustments.

The tropospheric delay produced by geodetic GPS inversions is the total zenith delay. Meteorologist are, however, only interested in the wet zenith delay, that can be determined by subtracting the hydrostatic part of the delay. If this separation is not carried out as a part of the GPS positioning process, the zenith hydrostatic delay can be determined based on a hydrostatic delay model if surface pressure (and temperature) observations at the site are available (Duan et al., 1996).

The zenith wet delay can be converted to Integrated Water Vapor, IWV, or to Precipitable Water Vapor, PW, which is directly useable for weather modelling and predictions. IWV is the mass of water vapor per unit area given in kg/m^2, and PW is the height of an equivalent column of water, given in mm. The two variables differ only by the density of liquid water; PW = IWV/ρ_l (Bevis et al., 1992).

These values can then be used for validating and/or improving numerical weather predictions (Elgered, 2001). A considerable amount of research is therefore going on, focusing on deriving water vapor not only from ground based GPS stations, but also from satellite occultations.

One example is the MAGIC[4] project which was a European research project run-

---

[4] MAGIC - Meteorological Applications of GPS Integrated Column Water Vapour Measurements in the Western Mediterranean. A cooperation between meteorological organisations in Spain, Italy, France and Denmark.

ning from 1998–2001, focusing on these subjects with the purpose of developing a near real-time system for introducing IWV derived from GPS data into weather predictions. The project included 51 permanent GPS stations and the data was processed by three processing centers using the GAMIT and GIPSY software in different modes. The derived zenith delays and IWV were evaluated by three other organizations, and the integration into numerical weather predictions was tested using a special MAGIC implementation of the DMI-HIRLAM-system (see Chapter 3). Comparisons based on 1.5 years of near real-time estimates of zenith total delays show a standard deviation of 12 mm when compared to zenith delays derived from radiosonde data (Haase et al., 2001a). More results from the MAGIC project are referenced in Chapter 3.

The NOAA[5] Forecast Systems Laboratory is estimating near real-time water vapor based on 59 GPS stations in USA. They obtain a difference of 1.2 mm precipitable water vapor when comparing hourly results with final daily solutions. This difference corresponds to a difference of about 8 mm in a zenith wet delay. The integrated precipitable water vapor values can be found on NOAAs web site along with other real-time meteorological observations (http://www.gpsmet.noaa.gov/jsp).

2.8 Handling the Tropospheric Delay for GPS Positioning

The tropospheric signal delay is spatially correlated for differential GPS positioning. When the two receivers are located close to each other the signals travel through basically the same part of the atmosphere. As the distance between the receivers increases so does the differential tropospheric delay.

Janes (1991) mentions that the differential tropospheric delay is about 0.3–0.5 ppm. This would cause an error of about 2 cm for a 50 km baseline.

The unmodelled differential tropospheric delay primarily affects the height difference for a GPS baseline, and Brunner and Welsch (1993) sets a rule of thumb for relative GPS positioning:

Error in height difference = 3 x differential tropospheric delay

The rule is valid with an elevation mask of 15°, and it means that a differential tropospheric delay of 1 cm causes an error of 3 cm in the height difference between the two stations.

This rule of thumb, and the 0.5 ppm, indicates how important handling of the troposphere is for high accuracy GPS positioning. If the differential delay is not taken into account, it can interfere with ambiguity resolution by extending the time to resolve ambiguities, thereby degrading the positioning performance. Even if the ambiguities are solved successfully, any residual tropospheric errors can be encapsulated into the station coordinates of the rover.

For differential positioning based on double differencing, handling of the tropospheric delay is particularly important when the atmospheric conditions are considerably different at the sites of the reference and rover. This can be the case for long baselines. But also for shorter baselines the differential tropospheric error can be considerable during the passage of a weather front, or if there is a large height difference between the two stations.

---

[5] NOAA is the National Oceanic and Atmospheric Administration in the US.

2.8.1 The tropospheric delay for post–processing

For high accuracy GPS positioning the raw observations are initially corrected for the tropospheric delay by means of a global tropospheric delay model. If the model used requires input of meteorological surface observations, real observations collected at the site of the reference and rover receiver can be used. However, the models are very sensitive to measurement or calibration errors of the meteorological equipment. In Beutler et al. (2000) some numerical examples are given on how meteorological observational errors influence the estimate of the tropospheric signal delay. For example, an error of 1% in relative humidity can cause an error in the zenith delay estimate of 4 mm, which then translates into an error in the height difference that is three times as much i.e. 1.2 cm.

Instead, standard meteorological parameters, as for instance given in Table 2.2, can be used.

Temperature	Pressure	Relative humidity
18° C	1013 mbar	50 %

Table 2.2: Example of standard meteorological values used for GPS positioning for at station located at sealevel. From (Beutler et al., 2000).

After the data has been corrected for the majority of the tropospheric delay by a global delay model, the observations are normally double differenced. For shorter baselines the residual tropospheric delay will then be negligible as mentioned in Chapter 1.

For longer baselines, or baselines with a large height difference, the residual tropospheric delay can have a significant effect, and it must be considered during the GPS positioning process. Within the Bernese software, for example, the residual tropospheric error can be estimated as an element in the carrier–phase adjustment process where also the ambiguities are estimated (Beutler et al., 2000).

2.8.2 The tropospheric delay for real-time positioning

When performing real-time differential code–based GPS positioning (DGPS), where positions with accuracies within 1–5 meters are obtained, the computation process in the rover must be simple due to the time and processor constraints. The tropospheric delay is therefore modelled using a fairly simple global model that does not require any meteorological parameters as input. Some examples of models are given by Mendes (1999).

For wide area DGPS systems (WADGPS), where the objective is to obtain improved positioning accuracy (0.5 meters) and reliability, a slightly different approach is used. For instance, for the European Geostationary Navigation Overlay Service (EGNOS), a global tropospheric delay model is combined with meteorological data for the site of the roving GPS receiver. The meteorological data is provided by a regional climate model, and the parameters are extracted from the model by latitude and day of year (Dodson et al., 1999).

Penna et al. (2001) tested this approach for five different sites in Great Britain using one year of GPS data. For the stations tested, the RMS of the difference between tropospheric zenith delays based on the EGNOS approach and zenith delays determined by post–processed GPS data ranged between 4 and 5 cm. The maximum error encountered during the year was 18 cm.

Another approach was suggested by Schueler et al. (2000a), where the tropospheric zenith delays for a given location was estimated by interpolation in a grid file with tropospheric zenith delays. The grid was based on a global numerical weather model improved by incorporating information from permanent GPS stations. The zenith delays in the two-dimensional grid files were determined for each grid point in the three dimensional weather model, and by interpolation in the new grid file, the zenith delay for a roving GPS receiver could be determined.

The method was evaluated by Schueler (2001) by determining zenith delays for 140 globally distributed GPS stations using 30 days of data. A global weather model from NOAA, with a resolution of 1° x 1°, formed the basis for the zenith delay files. Zenith delays determined by interpolation in the new files were compared with delays determined using post–processed GPS data, and the RMS of the difference between the two sets of results was 7 mm. However, the exact same GPS data was used for generating the zenith delay grid files and for testing the approach, so the result is overly optimistic. Schueler (2001) focused on the use of permanent GPS stations for meteorology, but the perspective of using the zenith delay grid files for kinematic GPS positioning is mentioned as well.

2.8.3 RTK and the tropospheric delay

In connection with RTK, the tropospheric delay is normally treated by global delay models followed by double differencing, just as in a post–processing situation. The tropospheric errors thus cancel out for shorter baselines, i.e. up to 20–30 km (Cannon, 1997).

For medium sized baselines, the residual error after both modeling and double differencing can, for instance, have a size of about 1.5 cm for a 67 km baseline, or 1.9 cm for a 192 km baseline under normal conditions as shown by Raquet (1998). Considering the rule of thumb given in Section 2.8, the height of the rover station will be influenced by an error of about 4.5 cm for the 67 km baseline.

For traditional RTK this residual effect cannot be handled, and it is thus the major reason for the distance limitation of the technique.

If a network of reference stations is available, combined network RTK corrections can be used to account for most of the residual tropospheric error experienced by the rover. This is the case for instance with the NetAdjust method. Corrections compensating only for the tropospheric residual error can also be generated based on a network of reference stations, as shown by Zhang and Lachapelle (2001). Both solutions provide a means to reduce the influence of the residual tropospheric delay in GPS positioning.

Another solution is to provide an estimate of the tropospheric delay, which is better than what can be provided in real-time by the global tropospheric delay models. If the initial estimate of the tropospheric delay is improved the residual tropospheric errors will also decrease.

Since real-time estimates of the zenith delays are available, as determined e.g. by the MAGIC project (Haase et al., 2001a), a solution could be to use those delays for RTK positioning. This might, however, introduce correlations in the positioning process if the same GPS data is first used for generating the zenith delays and then used in the positioning process to correct for the same zenith delays. A scenario like this could easily occur in connection with a network RTK system where permanent GPS reference stations could be providing data for both the RTK system and for the meteorological organization estimating zenith delays for weather prediction purposes.

Another solution is to integrate tropospheric delays determined from a numerical weather prediction into the network RTK processing. As long as GPS data is not used on an operational basis for weather predictions, tropospheric delays determined from weather predictions are independent observables when introduced in the GPS processing.

This approach has been suggested by a number of groups lately:

Schueler et al. (2000b) suggested the use of numerical weather predictions for WADGPS, and Pany et al. (2001b) discuss a number of methods for using numerical weather predictions for static positioning using the Bernese software. Tsujii et al. (2001) show how "on the fly" ambiguity resolution can be improved for kinematic post–processed positioning, when introducing a numerical weather prediction with a 40 km grid spacing. When zenith delays from the weather model were used instead of a global tropospheric delay model, improvements were obtained in the percentage of fixed ambiguities using both a 15° and a 10° elevation mask and a 1-hour data set. Finally, Behrend et al. (2001) tested the use of zenith delays determined from a numerical weather prediction with a 0.2° grid resolution for single point GPS positioning with the GIPSY software. Using the new approach, improvements were obtained in the positioning domain, but only when residual tropospheric errors were not estimated in the positioning process.

The last two investigations indicate that numerical weather predictions might be useful for RTK positioning, and the remaining part of this thesis deals with the introduction of numerical weather predictions for network RTK.

3

NUMERICAL WEATHER PREDICTIONS

Numerical Weather Predictions (NWPs) are three dimensional models of the conditions in the lowest part of the atmosphere, from the surface of the Earth up to an altitude of about 30 km. The models contain predicted information about temperature, humidity, wind speed and direction, amount of precipitation, cloud cover etc. The purpose of an NWP is to predict the future state of the atmospheric circulation from information on the present conditions by using numerical approximations of the dynamical equations describing the atmospheric circulation (Holton, 1992).

The present conditions of the atmosphere are given by numerous meteorological observations collected globally. An interpolation scheme is used for distribution of the irregularly spaced observations onto a regular grid that forms the basis for the numerical model development. This process, which also includes filtering and blunder detection of the observations, is referred to as data assimilation.

The meteorological observations originate from mainly radiosondes, but also transatlantic airplanes, and microwave radiometers are providing information about conditions in the atmosphere. The density of the data sources is limited, and this is the reason for the considerable meteorological interest in using GPS as a tool for probing the neutral atmosphere as described in Section 2.7.

In this chapter the DMI-HIRLAM-system for numerical weather predictions is briefly described, followed by a description of how to estimate tropospheric zenith delays using data from the NWP. The chapter ends with an evaluation of the expected accuracy of tropospheric zenith delays determined using DMI-HIRLAM NWPs.

3.1 The DMI-HIRLAM System

The HIRLAM (High Resolution Limited Area Model) system is an outcome of the international HIRLAM project that was initiated in 1985 as a cooperation between meteorological institutes in 8 different European countries, including Denmark. The purpose of the HIRLAM project is to develop and maintain a short-range weather forecasting system for operational use. The project is described on the HIRLAM web site: http://www.knmi.nl/hirlam/

The DMI-HIRLAM system is the Danish implementation of HIRLAM. It is running operationally at the Danish Meteorological Institute (DMI), and the system is optimized for short range forecasts up to about two days ahead (Sass et al., 2000).

DMI-HIRLAM is a three-dimensional grid-point model derived on a rotated grid of latitude and longitude. As the vertical reference a pressure–sigma system is used (Rasmussen et al., 2000). Sigma coordinates, σ, are normally given as $\sigma = P/P_s$, where P is pressure of the current level or layer in the model, and P_s is the pressure at the Earth's surface. So σ equals one at the surface of the Earth, and σ equals zero at the top of the atmosphere (Holton, 1992).

The forecast variables are the two horizontal wind components, surface pressure, temperature, specific humidity, specific cloud condensate, and turbulent kinetic energy. Continuous dynamic equations describing the rate of change of these parameters form the basis of the DMI-HIRLAM-system, and they are combined with models for atmospheric physics such as radiation, turbulence, condensation and convection (Sass et al., 2000).

Four different NWPs are generated to cover the areas for which DMI is responsible for providing weather forecasts, i.e. Denmark, the Faroe Islands and Greenland. The geographical extension of the models is shown in Figure 3.1.

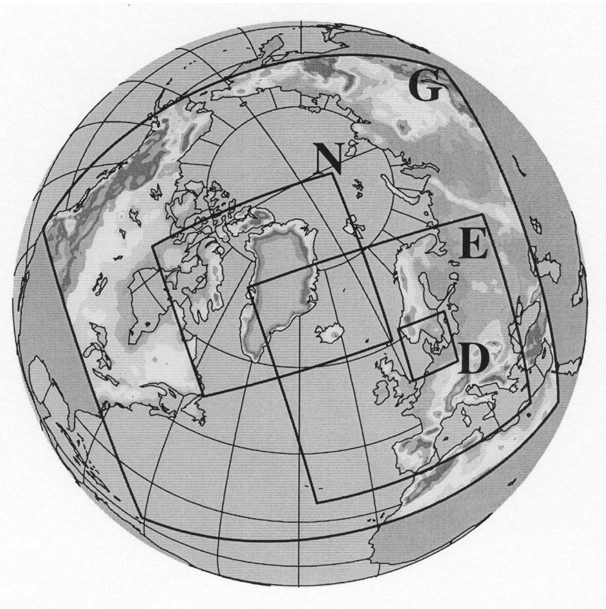

Figure 3.1: Geographical extension of the DMI-HIRLAM G, N, E, and D numerical weather predictions. From (Sass et al., 2000).

The boundary conditions for DMI-HIRLAM-G are given by the ECMWF-model from the European Center for Medium-Range Weather Forecasts in England. The boundary conditions for the models N, E and D are given by one of the larger DMI-HIRLAM models (Sass et al., 2000).

When running the forecast system, the data assimilation is concluded by a model analysis run for the initial time epoch. Then the prediction cycle is initiated, and the continuous system of equations is solved for predefined discrete time steps, and a weather forecast is then output for example every 6 hours (Sass et al., 2000).

3.2 Zenith Delays Determined From DMI-HIRLAM

Tropospheric zenith delays can be derived from an NWP by numerical integration of the refractivity through the model, where the necessary meteorological parameters are given. The expression for refractivity, Equation (2.13), which is also given below as Equation (3.1), must, however, be rewritten in order to be used with a HIRLAM NWP.

$$N = k_1\left(\frac{P_d}{T}\right)Z_d^{-1} + \left(k_2\frac{e}{T} + k_3\frac{e}{T^2}\right)Z_w^{-1}. \tag{3.1}$$

First the ideal gas equation applied to dry air is introduced (Spilker, 1996):

$$P_d = \rho_d R_d T Z_d \tag{3.2}$$

P_d is the partial pressure of dry air, ρ_d is the density of dry air, R_d is the gas constant for dry air, and Z_d is the compressibility factor for dry air. Introducing this expression for P_d, the dry part of Equation 3.1 can be rewritten:

$$N_d = k_1 \rho_d R_d \tag{3.3}$$

whereby the dry compressibility factor is eliminated. The total refractivity is then given as:

$$N = k_1 R_d \rho_d + \left(k_2 \frac{e}{T} + k_3 \frac{e}{T^2}\right) Z_w^{-1} \tag{3.4}$$

R_d is the gas constant for dry air, ρ_d is the mass density of the dry constituents, T is the temperature, and e is the water vapor pressure. k_1, k_2, and k_3 are the refractivity constants discussed in Chapter 2.

Consider the ideal gas equation for water vapor (Spilker, 1996):

$$e = \rho_w R_w T Z_w. \tag{3.5}$$

Inserting this into the expression for N gives the following:

$$N = k_1 R_d \rho_d + k_2 R_w \rho_w + k_3 \frac{\rho_w R_w}{T}. \tag{3.6}$$

Hereby also the wet compressibility factor has been eliminated.
Now, by introducing the density of dry air (Wallace and Hobbs, 1977):

$$\rho_d = \rho - \rho_w \tag{3.7}$$

and rewriting Equation (3.6), the following expression is obtained:

$$\begin{aligned} N &= k_1 R_d \rho - k_1 R_d \rho_w + k_2 R_w \rho_w + k_3 \frac{R_w \rho_w}{T} \\ &= k_1 R_d \rho + k_2 R_w \rho_w - k_1 R_d \rho_w + k_3 \frac{R_w \rho_w}{T} \\ &= k_1 R_d \rho + R_w \rho_w \left(k_2 - k_1 \frac{R_d}{R_w} + \frac{k_3}{T}\right). \end{aligned} \tag{3.8}$$

The ratio ϵ is now introduced:

$$\epsilon = \frac{R_d}{R_w} \Leftrightarrow R_w = \frac{R_d}{\epsilon} \tag{3.9}$$

so

$$\begin{aligned} N &= k_1 R_d \rho + \frac{R_d}{\epsilon} \rho_w \left(k_2 - k_1 \epsilon + \frac{k_3}{T}\right) \\ &= k_1 R_d \rho + \frac{R_d}{\epsilon} \left(k_2 - k_1 \epsilon + \frac{k_3}{T}\right) \frac{\rho_w}{\rho} \rho. \end{aligned} \tag{3.10}$$

Introducing the specific humidity, q, given as (Mendes, 1999):

$$q = \frac{\rho_w}{\rho} \tag{3.11}$$

implies:

$$N = k_1 R_d \rho + \frac{R_d}{\epsilon}\Big(k_2 - k_1\epsilon + \frac{k_3}{T}\Big)q\rho. \tag{3.12}$$

This expression for the refractivity can now be used for determining the zenith total delay by integrating N along the signal path as shown with Equation (2.17). The zenith total delay, ztd, is then given as:

$$ztd = 10^{-6}\int_{h_{ant}}^{h_{top}} k_1 R_d \rho \, dh + 10^{-6}\int_{h_{ant}}^{h_{top}} \frac{R_d}{\epsilon}\Big(k_2 - k_1\epsilon + \frac{k_3}{T}\Big) q \, \rho \, dh \tag{3.13}$$

The integration is carried out from the altitude of the GPS antenna, h_{ant}, to the top of the neutral atmosphere, h_{top}. Contributions to the dry part of the zenith delay are present up to an altitude of about 80 km, and the integration is therefore more correctly carried out to this altitude level.

Since the vertical reference in HIRLAM is a pressure system, it is more convenient to work with the integral if the integration is carried out with respect to pressure instead of altitude. This is handled by introducing the hydrostatic equation as given by e.g. Wallace and Hobbs (1977):

$$\delta p = -\rho \, g \, \delta h \tag{3.14}$$

where g is gravity. The integral in (3.13) is rewritten since the integration is also more conveniently carried out in the opposite direction, i.e. from the top of the atmosphere to the surface of the Earth where the GPS antenna is located. The final expression, which is given by Vedel et al. (2001), can be used directly with a HIRLAM NWP:

$$ztd = 10^{-6}\int_{P_{top}}^{P_{ant}} \frac{k_1 R_d}{g} dp + 10^{-6}\int_{P_{top}}^{P_{ant}} \frac{R_d}{\epsilon}\Big(k_2 - \epsilon k_1 + \frac{k_3}{T}\Big)\frac{q}{g} dp \tag{3.15}$$

P_{ant} is the pressure at the GPS antenna, given in Pascal, P_{top} is the pressure at top of the neutral atmosphere, k_1, k_2, and k_3 are the constants from Table 2.1 given in Kelvin/Pa, R_d is the gas constant for dry air, g is gravity in m/s^2, ϵ is the ratio given in Equation (3.9), T is temperature given in Kelvin, and q is the specific humidity in kg/kg.

By introducing the hydrostatic equation the neutral atmosphere is assumed to be in a hydrostatic equilibrium. This is not the case in reality, but the approximation is used for most weather predictions and it is sufficient for the level of accuracy that will be expected in this context (H. Vedel, DMI, personal communication).

When implementing Equation (3.15), pressure, temperature and specific humidity are extracted from the NWP by horizontal and vertical interpolation. The height is determined from the pressure, and then gravity can be determined using, for instance, the expression for normal gravity given by Torge (1989).

3.3 Verification of HIRLAM Zenith Delays

Based on Equation (3.15) an NWP like DMI-HIRLAM can be used for estimating tropospheric zenith delays. But before implementing the method, the expected accuracy level for the estimated zenith delays should be evaluated, keeping in mind that the purpose is to find a method that is more accurate than the Saastamoinen global delay model, which has an average accuracy of 3 cm.

The accuracy of an NWP can be estimated by comparing the predicted meteorological parameters with surface observations. But where an accuracy measure of the predicted surface values is highly relevant for estimating the quality of the NWP as a basis for weather forecasts, it is not necessarily a good indicator of the quality of the meteorological conditions higher up in the atmosphere. Since any errors at higher levels in the NWP will also have an influence on the estimated zenith delays, evaluation by surface values is not sufficient for this purpose.

Radiosondes are, in this context, a better tool, since they provide profiles of the atmospheric conditions up to a level of about 40 km. The horizontal spacing of radiosonde launch sites is, however, too large to give a general idea of NWP accuracy within the geographical area of interest for this thesis, which is Denmark and Southern Sweden.

Zenith delays determined from an NWP can also be compared with zenith delays estimated from GPS data, when the GPS data is processed as described in Section 2.7. In the next section some examples are given where other authors have compared zenith delays determined from NWPs with zenith delays determined from radiosonde and/or GPS data. In Chapter 4 these findings are verified using the test data examined in this thesis.

3.3.1 Verification by radiosondes and GPS data

Comparison of tropospheric zenith delays determined using both NWP and GPS data has for instance been carried out as a part of the MAGIC project mentioned in Chapter 2.

Haase et al. (2001b) describe a comparison between tropospheric zenith delays determined from GPS data collected in 41 GPS stations and zenith delays determined from NWPs. The DMI-HIRLAM system was used in a modified version to generate NWPs for an area covering the western part of the Mediterranean, and it was run with a resolution of 0.3° in the horizontal, 31 vertical levels, and with predictions being generated every 6 hours.

The difference between the tropospheric zenith delays determined with GPS and HIRLAM had a bias within -10 to +15 mm, and the standard deviation was less than 25 mm in most cases. The comparison was carried out using 1 1/2 years of data. Haase et al. (2001b) also found a clear seasonal trend in the differences, with a larger standard deviation during the summer months, when the water vapor content in the atmosphere is higher. Furthermore, the standard deviation is correlated with station height, where a lower altitude, and thereby a larger water vapor content, implied a higher standard deviation.

Vedel et al. (2001) also used MAGIC data to compare zenith delays determined from DMI-HIRLAM with delays determined by both GPS and radiosonde data. These tests were also based on 1 1/2 years of GPS and HIRLAM-data and on all available radiosonde launches from the corresponding time period.

The paper concludes the following mean and standard deviations for the various differences: Delays from radiosondes minus delays from HIRLAM: -1.6 ± 14.7 mm. Delays from GPS minus delays from radiosondes: 6.0 ± 11.7 mm. Delays from GPS minus delays from HIRLAM: 3.2 ± 17.1 mm

Since the standard deviation for the comparison between GPS and radiosondes is smaller than for the two cases where HIRLAM is involved, this indicates that the zenith delays derived from the NWP are slightly poorer than the delays derived from radiosondes or GPS. However, for this comparison a three sigma clipping was applied to the GPS results, so GPS based zenith delays appearing as gross errors were not included in the comparison (H. Vedel, DMI, personal communication). Judging by these tests, it is estimated that zenith delays can be determined from the MAGIC version of HIRLAM with an accuracy of about 2 cm.

Similar tests have been carried out by other groups using other NWPs with high or medium resolution, and the results are comparable. For instance, Pany et al. (2001a) conclude, when using the ECMWF NWP with a 56 km grid spacing and 61 vertical layers, that slant delays from the NWP could be determined with an accuracy of 10–15 mm. Schueler et al. (2000b) used a global NWP, with a resolution of 1° x 1° from NOAA, and found that zenith wet delays could be determined with a global average RMS of 14 mm.

3.3.2 On the temporal and spatial resolution of water vapor

When using NWPs to estimate tropospheric delays for GPS positioning, the inaccuracy encountered is almost entirely due to the wet part of the delay. The high temporal and spatial variability of the distribution of water vapor in the atmosphere is one of the limiting factors in weather predictions (Gutman and Benjamin, 2001). Inaccuracies in the NWPs map into the derived zenith delays, and the temporal and spatial variations of the water vapor distribution are therefore also the limiting factors when using weather predictions to estimate tropospheric zenith delays.

The temporal variation of water vapor is characterized for instance by Jarlemark (1997), who showed some examples of very rapid changes, the size of about 10 cm, in zenith wet delay following the passage of a weather front. The results were based on data from the water vapor radiometer at Onsala Space Observatory in Sweden. Dodson et al. (1996) used another data set from Onsala and discuss a situation where the zenith wet delay dropped 6 cm within two hours. Considering that the size of the zenith wet delay is normally within 5–20 cm in the area, this is a very large variation over a very short time interval.

The spatial variation of water vapor is the limiting factor with respect to grid resolution when using NWPs for zenith delay estimation. Pany et al. (2001b) investigated the necessary horizontal resolution of an NWP to be used for estimating tropospheric delays for GPS positioning. Their conclusion is that the horizontal resolution of the NWP is not a relevant limiting factor for determining slant delays by ray tracing through a medium resolution weather model. A consistent error level is shown for NWPs with a grid spacing of 20 to 150 km.

Yang et al. (1999) reached almost the same conclusion. They found that the accuracy of water vapor distribution extracted from NWPs did not change significantly when using a 0.4° as compared to a 0.2° grid spacing, corresponding to about 30 or 15 km, respectively.

Yang et al. (1999) also looked at the temporal predictability of water vapor by using NWPs based on different prediction intervals. They concluded that even for 30 hour predictions, which was the longest prediction interval tested, the accuracy of the water vapor results did not deteriorate significantly.

Emardson and Johansson (1998) investigated techniques for interpolation of zenith wet delays between GPS stations. They used five months of GPS data (July - November) from three reference stations, and tested interpolation techniques to estimate wet delays for a fourth and fifth station. The best of the interpolation techniques showed that interpolation of zenith wet delay can be carried out with an accuracy of 1 cm when the spacing between the reference stations is about 100 km. Since the summer months, with their high variability in water vapor, are included, the results indicate that a 100 km NWP grid spacing might be sufficient to estimate the zenith wet delay with high accuracy if an appropriate interpolation technique is used.

Several authors, e.g. Haase et al. (2001b), Walpersdorf et al. (2001), and Gutman and Benjamin (2001), mention that zenith delays estimated from NWPs are more accurate for inland sites than for locations close to the coast, where there is more water vapor in the air.

3.3.3 DMI-HIRLAM as a basis for zenith delays

The investigations described above, where the HIRLAM model was used in connection with the MAGIC project, indicate that this model can be used to estimate zenith delays with an accuracy of about 2 cm, and thus with an accuracy that is better than the accuracy for zenith delays determined with the Saastamoinen global delay model.

An improvement in tropospheric zenith delay estimates of about 1 cm does not look like much, even though it corresponds to 30%. But it is important to note that the accuracy of zenith delays determined from a high resolution NWP will be more consistent in the sense that a zenith delay determined using HIRLAM will also be good within a few cm during abnormal weather conditions, which are not represented sufficiently well by the climate data. In those situations a delay determined from a global delay model, which is based on climate data, will be more likely to provide erroneous estimates of the zenith delays than an NWP.

An improvement in GPS positioning performance is therefore expected if NWPs are used for estimating zenith delays, rather than using global tropospheric delay models. This expectation is tested in the following two chapters.

4

HIRLAM ZENITH DELAYS FOR GPS POSITIONING

This chapter describes a series of tests that have been carried out in order to evaluate the suitability of zenith delays derived from DMI-HIRLAM-E for various types of GPS positioning. Initially the data sets used for the tests are described. This is followed by a verification of the HIRLAM zenith delays. Tests and results from using the HIRLAM derived zenith delays for static and kinematic post–processed positioning, and finally for network RTK, are then described.

4.1 Description of the Test Data Sets

All the tests carried out in this, and the following chapter, are based on the same set of NWPs and GPS data, which is described in the following section.

4.1.1 DMI-HIRLAM-E

Data from DMI-HIRLAM-E was made available by the Danish Meteorological Institute (DMI). The NWP covers most of Europe, and has a grid spacing of 0.15° x 0.15° corresponding to approximately 16 x 9 km in latitude and longitude, respectively, for the region of interest. Vertically, the model consists of 31 layers, extending from the surface of the Earth up to an altitude of about 30 km. The distance between the layers is small close to the surface of the Earth, and increases with increasing altitude.

Data was available from two different days with a relatively low ionospheric activity: September 5 and 14, 2000. For the two days analysis fields were available at midnight and noon UTC, and in-between the analyses 1, 2, 3, 4, 5, 6, 9, and 12 hour predictions were generated. This means that for 1:00 hours UTC a 1-hour prediction was available, for 2:00 hours UTC a 2-hour prediction was available etc., until the new analysis was generated at noon. For 13:00 hours UTC a new 1-hour prediction was available etc. Local time in the area is one hour ahead of UTC.

The HIRLAM data was provided in the binary GRIB format, and was decoded using the PICOGRIB software, which is available as freeware on the internet from http://sourceforge.net/projects/picogrib/.

A complete DMI-HIRLAM-E field[1] is voluminous, and the fields were therefore limited in horizontal extension by considering the size of the area that would be covered by a GPS signal received at a 5° elevation angle at any azimuth. The resulting horizontal extension of the limited DMI-HIRLAM-E model, which is used for the following tests, is 50°–65° North in latitude, and 1° West to 26° East in longitude. This size of the limited test fields is driven by the ray tracing tests that are described in Chapter 5. When working with zenith delays a much smaller extension of the model can be used.

---

[1] A weather field is here a complete prediction including all the predicted parameters for the entire grid, for a given instant in time.

Atmospheric conditions on the test days

Both days were characterized by low global ionospheric activity, the main reason for selection of these two days. On September 5 the Kp-index was 2-3, and on September 14 it was between 1 and 3. The Ap-index[2] was 7 and 4, respectively. The ionospheric activity for the two days is discussed in detail in Appendix A.

The weather, and thus the tropospheric activity, was moderate on September 5, with a light wind, scattered showers, and a temperature of 10 to 20°C. On September 14 there was slightly more activity with a low-pressure system moving through the area causing more wind and more rain. The temperatures were again within 10 to 20°C.

4.1.2 GPS data

Data from 11 permanent GPS stations in the SWEPOS network was made available by Lantmäteriverket, the National Land Survey of Sweden. Further data from three permanent GPS stations in Denmark was made available by Kort & Matrikelstyrelsen (KMS), the National Survey and Cadastre - Denmark. The locations of the stations are shown in Figure 4.1.

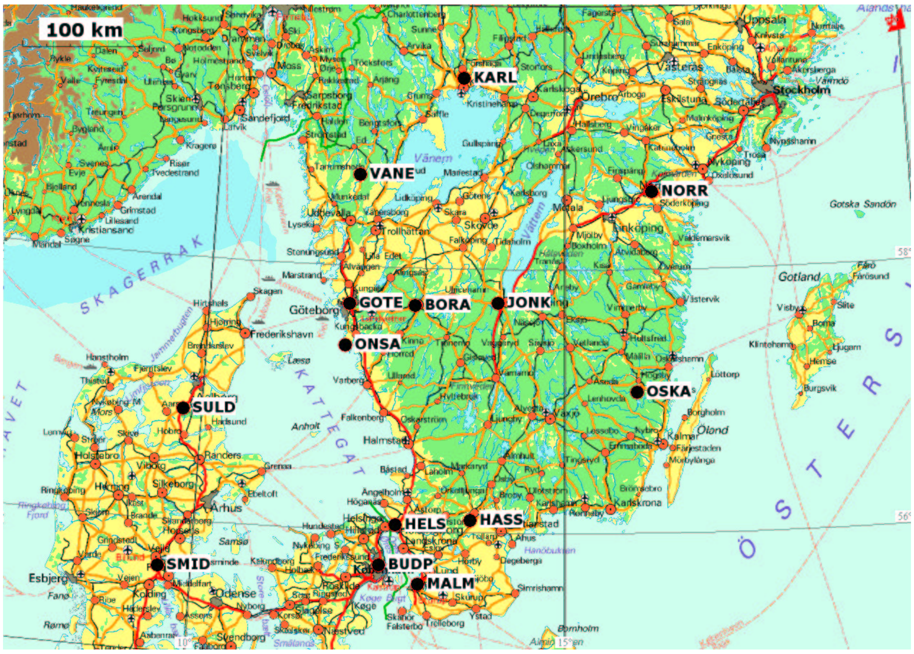

Figure 4.1: Locations of the 14 GPS stations.

The Swedish data was collected using Ashtech Z-XII dual frequency receivers and Dorne Margolin T and B antennas, and the Danish stations were equipped with Ashtech Z-XII dual frequency receivers and Ashtech antennas with choke rings.

All receivers were tracking GPS signals down to a 5° elevation angle, and the data was given with a 15 second data rate. In Appendix B, a list of station coordinates and height information are given along with a discussion of the data quality. The data quality is generally good, and the Danish stations seem to be less affected by multipath than the Swedish stations.

---

[2] The ionospheric Kp and Ap indices are described in Appendix A.

Accurate coordinates for the GPS stations are necessary in order to determine zenith delays based on the GPS data, so the Bernese software version 4.2 was used, along with precise orbits and one week of data (September 3–9, 2000) from all the stations to determine station coordinates in ITRF97. The coordinates for the IGS station ONSA were constrained to their IGS values, whereby the resulting coordinates for the other stations are in concordance with the reference frame used for the IGS orbits.

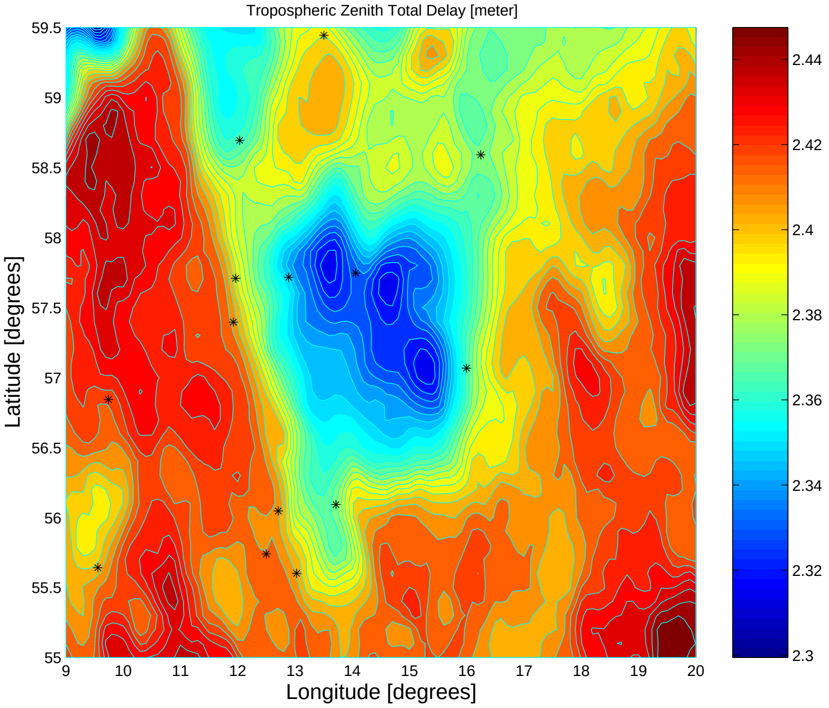

Figure 4.2: Zenith total delay determined from DMI-HIRLAM-E, September 5, 2000, at 01:00 UTC. The black stars indicate the locations of the GPS stations.

4.2 *Verification of Zenith Delays from HIRLAM*

The weather data given by DMI-HIRLAM-E was now used for determining tropospheric zenith delays for the GPS stations. This was carried out using the program ZTDPOINT, which is based on Equation (3.15). Implementation of Equation (3.15) is not straight forward, as indicated by the procedure described below:

The predicted meteorological parameters, temperature and humidity, are determined by horizontal interpolation in each layer of the NWP. The height, and thus the gravity, is determined from the pressure given for each of the NWP layers. The contribution to the delay from atop DMI-HIRLAM is determined using an expression given by Vedel et al. (2001), and the contribution to the delay from the nearest layer of DMI-HIRLAM to the GPS antenna is determined by a vertical interpolation/ extrapolation procedure. All these issues are described in detail in Appendix C along with descriptions of the necessary horizontal and vertical coordinate conversions, discussions on spherical or ellipsoidal pressures shells, and the use of NWP analysis versus prediction data.

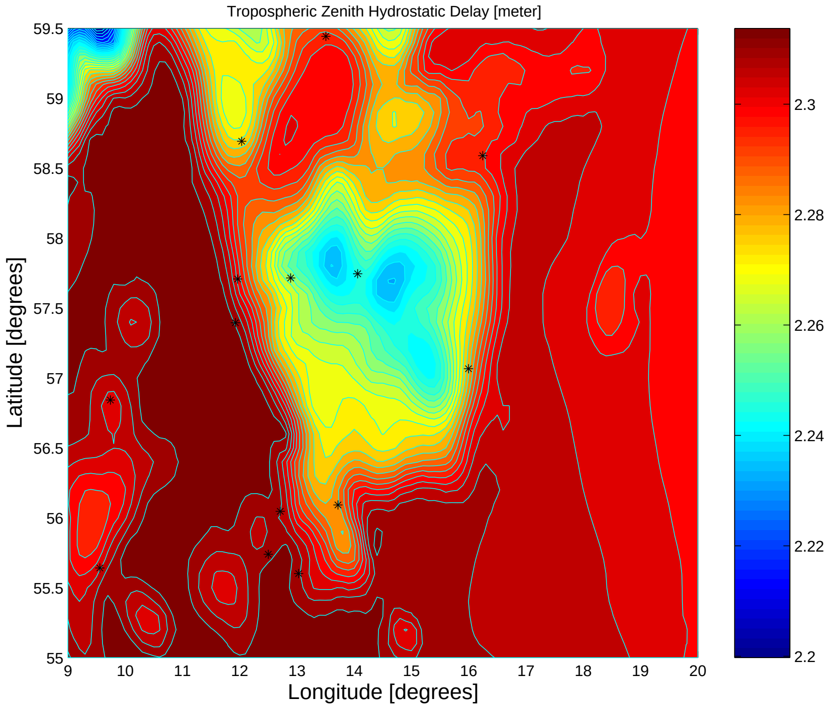

Figure 4.3: Zenith hydrostatic delay determined from DMI-HIRLAM-E, September 5, 2000, at 01:00 UTC. The black stars indicate the locations of the GPS stations.

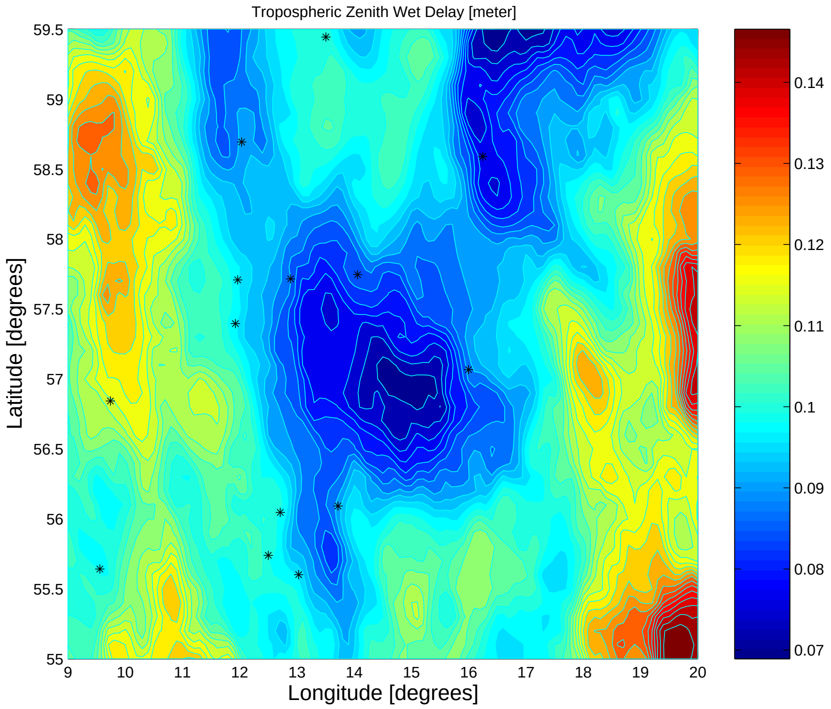

Figure 4.4: Zenith wet delay determined from DMI-HIRLAM-E, September 5, 2000, at 01:00 UTC. The black stars indicate the locations of the GPS stations.

Using ZTDPOINT, zenith delays were determined for the 14 GPS stations at all instances in time where DMI-HIRLAM-E analysis or prediction data was available from September 5 and 14.

The size and variation of the zenith total delay, the zenith hydrostatic delay, and the zenith wet delay for the area of interest are shown in Figures 4.2, 4.3, and 4.4.

The correlation between hydrostatic delay and the height above sea level can be seen from Figure 4.5 where the orthometric height for the surface model used for generating DMI-HIRLAM-E is shown for the same geographical area.

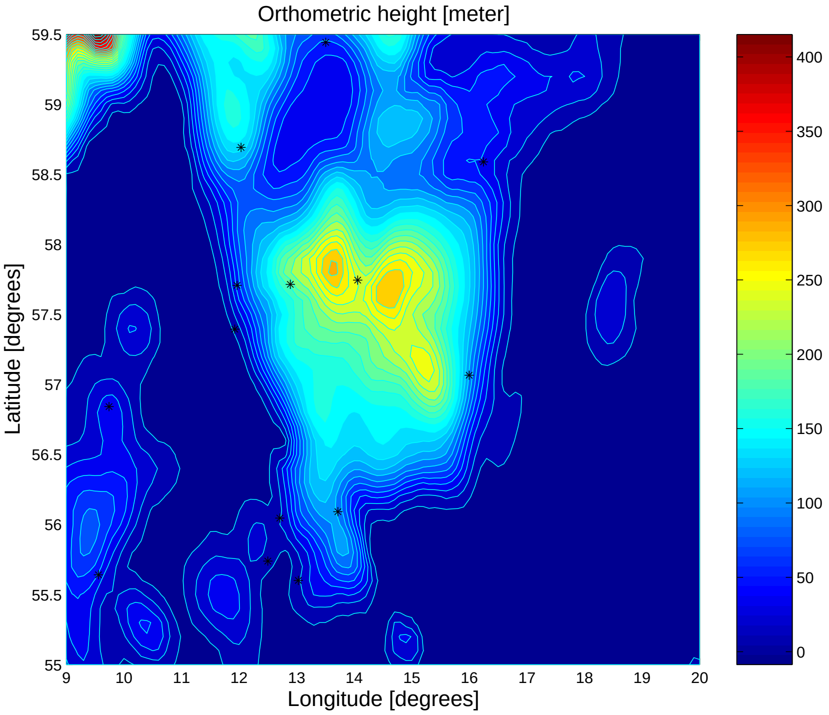

Figure 4.5: orthometric height for the surface model used in DMI-HIRLAM-E. The black stars indicate the locations of the GPS stations.

The surface temperature and specific humidity[3] from the same NWP field is plotted in Figures 4.6 and 4.7, respectively. Note that the resolution of the specific humidity is very coarse because the parameter is only given with one significant digit in the NWP files. A discussion on this is found in Appendix C.

In Figures 4.8 and 4.9, the zenith total delays are given for each station throughout the two days. The stations are listed according to orthometric height, and the heights are given in Appendix B. The size of the zenith delays is different for the two days, and the difference is correlated with the higher variability and larger amount of water vapor in the neutral atmosphere on September 14. The considerable decrease in zenith delay for station BUDP between 6:00 and 9:00 UTC on September 14 is caused by a decrease in specific humidity and surface pressure in the NWP.

---

[3] The surface values of temperature and specific humidity are actually given for a height of 2 meters above the surface.

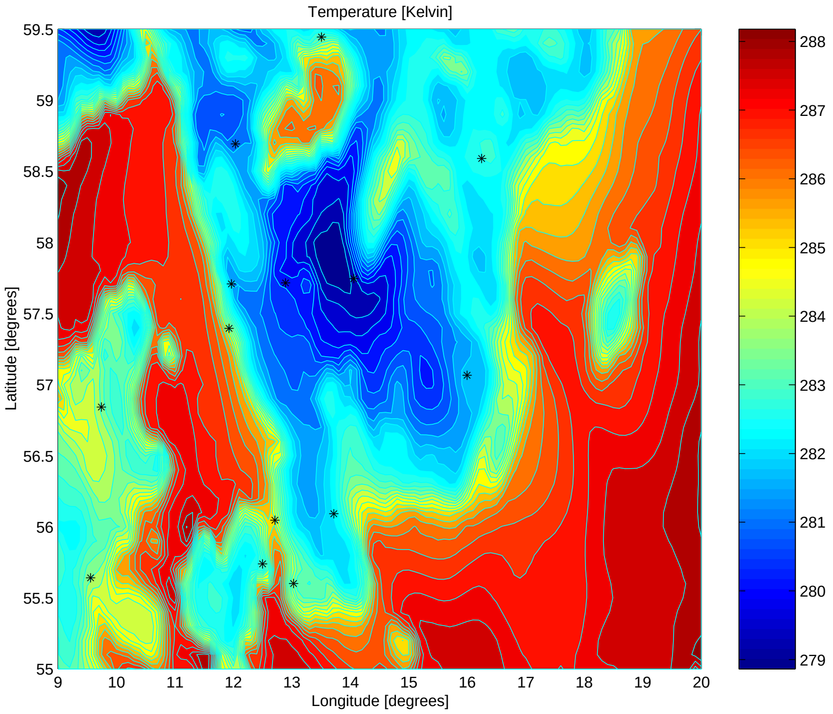

Figure 4.6: Surface temperature from DMI-HIRLAM-E, September 5, 2000, at 01:00 UTC. The black stars indicate the locations of the GPS stations.

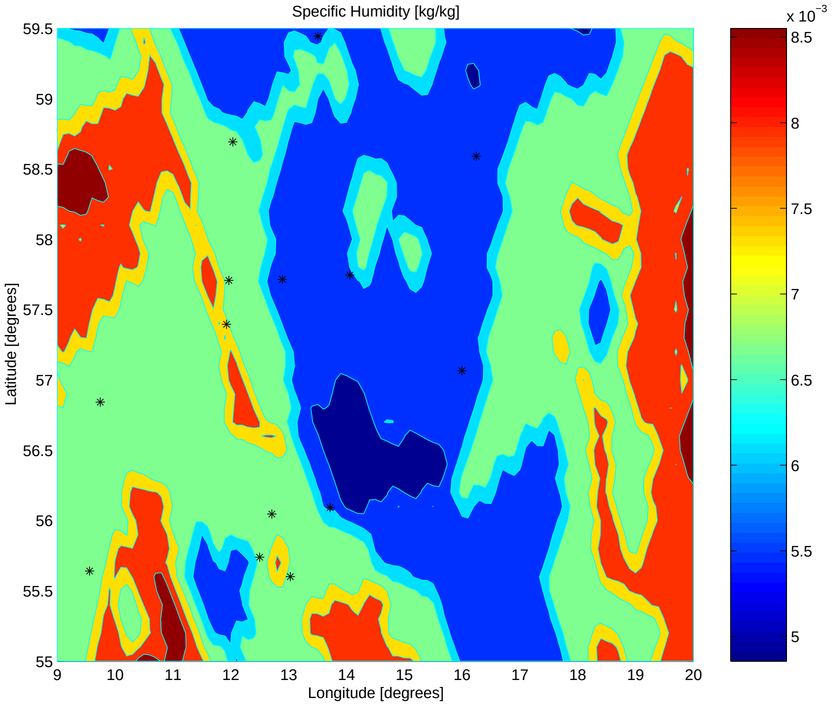

Figure 4.7: Surface specific humidity from DMI-HIRLAM-E, September 5, 2000, at 01:00 UTC. The black stars indicate the locations of the GPS stations.

4.2.1 Comparison with Bernese

Before introducing the HIRLAM zenith delays in a GPS positioning process, it is appropriate to evaluate the quality of the delays determined with ZTDPOINT. In

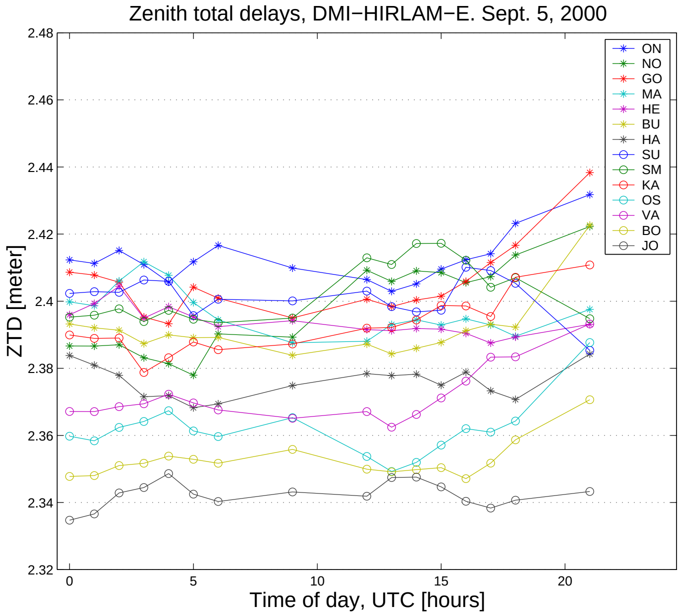

Figure 4.8: Zenith total delays from DMI-HIRLAM-E, September 5, 2000. Stations listed according to altitude.

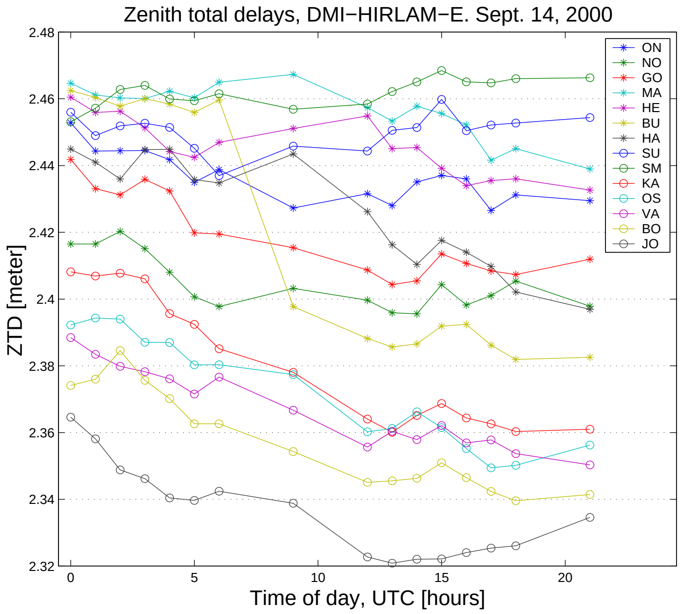

Figure 4.9: Zenith total delays from DMI-HIRLAM-E, September 14, 2000. Stations listed according to altitude.

this section the zenith delays determined using the HIRLAM NWPs are evaluated by a comparison with GPS-derived zenith delays.

Bernese processing of the GPS data from the two days was carried out using the Saastamoinen global tropospheric delay model to generate a priori values for the tropospheric delay. The residual tropospheric effect was estimated as a part of the adjustment process, and by adding it to the Saastamoinen delays, the total tropospheric delay for the given station and time epoch was obtained. The total tropospheric delays were estimated by Bernese for every hour.

In order to avoid any residual effects from other error sources in the positioning process, it is important that these effects are handled appropriately. The noise and multipath level is expected to be low because of the high quality receivers and antennas used. The residual orbit error is negligible for these baseline lengths when final precise orbits from the IGS are used. The station coordinates are in concordance with the orbits, as previously mentioned, so no geometrical misalignment, caused by reference frame mix up, is introduced. The Earth tide is modelled, and the ionosphere is handled by introducing the ionosphere–free linear combination of the observables. Any remaining error effects should thus be caused by either differential ocean loading effects, that are expected to be negligible, or by any higher order ionospheric effects, that are also expected to be negligible because of the low ionospheric activity.

A 10° elevation mask was used for the processing. Selection of elevation mask is, in this connection, a compromise between eliminating some of the noisy low elevation data, and introducing a bias in the results. The presence of a bias has been illustrated for instance by Emardson et al. (1998), where the influence of the GPS elevation mask on water vapor estimated from zenith wet delays derived from GPS data was shown. When they changed the elevation mask from 15° to 10° an offset of about 2 kg/m^2 in integrated water vapor, corresponding to approximately 1.3 cm in wet zenith delay, was almost removed[4].

With a lower elevation mask, for instance 5°, the amount of noise is increased and difficulties in the positioning process would be expected.

The zenith total delays were thus determined for the 14 stations using both GPS data and DMI-HIRLAM-E, and the difference between the solutions was calculated. These differences are plotted in the Figures 4.10 and 4.11 for the two days respectively. Time of day in hours is given on the x-axis.

On September 14 station BUDP appears as an outlier after 9:00 hours UTC, when the size of the difference increases considerably. The jump originates from the decrease in NWP zenith delay also shown in Figure 4.9, and no corresponding decrease is seen from the GPS results.

When station BUDP is not included in the statistics the mean of the differences for September 14. is -0.007 meters, the standard deviation is 0.012 meters, and the RMS is reduced to 0.014 meters.

In Tables 4.1 and 4.2 statistics are given for each time epoch throughout the test days. Note that station BUDP is not included in the results for the afternoon of September 14.

For DMI-HIRLAM-E, data assimilation is carried out twice daily, at midnight and noon UTC, and it is interesting to notice that the magnitudes of both the mean and the standard deviation of the differences vary before and after data assimilation. This is an indication of the influence of NWP data assimilation on the following NWP predictions.

12 hour predictions were available at 12:00 UTC for both days along with the analysis, and a comparison of zenith delays determined using the analysis and the prediction is given in Appendix C.

---

[4] The relationship between zenith wet delay, zwd, and integrated water vapor, IWV, can be approximated by: ZWD = 6.5 * IWV (Jarlemark, 1997).

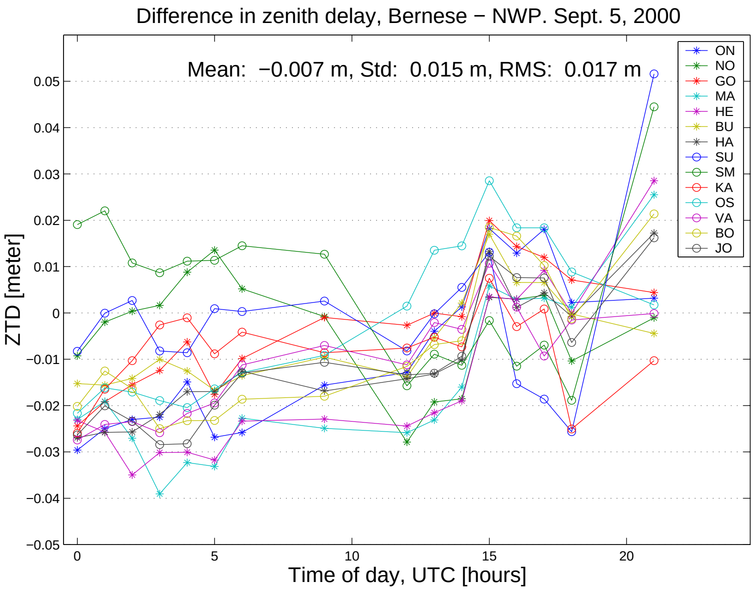

Figure 4.10: Difference between ZTDs determined with the Bernese software and derived from HIRLAM.

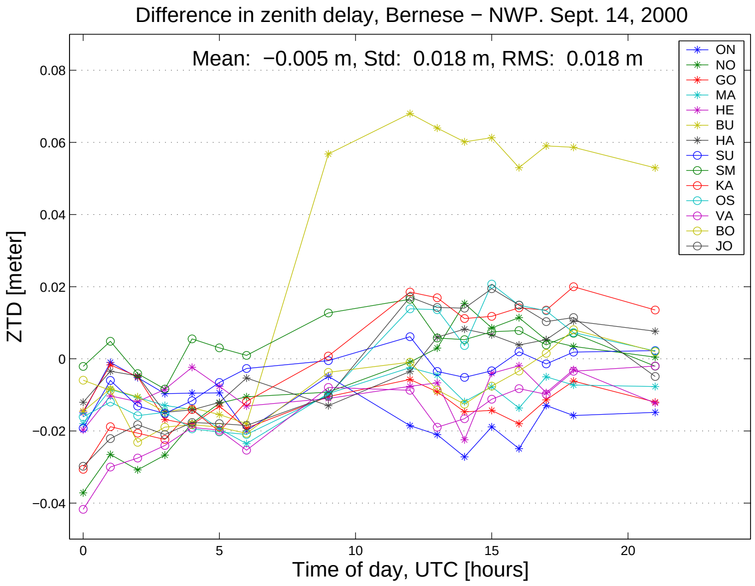

Figure 4.11: Difference between ZTDs determined with the Bernese software and derived from HIRLAM.

Generally, the overall RMS of the differences is better than 2 cm for both days and thus in concordance with the accuracy expected based on the analyses referenced in Chapter 3.

The difference between the total zenith delays determined with Bernese and the delays determined with the Saastamoinen model are given in Table 4.3. The RMS values are in agreement with the 3 cm accuracy of the Saastamoinen model as estimated by Mendes (1999).

Both the mean and standard deviation are considerably larger when comparing

UTC hour of day	Mean [meter]	Std. Dev. [meter]	RMS [meter]
00:00	-0.019	0.013	0.022
01:00	-0.014	0.013	0.019
02:00	-0.016	0.012	0.020
03:00	-0.016	0.014	0.021
04:00	-0.013	0.015	0.019
05:00	-0.014	0.016	0.020
06:00	-0.010	0.012	0.015
09:00	-0.010	0.010	0.014
12:00	-0.015	0.009	0.017
13:00	-0.009	0.010	0.013
14:00	-0.007	0.010	0.012
15:00	0.011	0.008	0.014
16:00	0.005	0.008	0.010
17:00	0.006	0.008	0.010
18:00	-0.004	0.009	0.010
21:00	0.010	0.015	0.018

Table 4.1: Mean, standard deviation, and RMS of difference between GPS and HIRLAM zenith delays for the 14 stations. Results from September 5, 2000.

UTC hour of day	Mean [meter]	Std. Dev. [meter]	RMS [meter]
00:00	-0.020	0.011	0.023
01:00	-0.012	0.011	0.016
02:00	-0.016	0.010	0.018
03:00	-0.017	0.006	0.018
04:00	-0.014	0.007	0.015
05:00	-0.014	0.006	0.015
06:00	-0.015	0.007	0.017
09:00	-0.002	0.018	0.018
12:00	0.001	0.012	0.011
13:00	0.000	0.012	0.012
14:00	-0.002	0.015	0.015
15:00	0.001	0.013	0.012
16:00	0.001	0.013	0.013
17:00	0.001	0.009	0.009
18:00	0.003	0.010	0.010
21:00	-0.002	0.008	0.008

Table 4.2: Mean, standard deviation, and RMS of difference between GPS and HIRLAM zenith delays for the 14 stations. Results from September 14, 2000. Station BUDP excluded from afternoon results.

	Mean [meter]	Std. Dev. [meter]	RMS [meter]
September 5	0.014	0.018	0.023
September 14	0.040	0.032	0.051

Table 4.3: Mean, standard deviation and RMS of difference between GPS and Saastamoinen zenith delays for the two GPS data sets.

the Saastamoinen model with the Bernese delays, than when the HIRLAM delays are compared to Bernese. Consequently, on the basis of these two data sets, HIRLAM zenith delays seem to be considerably better than delays determined with the Saastamoinen model.

4.2.2 Correcting GPS data with HIRLAM zenith delays

The use of NWP zenith delays was tested for static positioning using the GIPSY software by Behrend et al. (2001). They obtained an improvement in positioning performance, but only when the residual tropospheric errors were not estimated in the positioning process.

The major advantage of using NWP derived zenith delays is therefore expected to be in connection with commercial software packages where the residual tropospheric effect is not estimated. In such software the tropospheric delay is only handled by a global tropospheric delay model followed by double differencing, and any residual tropospheric effects will therefore propagate into the height component of the position determined for the roving GPS receiver.

The tests described in the following were carried out using the commercial software package GPSurvey from Trimble Navigation Limited, and a beta version of the NetCor software from the University of Calgary, which forms the basis for the commercial MultiRef software for network RTK. The disadvantage of working with this type of software is that the source code is not available, so the NWP tropospheric correction approach cannot be implemented in the source code. It is thus necessary to apply the NWP derived tropospheric delays to the data before feeding it into the processing software.

This was accomplished by initially determining the zenith hydrostatic and wet delays for each GPS station using all NWPs available. Since the NWPs are given each full hour, a linear temporal interpolation was carried out in order to determine zenith delays for every 15 seconds. The elevation angle of each observed satellite signal was determined, and the NMF hydrostatic and wet mapping functions (Niell, 1996) were used to map the zenith delay down to the appropriate elevation angle. Finally these slant delays were subtracted from the raw code and phase observations given in the GPS data files.

Hereby a set of tropospherically corrected GPS data files, that can be read by the two software packages, were generated.

This way of dealing with the tropospheric delay can cause an increased level of detected cycle slips when the data is cycle slip screened by the processing software. At the lowest elevation angles the change in tropospheric delay within 15 seconds can be larger than the wavelength for L1 or L2. For example, the speed of a GPS satellite is 4 km/s, corresponding to an angular velocity of approximately 0.009° pr. second, or 0.13° for 15 seconds. The mapping factor, determined with any of the mapping functions previously described, changes with about 0.2 when the elevation angle of the received satellite signal is changed with 0.1° from 5° to 5.1°. With a zenith delay of 2.4 meter, the change in total tropospheric delay is then approximately 0.2 x 2.4 meters = 0.5 meter, within the 15 seconds. Thus a jump of more than two cycles is introduced in the phase observations, and if the cycle slip detection function in the software is not tolerant to this kind of "dynamics" in the phase observations, the jump will be interpreted as a cycle slip. The purpose of the following tests is to investigate whether the HIRLAM zenith delays are better than zenith delays determined using a global tropospheric delay model. So in order to avoid problems with the increased level of cycle slips for low elevation angles, all tests in the rest of this chapter were carried out using a 15° elevation mask.

4.3 HIRLAM Zenith Delays for Static Positioning

In order to test the impact of using the HIRLAM zenith delays instead of a global delay model, the GPSurvey software version 2.35 (Trimble, 1996) was used as an example of a commercial software package.

Static GPS positioning was carried out using precise orbits and the ionosphere–free linear combination of the L1 and L2 observables to resolve the L1 ambiguities. Furthermore the processing was carried out using both the tropospherically corrected data and the raw data files. In the latter case the Saastamoinen global delay model was activated in the programme. The ONSA station, located in the middle of the network, was kept fixed, and the rest of the stations were treated as rover stations. See Figure 4.12.

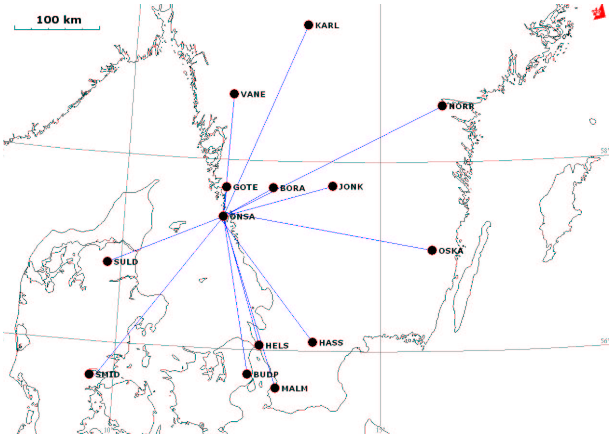

Figure 4.12: Baselines processed with GPSurvey.

The processing was carried out using two sets of 6 hour data for each day, the morning data set from 0:00 to 6:00 hours UTC, and the afternoon data set from 12:00 to 18:00 hours UTC.

The positions determined using the two sets of data files were compared with the known station coordinates determined using the Bernese software, as described in Section 4.1. The coordinate differences were then used to determine the 3D error, σ_{3D}, as shown below

$$\sigma_{3D} = \sqrt{dX^2 + dY^2 + dZ^2} \tag{4.1}$$

This was done for both the NWP and the Saastamoinen solutions, and then by subtracting the NWP σ_{3D} from the Saastamoinen σ_{3D} a measure of the improvement was obtained.

Figures 4.13 and 4.14 show the improvement in position accuracy for the 13 baselines and the two data sets from September 5. Negative values imply that the Saastamoinen model gave better results.

With the morning data set from September 5, the ambiguities for three of the baselines (ON-BU, ON-MA, ON-SM) were not fixed when using the Saastamoinen model, and when using the NWP approach, ambiguities for four of the baselines were not fixed (ON-HE, ON-BU, ON-MA and ON-SM). When the ambiguities are not fixed to integers, but are left at real floating values, the position solution becomes more unreliable, since floating ambiguities are not constrained in the same way as integer values are by the fact that they can only vary with integer numbers.

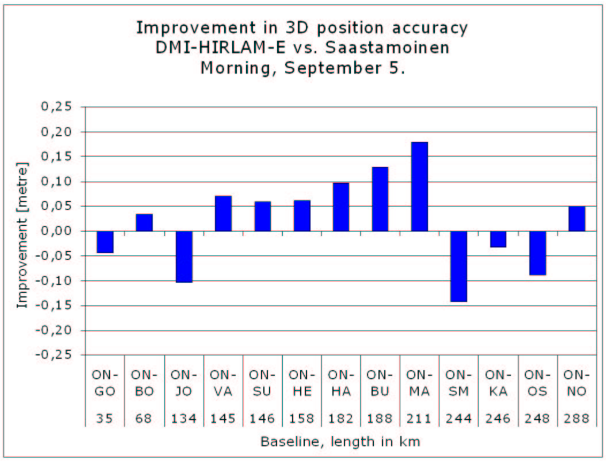

Figure 4.13: Improvement in 3D position accuracy using the NWP approach versus the Saastamoinen model. September 5, 2000, from 0:00 to 06:00 UTC.

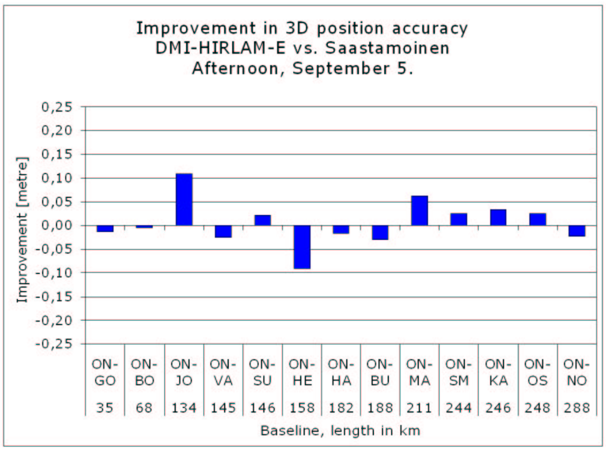

Figure 4.14: Improvement in 3D position accuracy using the NWP approach versus the Saastamoinen model. September 5, 2000, from 12:00 to 18:00 UTC.

Figures 4.15 and 4.16 show the improvement in position accuracy using the two data sets from September 14. The ambiguities were fixed for all these solutions.

Note that it was not possible to process the afternoon data from station SULD on September 14. Also note the large negative improvement for station BUDP with the afternoon data on September 14. This is directly related to the inaccuracy of the HIRLAM zenith delays as previously discussed and shown in Figure 4.11.

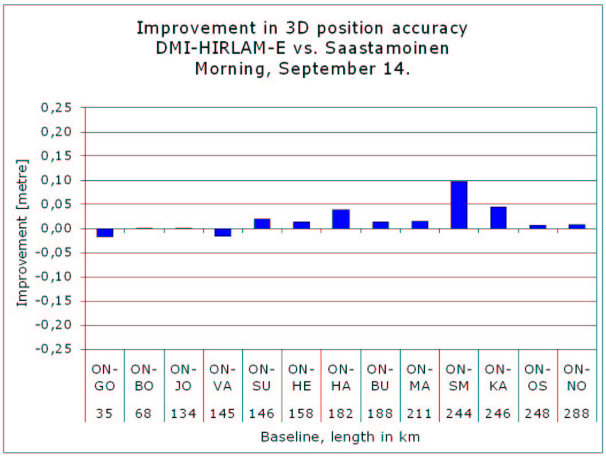

Figure 4.15: Improvement in 3D position accuracy using the NWP approach versus the Saastamoinen model. September 14, 2000, from 0:00 to 06:00 UTC.

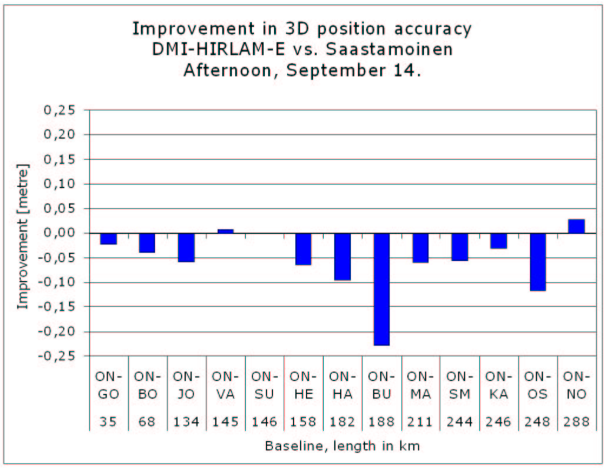

Figure 4.16: Improvement in 3D position accuracy using the NWP approach versus the Saastamoinen model. September 14, 2000, from 12:00 to 18:00 UTC.

The results in these figures can be compared with the estimated accuracy of the zenith delays given in Tables 4.1 and 4.2.

Figures 4.15 and 4.16 from September 14 show a difference in the results using the morning and afternoon data, but there is also a considerable difference in the mean of the zenith delays shown in Table 4.2. In the morning, the mean of the differences is about -1.5 cm, and in the afternoon it is only a couple of mm.

According to Zhang (1999) the residual tropospheric error is generally larger than multipath and receiver noise when the GPS baselines are longer than 100 km. An improved tropospheric correction approach is therefore not expected to have any influence for baselines shorter than about 100 km.

It was expected, however, that the improvement would increase with an increasing baseline length or an increasing height difference between reference and rover. The improvements in 3D position accuracy are listed according to baseline length in the figures, and no indication of a correlation of improvement with baseline length can be seen from the diagrams.

Considering the height differences there is no trend in the results either. The largest height differences are for the baselines from ONSA to BORA and ONSA to JONK, and they do not show a larger improvement than the rest of the baselines. A height difference of about 200 meters is, however, not significant, and when the baselines are only 68 and 134 km in length, as in the present case, the difference in atmospheric conditions is not large enough to cause a significant improvement by using NWPs rather than the Saastamoinen model.

The previous four plots all show the improvement in 3D position accuracy, but as mentioned in Chapter 2 the size of the tropospheric delay has a larger influence on the height component than on the horizontal components of the position solution. The improvement in height determination was therefore also analyzed. The results for the height are, however, similar to the 3D position results, and they are not discussed separately.

4.4 HIRLAM Zenith Delays for Kinematic Positioning

Considering the results obtained for the static GPS positioning a larger improvement is expected for kinematic positioning, because in this type of positioning only one data epoch is used for determining the position of the rover. A good initial estimate of the tropospheric delay is more important since the effect of any residual errors are not as easily minimized by the adjustment process as in static positioning. It is therefore interesting to test whether any improvement is seen when processing the data in kinematic mode.

Tsujii et al. (2001) did, as mentioned in Chapter 2, investigate how “on the fly” (OTF) ambiguity resolution was affected by NWP derived zenith delays. They only used one hour of data, but the results showed a large improvement in the number of fixed ambiguities.

Here a kinematic scenario was simulated using the static raw and tropospherically corrected data sets from the 14 GPS stations. The GPSurvey software was configured to process the data in kinematic mode, using OTF ambiguity resolution. Otherwise the same processing parameters were used as in the static case, and again station ONSA was fixed.

Before analyzing the results it should be noted that this is not the ideal way of testing kinematic positioning performance. Most GPS software, when used for kinematic positioning, contain filters that track the dynamics of the roving GPS antenna by monitoring the phase rate of change. How these filters, if they are implemented in GPSurvey, correspond to static data is not known by the author, but it is anticipated that they will react in the same way to the raw and tropospherically corrected data sets used here, whereby it should be possible to compare the two sets of position results.

Processing was carried out using the same 6 hour data sets as for the static tests. With the 15 second data interval a total of 1440 positions were determined for each data set. These positions were compared with the “true” positions determined with the Bernese software, and the 3D RMS of the position differences are given in Tables 4.4 and 4.5. The baselines are listed according to baseline length, and the lengths are the same as given in Figure 4.16.

The RMS of the 3D coordinate differences are much higher than when processing static data. This is typical for kinematic positioning, because only one epoch of data is available for each position to be determined. The position solution is therefore much more sensitive to multipath and cycle slips, and it is not uncommon to find ambiguities that are fixed to wrong integer values, implying a decrease in position accuracy when processing kinematic data for longer baselines. Previous experience with GPSurvey has shown that this software does have difficulties in solving ambiguities for longer baselines in kinematic mode (Jensen, 1999) and (Jensen and Cannon, 2000).

RMS of 3D Coordinate Differences in cm

September 5, 2000	Morning		Afternoon	
Baseline	Saastamoinen model	HIRLAM and NMF	Saastamoinen model	HIRLAM and NMF
ONSA-GOTE	8.5*	11.8	13.1	4.9
ONSA-BORA	43.8	10.7	23.8	8.4
ONSA-JONK	7.0	12.8	20.4	18.9
ONSA-VANE	17.3	17.7	24.6	10.6
ONSA-SULD	18.9*	48.3	19.7	25.3
ONSA-HELS	45.4	40.7	31.1	19.6
ONSA-HASS	64.9*	60.4*	23.3	16.8
ONSA-BUDP	47.2*	28.6	23.7	15.8
ONSA-MALM	80.8*	64.3	23.1	14.9
ONSA-SMID	17.2	21.1	21.0	22.9
ONSA-KARL	48.0*	25.4*	24.3	23.2
ONSA-OSKA	9.2	11.2	37.8*	24.6
ONSA-NORR	43.1	21.4	31.0	24.0

Table 4.4: RMS of 3D coordinate differences for 2 x 6 hours of kinematic positions, September 5, 2000. $*$ indicates a partly floating solution.

When analyzing Table 4.4 is it seen that for 18 of the 26 kinematic processings listed, the 3D position error is smaller when using the HIRLAM approach rather than the Saastamoinen model. Also, the number of float solutions is reduced considerably when HIRLAM is used. In particular with the results for the afternoon data, there is a general improvement when using the HIRLAM zenith delays as compared to the Saastamoinen delays. These good results are explained by the small standard deviation of the zenith delays shown in Table 4.1. Also note from the afternoon data that there is an indication of a correlation with baseline length.

For the morning data set from September 5 there are more solutions with floating ambiguities than for the other data sets. This could be caused by higher order ionospheric effects that are not treated with the ionosphere–free linear combination. In Appendix A, Figure A.6 shows that there was a relatively high ionospheric activity the night before, on September 4. Even though the geomagnetic field is not affected in general, as shown in Figure A.8, there are a few fluctuations in the field strength of the observations from the Brorfelde observatory around 3:00–4:00 hours UTC on September 5, and these could be leftovers from the activities the night before.

The positioning results for the meteorologically more turbulent day, September 14, are not as good, and as shown in Table 4.5 better position results are only obtained with the NWP for 5 of the 25 processings. The ambiguities have, however, been fixed in all cases where the HIRLAM approach is used, but this does not change the fact that the results in the position domain for September 14 show a general degradation in position accuracy when using the HIRLAM approach.

RMS of 3D Coordinate Differences in cm

September 14, 2000	Morning		Afternoon	
Baseline	Saastamoinen model	HIRLAM and NMF	Saastamoinen model	HIRLAM and NMF
ONSA-GOTE	5.1	6.6	4.8	9.0
ONSA-BORA	10.2	8.4	2.7	11.0
ONSA-JONK	5.9	6.2	2.6	10.7
ONSA-VANE	8.6	8.8	10.8	19.0
ONSA-SULD	5.1	8.8	-	-
ONSA-HELS	9.8	13.3	3.3	10.4
ONSA-HASS	16.8*	18.2	2.5	14.8
ONSA-BUDP	2.5	2.4	2.4	26.0
ONSA-MALM	14.0*	12.7	3.0	9.8
ONSA-SMID	39.1	15.7	3.2	10.9
ONSA-KARL	16.7	26.3	34.4	21.5
ONSA-OSKA	3.5	7.9	24.4	25.0
ONSA-NORR	19.5	26.7	17.3	30.0

Table 4.5: RMS of 3D coordinate differences for 2 x 6 hours of kinematic positions, September 14, 2000. $*$ indicates a partly floating solution.

The Saastamoinen model is used with standard meteorological parameters in GPSurvey, so there is no temporal variation of the zenith delays, even if the weather conditions are turbulent. This is not the case for the NWP derived zenith delays, which will change directly with changes in the weather conditions. The variability of the NWP zenith delays, caused by the more active atmospheric conditions, might be causing disturbances in the GPSurvey positioning process, which are not occurring with the Saastamoinen model.

GPSurvey does not determine a kinematic position for a given time epoch if the number of satellites is below 5, if the satellite geometry is evaluated to be poor, if there is a gap in phase data caused for instance by a cycle slip, or if the noise level is too high. When analyzing the kinematic position results, a large difference was found in the number of positions determined using the Saastamoinen and the HIRLAM approach. Generally, more positions were rejected by the software when using the Saastamoinen approach, and for instance for the morning data of September 5, only 84 % of the possible number of positions were determined with the Saatamoinen approach, as compared to 95% when using the HIRLAM approach. This could indicate that the residual noise level in the observations is smaller with the NWP approach, even though this is not seen in the resulting station coordinates.

The results shown in the tables above are all based on an identical number of time epochs. If a position was not determined for any instant in time by the Saastamoinen approach, this position was not used for determining the RMS of the HIRLAM approach either. No distinction was made, however, between kinematic positions based on float or fixed ambiguities.

4.5 HIRLAM Zenith Delays for Network RTK

In the real-time network RTK scenario, the GPS data was processed with the NetCor software made available by the University of Calgary. The NetAdjust method, which is described in Chapter 1, is implemented in the NetCor software. The software initially solves the ambiguities between the reference stations in the network, and then generates RTK corrections based on Equations (1.9) and (1.10). The software works sequentially, so that only data from the present and any previous data epochs are used for processing the current data epoch. It is thus possible to simulate real-time even if the data is logged in data files. The tests described in this section are also described in (Jensen et al., 2002).

A reduced data set was used for these tests. The ten GPS stations shown in Figure 4.17 were used, and a total of 6 hours of data from September 5 (00:00 - 03:20 UTC) and September 14 (01:00 - 03:45 UTC) was used.

The data was processed using both the Goad and Goodman global tropospheric delay model, described by Leick (1990) and implemented in NetCor, and the NWP approach where the tropospherically corrected data files were used.

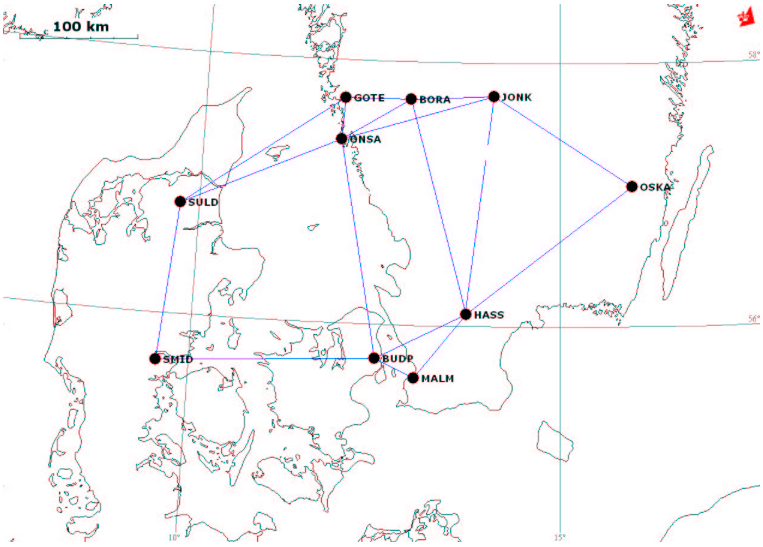

Figure 4.17: The 10 stations used, and the reference baselines processed, with the NetCor tests.

4.5.1 Ambiguity resolution

Before any RTK corrections can be determined with the NetAdjust method, the double difference ambiguities for the baselines between the reference stations in the network have to be resolved.

The ambiguity resolution in the NetCor software is based on the LAMBDA method (Teunissen and Tiberius, 1994) and the principles by Sun et al. (1999), who describe an ambiguity resolution method that is especially suitable for a network of reference stations with known coordinates. The method is based on reference station chaining, the fact that the sum of the double difference integer ambiguities in a closed loop is zero.

NetCor determines ambiguities for the minimum number of baselines necessary to generate RTK corrections for the area covered by the reference network. With the 10 stations shown in Figure 4.17 this gives a total of 17 necessary baselines.

In the tests carried out the ambiguities were solved faster, i.e. required less data, when using the NWP approach than when the Goad and Goodman model was used. In Table 4.6 the average time to solve the L1 ambiguities is given for the two data sets, using the two different tropospheric correction approaches.

	Number of $\nabla\triangle N_{L1}$	Fixed $\nabla\triangle N_{L1}$	Percentage fixed	Average time to fix
September 5, 2000				
Goad and Goodman	561	201	36 %	49 epochs
HIRLAM and NMF	561	188	34 %	43 epochs
September 14, 2000				
Goad and Goodman	490	207	42 %	75 epochs
HIRLAM and NMF	490	194	40 %	65 epochs

Table 4.6: Total number of L1 double difference ambiguities, Number of fixed L1 double difference ambiguities, percentage fixed, and average time to fix the L1 ambiguities with the two different approaches for tropospheric correction.

Generally, more ambiguities were solved with the Goad and Goodman model. The ambiguities which were solved with the Goad and Goodman model but not with the NWP approach, are generally the ones that required a longer time to resolution. This explains why the average time to fix is shorter with the NWP approach.

As will be indicated later in this section (Table 4.8), some of the ambiguities which were solved with the Goad and Goodman model but not with the NWP approach were solved to wrong integer numbers. So even though more ambiguities were solved with the Goad and Goodman approach, the result was not more useful.

43 epochs of 15 seconds correspond to 11 minutes, which is a fairly long time to resolution. But the lengths of the baselines in this network are in the high end of what the software has been developed to handle. Furthermore, the software only reads broadcast orbits, as it is meant for real-time positioning, so residual orbit errors of 1–2 cm are also present in the data processing.

4.5.2 Misclosures

The misclosures, which are the last part in the two Equations (1.9) and (1.10), are described in Chapter 1 and are given as:

$$B_n l_n - \lambda\nabla\triangle N_n \tag{4.2}$$

where B_n is the double difference matrix, l_n are the measurement minus range observations from the network, λ is the wavelength, and $\nabla\triangle N_n$ are the double difference ambiguities.

As previously mentioned, the misclosures are basically the difference between the double difference observations and the double differenced ambiguities, and they are thus a measure of the unmodelled errors in the observations and can be used to give an indication of the quality of the error modelling.

The NetCor software was used to generate the ionosphere–free (IF) misclosures where the double differenced IF combination of the L1 and L2 observations, and the double differenced IF ambiguities, are used for the l_n vector and for the $\nabla\triangle N_n$ matrix, respectively. Therefore the residual first order ionospheric effects are removed, so that the misclosures mainly indicate the size of the residual orbit and tropospheric errors. Since the ionospheric activity was low on the two days, any higher order ionospheric effects are expected to be negligible, and the influence of noise and multipath is also expected to be negligible.

When generating the IF misclosure for the two different tropospheric approaches, the residual orbit errors would be the same for the two processing scenarios, and any difference in IF misclosures shows the difference in residual tropospheric error.

The plots given in Figures 4.18 and 4.19 show the RMS of the IF misclosures for each baseline plotted as a function of baseline length for the two data sets. Misclosures were generated for all double difference combinations for every data epoch, but only when both the L1 and L2 ambiguities were fixed.

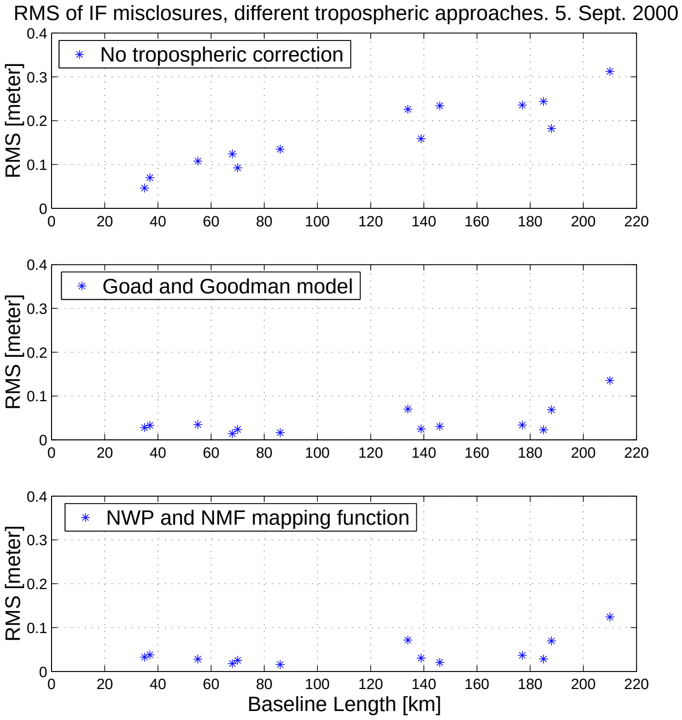

Figure 4.18: RMS of ionosphere–free misclosures obtained when not correcting for the troposphere, when using the Goad and Goodman model, and when using the NWP approach. September 5 data set.

The figures contain three plots. The first plot shows the RMS of the IF misclosure when no tropospheric correction was applied, the second plot shows the RMS of the misclosures when the Goad and Goodman model was used, and the last plot shows the results of using HIRLAM data and the NMF mapping function.

By analyzing Figure 4.18, a clear linear trend is seen in the first plot indicating that the residual tropospheric and orbit errors increase with increasing baseline length.

This trend is not so significant for the next two plots, when the majority of the tropospheric delay has been removed. For the baselines longer than 100 km, a slight linear trend is, however, indicated in the plot, and parts of this trend is caused by the residual orbit errors that normally have a size of about 0.1 ppm (Rothacher, 1997), i.e. 2 cm for a 200 km baseline, with broadcast orbit information. In addition, the residual tropospheric error will increase for the longer baselines when the double differences can no longer account for the inaccuracy in the tropospheric correction method used.

In Figure 4.19 the linear trend is also visible for the plots where the tropospheric corrections have been applied, and also the RMS values are larger for this data set. This indicates that both correction approaches leave larger unmodelled errors with higher tropospheric activity.

It should be noted that any ambiguities that are fixed to wrong integer numbers will also affect the misclosures. However, the exact same set of ambiguities were used for generating the misclosures in all three cases shown in the plots, and any wrong integers will only affect the size of the misclosures, not the relative difference between the misclosures generated using the different tropospheric correction approaches.

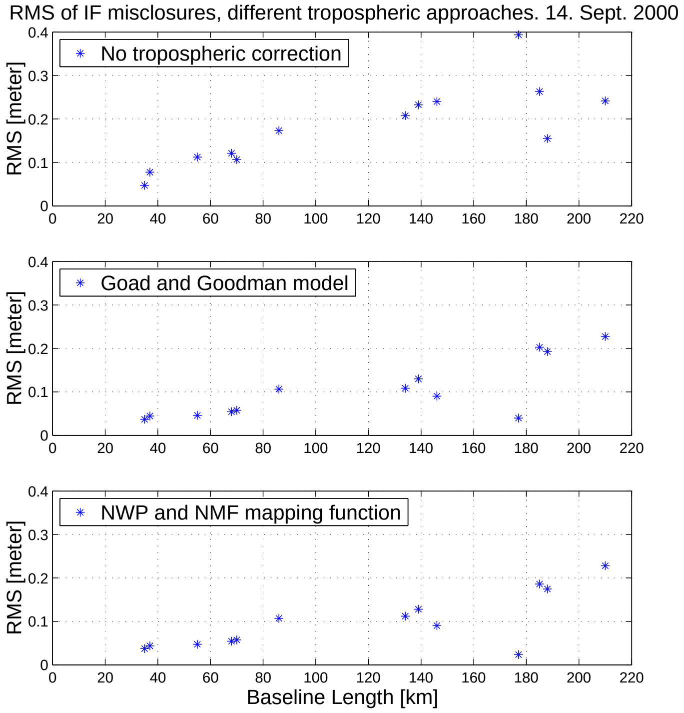

Figure 4.19: RMS of ionosphere–free misclosures obtained when not correcting for the troposphere, when using the Goad and Goodman model, and when using the NWP approach. September 14 data set.

Figures 4.18 and 4.19 show no difference between the Goad and Goodman model and the HIRLAM-approach, and therefore the RMS of the total number of misclosures for all baselines using the two different tropospheric approaches is given in Table 4.7. A minor improvement in the IF misclosures is obtained when using the NWP approach, but it is at the sub-cm level and is not significant.

In both Figures 4.18 and 4.19, and in Table 4.7, the misclosures for the 165 km

	Ionosphere–free misclosures		
	Mean [cm]	Std. Dev. [cm]	RMS [cm]
September 5, 2000			
Goad and Goodman	-0.1	4.9	4.9
HIRLAM and NMF	0.0	4.8	4.8
September 14, 2000			
Goad and Goodman	-0.8	9.3	9.3
HIRLAM and NMF	-0.3	9.2	9.2

Table 4.7: Mean, standard deviation and RMS of IF misclosures using the two tropospheric correction approaches and the two data sets.

baseline between GOTE and SULD are not included because of problems with the ambiguity resolution for this baseline on both days.

The misclosures are larger than what was expected. A total RMS value of about 1–3 cm, including residual orbit and tropospheric errors, was expected, and one of the explanations for these high RMS values can be that some ambiguities were fixed to wrong integer values. Consequently the NetCor ambiguities were compared with corresponding ambiguities determined with the Bernese software.

The raw GPS data was reprocessed with Bernese using a slightly modified version of the source code. The modified code, which was made available by L. P. Fortes, the University of Calgary, adds a phase offset, corresponding to the size of the ambiguities for the initial data epoch, to the rest of the ambiguities. This offset is normally subtracted for ambiguities determined with the Bernese software.

The single difference ambiguities from Bernese were then used to generate double difference ambiguities for the same double difference combinations as generated by the NetCor software, and the resulting Bernese L1 double difference ambiguities were compared to the L1 double difference ambiguities from the NetCor software.

The comparison was carried out for the ambiguities determined using both the Goad and Goodman and the HIRLAM approach, and the results are shown in Table 4.8.

	Fixed NetCor $\nabla\triangle N_{L1}$ identical with Bernese $\nabla\triangle N_{L1}$
September 5, 2000	
Goad and Goodman	54.9 %
HIRLAM and NMF	55.2 %
September 14, 2000	
Goad and Goodman	50.2 %
HIRLAM and NMF	51.1 %

Table 4.8: Percentage of fixed NetCor L1 ambiguities that are identical with the corresponding fixed Bernese L1 ambiguities.

The table shows that 50–55% of the ambiguities that were fixed, were fixed to the same integer values with the two software packages. No investigation has been made as to which ambiguities are correct, but it is anticipated that the results from Bernese are more accurate since the Bernese software is making use of the entire 24 hour data set, and, most importantly, the Bernese data processing is carried out using precise orbits which is not the case for the real-time NetCor processing.

However, some of the ambiguities might have been solved incorrectly with Bernese, and the numbers above can therefore not be used to conclude that NetCor only solved 55% of the ambiguities correctly. But any NetCor ambiguities fixed to wrong integer values would affect the misclosures discussed above, and wrong ambiguities are therefore most likely the explanation for the large RMS of the misclosures given in Table 4.7.

Table 4.8 also shows that slightly more ambiguities were fixed identically with the two software packages when using the HIRLAM approach than when using the Goad and Goodman model.

In order to reinvestigate the performance of the HIRLAM approach, a modified version of NetCor was made available by the University of Calgary. This version reads pre-determined ambiguities from input files, and it was thus possible to use Bernese ambiguities for the baselines between the reference stations in the network as the basis for the generation of misclosures and RTK corrections. This exercise has previously been carried out by the author with successful results using an older version of the NetAdjust software (Jensen, 1999). Unfortunately, it was not possible to generate any results with the new NetCor software before the deadline of this thesis.

4.5.3 RTK corrections

From the two NetAdjust equations (1.9) and (1.10) given in Chapter 1, it is seen that the misclosures are the dominant part of the RTK corrections generated with the NetAdjust method. The first part of the equations mainly consists of the variances and cross-covariances, and they are normally relatively small numbers compared to the misclosures.

When the difference between misclosures determined using the Goad and Goodman model and the HIRLAM approach is as small as in Table 4.7, there will not be any significant difference in the RTK corrections either. Therefore the final positions determined for the roving receiver will be almost identical with the two different tropospheric approaches, and no investigation was carried out in order to evaluate the influence of the HIRLAM approach in the RTK position domain.

In this case the improvement of using NWPs for network RTK is consequently only seen for the ambiguity resolution within the reference network.

5

HIRLAM SLANT DELAYS FOR GPS POSITIONING

For GPS signals received at lower elevation angles, the tropospheric delay can be estimated from an NWP in two different ways:

1. By determining the zenith delay, assuming that the satellite signal follows a straight line through the atmosphere, which is very close to being true in the zenith direction. Then apply a mapping function to determine the delay at lower elevation angles. This is the approach followed throughout Chapter 4.
2. By implementing a ray tracer that simulates the path of a satellite signal, at any elevation angle, through the NWP. This approach will be investigated in this chapter.

The first approach (zenith delay and mapping function) is computationally faster than using a ray tracer, but the results are also less reliable since mapping functions are based on the assumption of an isotropic atmosphere around the GPS antenna. With a ray tracer the spatial variability of the atmosphere is taken into account, and it will thereby give a better estimate of the tropospheric delay for lower elevation angles.

Selection of which procedure to use will thus be a trade off between computation time and accuracy. Using a combination of the two approaches is a reasonable solution, where the first option is used for satellite signals received at higher elevation angles, and a ray tracer is used for lower elevation angles where the mapping function is less reliable. This is the background for the tests described in the following, and these tests are also described in (Jensen, 2002).

5.1 Ray Tracing

A ray tracer models the signal path by splitting it into numerous small pieces, starting at the position of the satellite and working its way down through the neutral atmosphere. For each endpoint of a ray piece the refractivity is determined and taken into account when computing the next part of the ray path. The signal delay is finally determined as the difference between the total phase, in distance units, and the geometric distance between satellite and ground point.

For the tests to follow, a ray tracing code called ROSAP was made available by S. Syndergaard from the Institute of Atmospheric Physics at the University of Arizona. A description of the ray tracer is given by Høgh et al. (1995).

The ray tracer was originally developed to be used in connection with radio occultations, so several modifications were made in order to use in it here in connection with ground-based GPS stations and DMI-HIRLAM-E. Also, a geoid model (see Appendix C) was implemented since the ray tracer works with ellipsoidal heights, and the NWP is based on orthometric heights.

With these modifications of ROSAP it calls a HIRLAM extraction routine with the position (latitude, longitude and ellipsoidal height) of the given ray point. The ellipsoidal height is converted to an orthometric height using the geoid model and the interpolation routine which is also described in Appendix C. Then the latitude and longitude are rotated to match the coordinate system used by HIRLAM. The pressure, temperature and specific humidity for the given ray point is determined by horizontal and vertical interpolation as described in Appendix C, the refractivity for the point is finally determined using Equation (2.16), and the refractivity is returned to ROSAP.

Note that a different expression for the refractivity is used for the slant delays than for the zenith delays, because the error introduced by the ideal gas assumptions used in Chapter 3 would be considerable for low elevation angles, where a larger part of the signal path is going through the non-ideal moist air in the lower part of the atmosphere.

A global atmospheric model has been implemented above HIRLAM in order to account for the signal delay caused at altitudes above the NWP. The global atmospheric model used is a modified version of the MSIS90 model developed by Hedin (1991), and the code for the model was also made available by S. Syndergaard from the University of Arizona.

The uncertainty of the ray tracer is in (Høgh et al., 1995) evaluated to be within 1 mm, given a perfectly continuous neutral atmosphere. When integrating through an NWP the refractivity will not be perfectly continuous if there are local irregularities in the NWP (S. Syndergaard, personal communication). The accuracy of the delay determined by the ray tracer is thus more dependent on the accuracy and the continuity of the parameters in DMI-HIRLAM-E, than on the numerical resolution of the ray tracer itself.

The ray tracer was tested on zenith delays for the GPS stations ONSA and BORA, and the ray traced zenith delays corresponded to the zenith delays determined in Chapter 4 to within 4–7 mm. Mapping this zenith difference down to a 5° elevation angle, gives 4–8 cm, which is an indication of the type of accuracy that can be obtained with a ray tracer. A thorough verification of the accuracy of the slant delays determined with the ray tracer has, however, not been carried out.

5.2 *Ray Tracing Versus Mapping Functions*

As described in Chapter 2, the NMF mapping function (Niell, 1996) is recommended when no meteorological data is available. If meteorological observations are available, the Ifadis (Ifadis, 1986) and MTT (Herring, 1992) mapping functions are recommended.

In this section some tests are described that have been carried out with the purpose of estimating the "critical" elevation angle, i.e. the elevation angle where the mapping functions become too unreliable and the use of a ray tracer is a better alternative.

The three mapping functions mentioned above were implemented, and the hydrostatic and wet mapping factors were applied to the hydrostatic and wet zenith delays determined using DMI-HIRLAM-E and the programme ZTDPOINT, used in Chapter 4.

Figure 5.1 shows the signal delay determined using the three mapping functions for the BORA GPS station. Meteorological surface parameters for the Ifadis and

MTT mapping functions were determined by interpolation in DMI-HIRLAM-E.

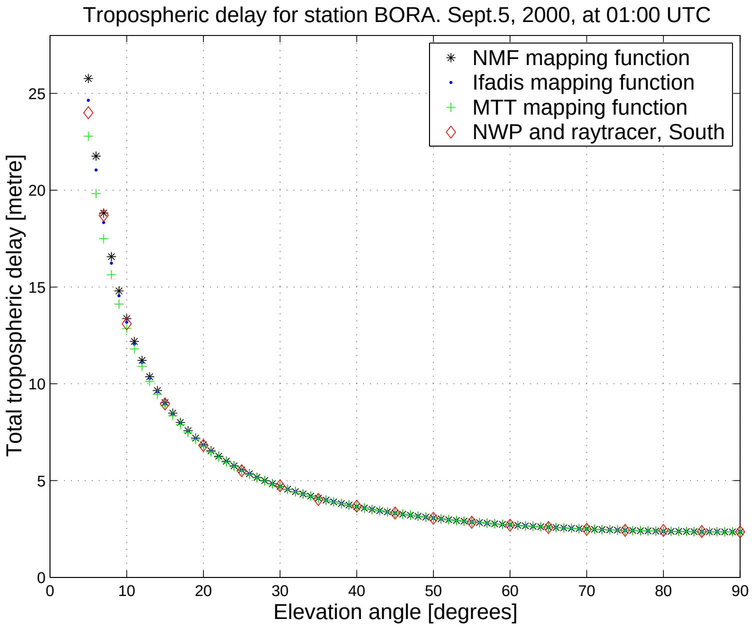

Figure 5.1: Signal delay for various mapping functions and ray tracer.

5.2.1 Comparison of slant delays

For carrier–phase based GPS positioning, an elevation mask of 10°–15° is often used in order to eliminate some of the noisy low elevation signals. A higher elevation angle is used for static applications and a lower elevation angle is used for kinematic applications where the increased amount of data is more important than the noise level, since the position accuracy under any circumstances will be lower than for static applications. The noise for the low elevation signals is mainly caused by multipath and residual tropospheric effects. If a better "model" of the tropospheric delay is available, it will be possible to include more of the satellite signals received at lower elevation angles in the positioning process. In the following, satellite signals received down to a 5° elevation angle are therefore considered.

In Figure 5.2 the tropospheric delay estimated by DMI-HIRLAM-E zenith delays and the three mapping functions are plotted again, this time only for elevation angles of 5°, 7°, 10°, 15° and 20°.

Also plotted in Figure 5.2 are the delays estimated with the ray tracer. The ray tracer was used for satellite signals received from East, South and West of station BORA at the different elevation angles, whereby it is possible to analyze the difference in delay caused by the asymmetric atmospheric properties around the antenna.

For this data set, the variation in signal delay with varying azimuth is only a few mm for 15°, but up to 1.2 meters for the 5° elevation angle, which is, however, within the difference between the mapping functions. The delay determined East of the station at 20° is marked by the ray tracer as being inaccurate, and this is most likely caused by discontinuities in HIRLAM.

The same procedure was carried out using HIRLAM data, ray tracing and mapping functions for station ONSA. This station is located at a lower altitude than BORA,

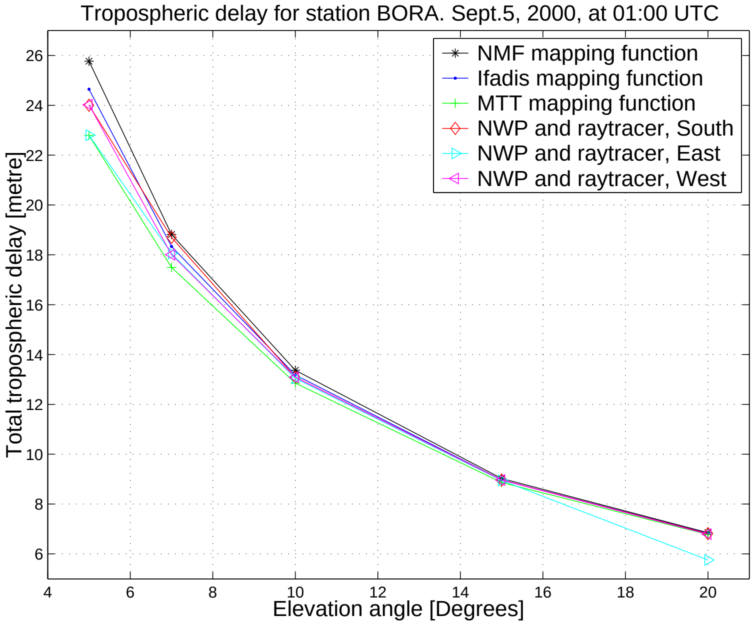

Figure 5.2: Tropospheric delays determined with mapping functions and ray tracer at three different azimuths for station BORA.

so the signal delays are larger. Furthermore station ONSA is located right on the coast.

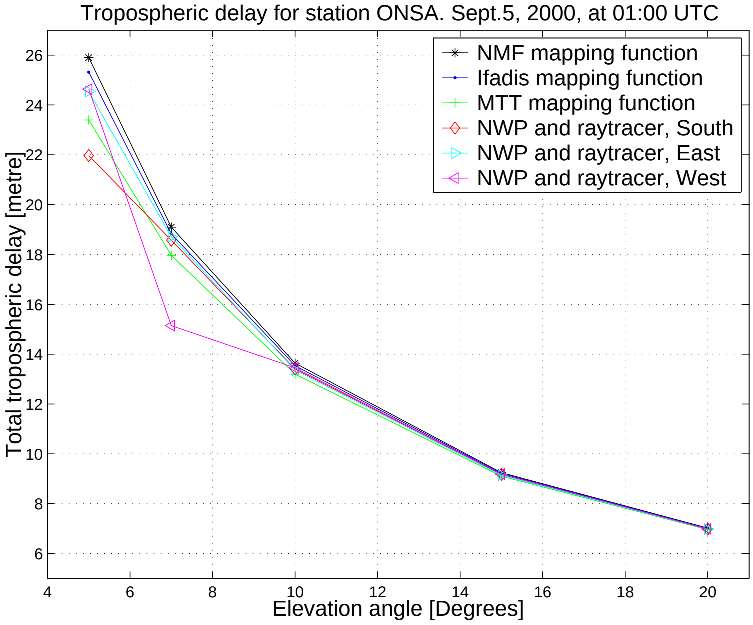

Figure 5.3: Tropospheric delays determined with mapping functions and ray tracer at three different azimuths for station ONSA.

Figure 5.3 shows the slant delays for station ONSA. In this plot there is also an outlier, and again it is being flagged by the ray tracer, most likely because of irregularities in DMI-HIRLAM-E. In this situation it is a signal from the West that is causing problems. The ONSA station is located on the Swedish coast of the Kattegat

Sea (see Figure 4.1), and for signals acquired from the West the last part of the signal path is over water. There are no meteorological observations from the Kattegat Sea in HIRLAM, and it is anticipated that the accuracy of HIRLAM is slightly degraded over the sea because of the lack of surface observations.

Table 5.1 shows the RMS of the differences between signal delays determined using the ray tracer, in the East, South and West directions, and the mapping functions for the two stations. In order to directly compare the results for the different elevation angles, the RMS values of the differences have been mapped to the zenith by multiplying with $sin(elv)$. The two outliers are not included in the RMS values in the table.

	RMS of difference mapped to zenith		
Elevation	NMF	Ifadis	MTT
20°	0.014 m	0.007 m	0.008 m
15°	0.019 m	0.009 m	0.021 m
10°	0.047 m	0.022 m	0.037 m
7°	0.071 m	0.034 m	0.093 m
5°	0.207 m	0.144 m	0.100 m

Table 5.1: RMS of difference in signal delays (mapping function minus ray tracer) mapped to the zenith.

All three mapping functions are defined to be used down to elevation angles below 5°, but at these low elevation angles the variability in the water vapor content limits the performance of the mapping functions, and there is a significant variation in the results obtained with the various mapping functions for elevation angles below 10°.

For the results given at 7°, the RMS for both the NMF and the MTT mapping functions are close to one decimeter. Errors of this size, corresponding to about half a wavelength for the L1 and L2 frequencies, are considerable even though most of the error will be eliminated when the tropospheric modelling is followed by double differencing. These tests indicate that the use of a ray tracer, compared to using a mapping function, might be beneficial for elevation angles of 10° and lower. Further tests will be necessary, however, in order to evaluate the influence that the ray tracer has on GPS positions.

The ray tracer is computationally demanding, and for implementational purposes it might be more beneficial to assume that the satellite signal follows a straight line path through the NWP. The slant signal delay can be determined by integrating the refractivity along the geometric signal path, and even though the ray bending is ignored by this approach, the results might be better than using the mapping functions. This is because the variability in the atmospheric conditions, given in the NWP around the GPS antenna, would be represented by the varying refractivity along the signal path. This has, however, not been tested.

6

PERSPECTIVES

With this chapter the work described in the previous three chapters is put into a larger perspective. First by discussing some implementational issues, and then by considering the relevance of this work in relation to the planned improvement of the GPS system and the future European Galileo system.

6.1 Implementational Considerations

The tests carried out in Chapter 4 are not clearly showing that numerical weather predictions can improve GPS positioning. Some of the results are positive, but more research is necessary in order to conclude whether or not there are operational benefits with the method. Some implementational considerations are given below, since the test scenario used for this thesis has proven not to be the ideal approach to deal with NWPs.

6.1.1 Post–processing software

For the tests carried out using the GPSurvey software the data was corrected externally for the tropospheric delay. It is anticipated, however, that it would be fairly simple to implement the method in a post–processing software for either static or kinematic positioning. The zenith delays derived from an NWP can be generated separately for instance on a hourly basis, and be made available by a meteorological organization. These zenith delays can then be introduced into the post processing process where the software normally uses a global tropospheric delay model. This approach is similar to the procedure for using precise orbits, which are introduced as separate data files read by the program.

At the point in the positioning process where the tropospheric correction is applied the zenith delays can be read from the external files. Then slant delays can be derived using a mapping function and the elevation angle of the signal, which is already determined by the processing software. This would only require the addition of a small subroutine to the software package.

The processing software could also be set up to determine the zenith delays if the entire NWPs are made available. This would, however, require more comprehensive changes to the software, and also increase the processing time and the disk space requirements, since weather fields are voluminous.

A more advanced solution is to implement a ray tracer in the software and let this function determine the signal delay for each satellite and epoch in time. The results in Chapter 5 indicate that this will only be necessary for signals received below a 10° elevation angle. Implementing a ray tracer will increase the processing time considerably and this might not be beneficial from a user perspective.

Conclusively, the procedure with zenith delays given in data files seems like the most straightforward solution, and it would be eased if a standard data format was developed for the purpose.

6.1.2 Network RTK

Implementing NWPs in connection with network RTK would also be fairly straightforward, at least with respect to processing of data from the reference stations.

As also described above zenith delays can be made available as 1-hour predictions by a meteorological organization. These can be incorporated into the ambiguity resolution process for the reference network. The reference stations would all be connected to a processing center, and this center could automatically acquire both zenith delays and predicted orbits via the internet, and then use these files in the ambiguity resolution process.

Introducing NWPs for the actual correction generation process might be more complicated, and it is probably more beneficial to estimate a tropospheric correction based on the data from the reference stations as suggested by Zhang and Lachapelle (2001).

However, the NWPs can be used as a basis for determining coefficients for a local or regional continuos tropospheric model. This model could be developed based on climate data, and depending on the construction of the model, any coefficients can be estimated on an hourly basis in the processing center, based on NWPs provided by a meteorological organization. Such a model could include modelling of the horizontal gradients for a given rover site, where the gradients are estimated based on the NWP. It could also include a measure of the temporal variation of the zenith delay based on knowledge about movements of, for instance, low pressure systems as given by the NWP.

The model could then be used to determine tropospheric delays for the location of the roving GPS receiver, and the delay could be transmitted to the rover along with the other RTK corrections, assuming that some suitable data transmission format is available for the purpose.

Distribution of the corrections to the roving receiver is not discussed in this thesis. The subject is being treated by the RTCM subcommittee working with developing the new RTCM format as mentioned in Chapter 1. See for instance Euler et al. (2001). The subject has also be treated by for instance Fotopoulos (2000) and Raquet and Lachapelle (2001).

6.2 Modernized GPS

With the purpose of making GPS positioning and timing more accurate and robust, and also in order to make the system more resistant to radio interference and jamming, a GPS modernization plan was launched by the US White House in 1998.

The first step in the modernization process was taken in May 2000 when Selective Availability, degradation of information in the navigation message, was deactivated.

The next step will be the implementation of a civil C/A code on the L2 frequency, whereby the L2 signal can be used for civil code–based positioning. Also, the new code on L2 will improve the use of carrier–phase observations from L2 since the noisy codeless or semi-codeless techniques can be replaced by a direct code–correlation tracking technique as is being used for tracking the L1 carrier today.

More reliable phase observations on L2 will improve handling of the ionosphere within the positioning process, since the ionosphere–free linear combination of the observables can be used with more confidence. As discussed in Chapter 2, the effects of scintillation will also be minimized. The new civil code on L2 will, however, be most beneficial for code based positioning since it will be possible to perform civil dual frequency code–based positioning.

The next step in the GPS modernization plan is a new civil frequency which is now referred to as L5, and this is where the major improvements for carrier–phase based positioning are expected. With an extra frequency a new range of linear combinations of the phase observations can be generated. One example is the L2–L5 wide lane observable that, with the planned center frequency for L5 of 1176.45 MHz, will give a wavelength of 5.8 meters. It will be fairly easy to solve the wide lane ambiguities with such a long wavelength. Afterwards this can be used for easier resolution of the traditional L1–L2 wide lane ambiguities, and thereby finally the process of solving the L1 ambiguities is improved. The procedure for using L5 for L1 ambiguity resolution is described by Misra and Enge (2001).

Also, the addition of the L5 frequency opens up new ionosphere–free linear combinations that are expected to improve GPS positioning accuracy (Odijk et al., 2002).

The last part of the modernization plan affecting civil GPS positioning is an update of the procedures in the GPS control segment. This will result in better quality of the broadcast orbits and clock corrections, and this upgrade runs parallel with the other modernization efforts.

The present time schedule for the civil parts of the GPS modernization plan, as given by Misra and Enge (2001) and Shaw et al. (2000), is the following:

- 2003: the first satellites transmitting C/A code on L2 will be launched.
- 2005: the first satellites transmitting on the L5 frequency will be launched.
- 2010: C/A code transmitted from all satellites on both L1 and L2.
- 2015: Full modernized constellation with the L5 signal transmitted from all satellites.

When complete, the system will not only provide a better accuracy for civil GPS users, but the new signal will also be transmitted with a higher signal power, reducing the risk of interference and jamming. Furthermore, if one of the three frequencies is lost by interference or jamming, the other two signals will still be available and can be used for positioning.

The US government has also started working on the GPS III project, which deals with a redesign of the GPS system architecture including for instance new satellite designs, a new satellite constellation and various augmentation opportunities. The purpose of GPS III is to ensure that GPS meets the user needs and is competitive within the next 30 years, and whatever is developed is expected to be implemented beginning in 2010 (Shaw et al., 2000).

The advantages of a modernized GPS for high end users are thus primarily given by more combination possibilities of the phase observables which will improve handling of the ionosphere and ambiguity resolution. The troposphere is however still an error source to consider in the positioning process, and even though the influence of the tropospheric error will be less important with the new linear combinations, the troposphere still has to be treated separately by some kind of modelling. Here NWPs, as well as continuous regional tropospheric delay models determined from climate data, can be an option.

6.3 Galileo

On March 26, 2002 the European Council of Transport Ministers agreed on funding the development phase of the Galileo system. Galileo will be a global satellite system for positioning and navigation, developed and maintained by the European Union and the European Space Agency.

Unlike GPS, Galileo will be a system under civil control, and the two systems further differ by the introduction of user fees in connection with Galileo. Some Galileo services will be free of charge, but other services, where a guarantee or a certain quality is expected, will have to be paid for (EC, 2002a). It is also expected that the private sector will play a role in financing Galileo at the later stages in the project, and public participation is therefore expected to decrease gradually until 2015.

The present time schedule for Galileo as given by (EC, 2002a) is as follows:

- 2002–2005: Development and validation phase.
- 2006–2007: Deployment phase[1].
- 2008–: Commercial operation phase.

Funding has only been provided for the development and validation phase, so there is no guarantee for an operational system in 2008.

The system design has not yet been finalized, but some initial investigations point to a system with 30 satellites placed in orbits at an altitude of about 24000 km above the surface of the Earth (EC, 2002b). Also, it looks like the orbits will have a higher inclination angle than the GPS orbits, 57° or 60°, to ensure a reasonable coverage at higher latitudes, for instance in Sweden and Finland.

Four frequency bands have been allocated for Galileo in the radio frequency spectrum and some considerations as how to use these frequencies are given by Benedicto and Ludwig (2001). Navigation codes will be modulated onto the carriers, and some of them might be encrypted for the services based on user fees. The signal structure and transmission data rate will have the possibility of transmitting other information as, for instance, digital maps or RTK corrections.

Presently there is much focus from both the European and American side on the possible risk of interference between GPS and Galileo signals. On the positive side there is also a lot of work going on already with the purpose of ensuring interoperability of GPS and Galileo, and this is where the real advantages of Galileo will be seen for the high end users.

With a well functioning and non-interfering interoperability of the two systems, the user will be able to obtain signals from twice as many satellites as with GPS or Galileo alone. This will greatly improve the positioning performance mainly in

---

[1] According to EC (2002b) the first Galileo satellite must be launched before February 13, 2006, otherwise the right to use the frequencies, obtained for Galileo at the World Radio Communications Conference in May 2000, will be lost.

terms of reliability and accuracy. With more satellite signals available, the influence of multipath and cycle slips will be reduced. Furthermore, with satellite signals transmitted on more frequencies, more ionosphere–free and wide lane combinations of the phase observables can be generated, and this will most likely improve high accuracy positioning in both real–time and post–mission.

The proposed plans for Galileo contain no specific studies on new ways of handling the tropospheric effect on the satellites signals. So, as mentioned above in connection with modernized GPS, there will also be a need for handling the tropospheric delay in the positioning process with Galileo. Again, regional tropospheric delay models might offer improved performance as compared to global models, if developed for instance for Europe.

But NWPs might also be a solution for Galileo, and since the system is under development a lot of possibilities are still open. One solution could be to distribute coefficients for a tropospheric delay model along with the broadcast information on satellite orbits and clocks. An ionospheric model is distributed with the GPS navigation message today, so broadcasting a tropospheric model from the Galileo satellites is not unreasonable. The tropospheric coefficients for such a model could be estimated based on NWPs, and could thus be available every 6 or 12 hours just as the orbit information today is uploaded to the GPS satellites every 12 hours.

Considering the combined GPS and Galileo case, the tropospheric error will be less interesting. With more satellites in view, and thereby more signals available, the improvements added by introducing better tropospheric modelling might be less important. For instance, it will be possible to work consistently with a higher elevation mask, since enough signals will be provided by satellites above say 20°. Hereby the low elevation signals where the tropospheric effect is largest, and where an improved tropospheric estimation process would have the largest influence, may be eliminated.

7

CONCLUSIONS

The objective of this thesis work was to develop a method for dealing with the tropospheric delay in connection with network RTK, giving a better positioning performance than the traditional global tropospheric delay models.

The new method is based on the use of numerical weather predictions (NWPs) for determining tropospheric zenith delays. These delays can then be introduced into the positioning process, and several tests have been carried out in order to evaluate the new NWP method.

Summary of test results

Initially the accuracy of the NWP zenith delays was estimated by a comparison with zenith delays determined from GPS data. The verification showed that the zenith delays were determined with an RMS of about 1.5 cm, which was according to expectations and better than the estimated accuracy of the Saastamoinen model, which is known as the best global tropospheric delay model.

The tropospheric zenith delays were then used in several test scenarios in order to investigate whether the new method would imply an improvement for the GPS positioning process.

With the static tests an improvement in the position domain was obtained for 25 of the 51 baselines processed, and for the kinematic tests an improvement in position was obtained for 23 of the 51 processing sessions. Thus an improvement was obtained for about 50% of the positions determined when using the NWP approach, as compared to the Saastamoinen model. For the other half of the results, no improvement was obtained.

With respect to network RTK, an improvement was seen in the number of observations necessary to solve the ambiguities within the network of reference stations. The number of fixed ambiguities was reduced with the NWP approach compared to using the Goad and Goodman model, but the following ambiguity check indicated that more ambiguities were solved to the correct integer values when using the NWP approach.

The misclosure analysis showed a slightly smaller residual noise level with the NWP approach. The improvement was small, however, and seems to be so small that it disappears within the other residual errors present in the data processing.

Conclusions

The new NWP approach for determining tropospheric zenith delays shows a considerable improvement in the estimated accuracy of the zenith delays as compared to zenith delays determined with the Saastamoinen model. High resolution NWPs, as the DMI-HIRLAM-E used for test purposes in this thesis, are therefore a new and very promising tool for the estimation of the tropospheric error in GPS positioning.

NWPs have been tested for kinematic positioning and for network RTK; two types of tests that have never previously been carried out. The test results show that the method does have a potential especially with respect to the ambiguity resolution.

The results may, however, be influenced by the fact that the NWP method is not applied in an optimal way, and more work is necessary, to conclude whether the method can provide an improvement in the position domain. Several suggestions to future work are given below.

7.1 Future work

Most importantly, the method must be implemented directly into the processing software, so that the tropospheric correction can be carried out as an integral part of the positioning process, and not as an external process which was the case with the tests carried out in this thesis.

Only a sparse amount of data has been used for these tests, and in order to draw any final conclusions as to whether or not the method is feasible, tests must be carried out using data from different seasons of the year and from different geographic locations.

The success rate of the method is directly related to the accuracy of the numerical weather model, and the method should therefore also be tested using NWPs with a higher resolution, as for instance DMI-HIRLAM-D, which has a resolution of 0.05° x 0.05°.

Also, it would be interesting to test the method using data from days with more atmospheric activity. With a higher activity level in the troposphere, it is expected that the method would show more significant improvement, since the global delay models do not model rapid temporal variations in the tropospheric delay very well.

It would also be interesting to test the method with GPS data collected during high ionospheric activity. If the data is processed in a scenario where the higher order ionospheric effects and the residual tropospheric errors are treated separately, an improvement by introducing NWPs will most likely be found in the positioning process.

Possibilities with the ray tracer have not been fully investigated in this thesis. NWPs combined with a ray tracer can for instance be used to develop local or regional mapping functions. These mapping functions could be parameterized with respect to the predicted atmospheric conditions around the GPS antenna, e.g. the water vapor content, whereby the isotropic assumption, normally used for mapping functions, can be eliminated.

This can also be used in connection with network RTK, where the NWPs can be used for developing tropospheric RTK corrections. Today this is carried out using GPS data from reference stations, but if the GPS data is affected for instance by multipath, these errors will propagate into the RTK corrections. Therefore a better estimate for the tropospheric error over the area is expected when using an external data set as NWPs.

As mentioned in the perspectives in Chapter 6, the troposphere will remain one of the largest error sources after GPS has been modernized and when Galileo becomes operational. So there is still a need for more research focusing on improving the methods for estimating the tropospheric delay for high accuracy satellite based navigation and positioning.

BIBLIOGRAPHY

Allnutt, J. (1989). *Satellite-to-Ground Radiowave Propagation.* IEEE Electromagnetic Waves Series 29. Peter Peregrinus.

Behrend, D., Cucurull, L., Cardellach, E., Rius, A., Sedo, M. J., and Nothnagel, A. (2001). The Use of NWP Products in Near Real-Time GPS Data Processing. In *Proceedings of the 14th International Technical Meeting of the Satellite Division of the Institute of Navigation (ION GPS-2001)*, pages 2499–2506.

Benedicto, J. and Ludwig, D. (2001). Galileo Defined, Proposed Architecture and Services for the New European Satellite Positioning System. *GPS World*, 12(9):46–49.

Beutler, G., Brockmann, E., Dach, R., Fridez, P., Gurtner, W., Hugentobler, U., Johnson, J., Mervart, L., Rothacher, M., Schaer, S., Springer, T., and Weber, R. (2000). *Bernese GPS Software.* Astronomical Institute, University of Berne.

Bevis, M., Businger, S., Chriswell, S., Herring, T. A., Anthes, R. A., Rocken, C., and Ware, R. H. (1994). GPS Meteorology: Mapping Zenith Wet Delays onto Precipitable Water. *Journal of Applied Meteorology*, 33:379–386.

Bevis, M., Businger, S., Herring, T. A., Rocken, C., Anthes, R. A., and Ware, R. H. (1992). GPS Meteorology: Remote Sensing of Atmospheric Water Vapor Using the Global Positioning System. *Journal of Geophysical Research*, 97(D14):15787–15801.

Blewitt, G. (1989). Carrier Phase Ambiguity Resolution for the Global Positioning System Applied to Geodetic Baselines up to 2000 km. *Journal of Geophysical Research*, 94(B8):10187–10203.

Blewitt, G. (1997). Basics of the GPS technique: Observation equations. In *Geodetic Applications of GPS, Lecture Notes for Nordic Autumn School*, number 16 in Reports in Geodesy and Geographical Information Systems, pages 9–54. National Land Survey of Sweden. Edited by Bo Jonsson.

Brunner, F. K. and Gu, M. (1991). An improved model for the dual frequency ionospheric correction of GPS observations. *Manuscripta Geodaetica*, 16:205–214.

Brunner, F. K. and Welsch, W. M. (1993). Effects of the Troposphere on GPS Measurements. *GPS World*, 4(1):42–51.

Budden, K. G. (1985). *The Propagation of radio waves. The teory of radio waves of low power in the ionosphere and the magnetosphere.* Cambridge University Press.

Campbell, W. H. (1997). *Introduction to Geomagnetic Fields.* Cambridge University Press.

Cannon, M. E. (1997). Carrier Phase Kinematic Positioning: Fundamentals and Applications. In *Geodetic Applications of GPS, Lecture Notes for Nordic Autumn School*, number 16 in Reports in Geodesy and Geographical Information Systems, pages 157–179. National Land Survey of Sweden. Edited by Bo Jonsson.

Davis, J. L., Herring, T. A., Shapiro, I. I., Rogers, A. E. E., and Elgered, G. (1985). Geodesy by radio interferometry: Effects of atmospheric modeling errors on estimates of baseline length. *Radio Science*, 20(6):1593–1607.

DoD (2001). Global Positioning System, Standard Positioning Service Performance Standard. Technical report, Assistant Secretary of Defense, US Department of Defense, Pentagon, Washington.

Dodson, A. H., Chen, W., Baker, H. C., Penna, N. T., Roberts, G. W., Jeans, R. J., and Westbrook, J. (1999). Assesment of EGNOS Tropospheric Correction Model. In *Proceedings of the 12th International Technical Meeting of the Satellite Division of the Institute of Navigation (ION GPS-99)*, pages 1401–1407.

Dodson, A. H., Shardlow, P. J., Hubbard, L. C. M., Elgered, G., and Jarlemark, P. O. J. (1996). Wet tropospheric effects on precise relative GPS height determination. *Journal of Geodesy*, 70:188–202.

Duan, J., Bevis, M., Fang, P., Bock, Y., Chriswell, S., Businger, S., Rocken, C., Solheim, F., Hove, T, v., Ware, R., McClusky, S., Herring, T. A., and King, R. W. (1996). GPS Meteorology: Direct Estimation of the Absolute Value of Precipitable Water. *Journal of Applied Meteorology*, 35:830–838.

EC (2002a). Galileo, the countdown has started. European Commission, Directorate-General for Energy and Transport. April 2002.

EC (2002b). Galileo, the European projet on radio navigation by satellite. Information note. European Commission, Directorate-General for Energy and Transport. March 2002.

Elgered, G. (2001). An Overview of COST Action 716: Explotation of Ground-Based GPS for Climate and Numerical Weather Predition Applications. *Physics and Chemistry of the Earth*, 26A(6-8):399–404.

Emardson, T. R., Elgered, G., and Johansson, J. (1998). Three months of continuous monitoring of atmospheric water vapor with a network of Global Positioning System receivers. *Journal of Geophysical Research*, 103(D2):1807–1820.

Emardson, T. R. and Jarlemark, P. O. J. (1999). Atmospheric modelling in GPS analysis and its effect on the estimated geodetic parameters. *Journal of Geodesy*, 73:322–331.

Emardson, T. R. and Johansson, J. (1998). Spatial interpolation of the atmospheric water vapor content between sites in a ground-based GPS network. *Geophysical Research Letters*, 25(17):3347–3350.

Euler, H. J., Keenan, C. R., Zebhauser, B. E., and Wübbena, G. (2001). Study of a Simplified Approach in Utilizing Information from Permanent Reference Stations Arrays. In *Proceedings of the 14th International Technical Meeting of the Satellite Division of the Institute of Navigation (ION GPS-2001)*, pages 379–391.

Falthammar, C. G. (1994). *Space Physics.* Department of Plasma Physics, Alfvèn Laboratory, Royal Institute of Technology, Stockholm.

Fedrizzi, M., Eurico, R. d. P., Kantor, I. J., Langley, R. B., Santos, M. C., and Komjathy, A. (2002). Mapping the low-latitude Ionosphere with GPS. *GPS World*, 13(2):41–47.

Fortes, L. P., Cannon, M. E., and Lachapelle, G. (2000). Testing a Multi-Reference GPS Station Network for OTF Positioning in Brazil. In *Proceedings of the 13h International Technical Meeting of the Satellite Division of the Institute of Navigation (ION GPS-2001).*

Fortes, L. P., Lachapelle, G., Cannon, M. E., Marceau, G., Ryan, S., Wee, S., and Raquet, J. (1999). Testing of a Multi-Reference GPS Station Network for Precise 3D Positioning in the St. Lawrence Seaway. In *Proceedings of the 12th International Technical Meeting of the Satellite Division of the Institute of Navigation (ION GPS-99). Nashville, Tennessee*, pages 1259–1269.

Foster, J. C. (2000). Quantitative Investigation of Ionospheric Density Gradients at Mid Latitiudes. In *Proceedings of the 2000 National Technical Meeting of the Satellite Division of the Institute of Navigation (ION NTM 2000)*, pages 447–453.

Fotopoulos, G. (2000). Parameterization of Carrier Phase Corrections Based on a Regional Network of Reference Stations. In *Proceedings of the 13th International Technical Meeting of the Satellite Division of the Institute of Navigation (ION GPS-2000).*

Fotopoulos, G. and Cannon, M. E. (2001). An Overview of Multi-Reference Station Methods for cm-level Positioning. *GPS Solutions*, 4(3):1–10.

Frei, E. and Beutler, G. (1990). Rapid Static Positioning Based on the Fast Ambiguity Resolution Approach "FARA": Theory and First results. *Manuscripta Geodaetica*, 15:325–356.

Goad, C. C. (1996). Short distance GPS models. In *GPS for Geodesy*, Lecture Notes in Earth Sciences, pages 239–262. Springer-Verlag. Edited by Kleusberg, A. and Teunissen P. J. G.

Gutman, S. I. and Benjamin, S. G. (2001). The Role of Ground-Based GPS Meteorological Observations in Numerical Weather Prediction. *GPS Solutions*, 4(4):16–24.

Haase, J., Calais, E., Talaya, J., Rius, A., Vespe, F., Santangelo, R., Huang, X. Y., Davila, J. M., Ge, M., Cucurull, L., Flores, A., Sciarretta, C., Pacione, R., Boccolari, M., Pugnaghi, S., Vedel, H., Mogensen, K., Yang, X., and Garate, J. (2001a). The Contributions of the MAGIC Project to the COST 716 Objectives of assessing the Operational Potential of Ground-Based GPS Meteorology on an International Scale. *Physics and Chemistry of the Earth*, 26A(6-8):433–437.

Haase, J., Vedel, H., Ge, M., and Calais, E. (2001b). GPS Zenith Tropospheric Delay (ZTD) Variability in the Mediterranean. *Physics and Chemistry of the Earth*, 26A(6-8):439–443.

Han, S. and Rizos, C. (1996). GPS Network Design and Error Mitigation for Real-Time Continuous Array Monitoring Systems. In *Proceedings of the 9th International Technical Meeting of the Satellite Division of the Institute of Navigation (ION GPS-96). Kansas City, Missouri*, pages 1827–1836.

Han, S. and Rizos, C. (1997). Comparing GPS Ambiguity Resolution Techniques. *GPS World*, 8(10):54–56.

Hargreaves, J. K. (1992). *The Solar-Terrestrial Environment*. Cambridge University Press.

Hatch, R. and Euler, H. J. (1994). Comparison of Several AROF Kinematic Techniques. In *Proceedings of the 7th International Technical Meeting of the Satellite Division of the Institute of Navigation (ION GPS-94)*.

Hedin, A. E. (1991). Extension of the MSIS Thermosphere Model into the Middle and Lower Atmosphere. *Journal of Geophysical Research*, 96(A2):1159–1172.

Herring, T. A. (1992). Modelling atmospheric delays in the analysis of space geodetic data. In *Proceedings of Symposium on Refraction of Transatmospheric Signals in Geodey*, number 36 in Publications on Geodesy, New Series. Netherlands Geodetic Commission, Delft. Edited by J. C. de Munck and T. A. Th. Spoelstra.

Hofmann-Wellenhof, B., Lichtenegger, H., and Collins, J. (1992). *GPS, Theory and Practice*. Springer-Verlag, 2nd edition.

Høgh, P., Hauchecorne, A., Kirchengast, G., Syndergaard, S., Belloul, B., Leitinger, R., and Rothleitner, W. (1995). Derivation of Atmospheric Properties Using Radio Occultation Technique. Technical Report 95-4, Solar-Terrestrial Physics Division, The Danish Meteorological Institute. Copenhagen.

Holton, J. R. (1992). *An Introduction to Dynamic Meteorology*. International Geophysics Series. Academic Press, 3rd edition.

Hopfield, H. S. (1969). Two-quartic Tropospheric Refractivity Profile for Correcting Satellite Data. *Journal of Geophysical Research*, 74(18):4487–4499.

Hopfield, H. S. (1971). Tropospheric effect on electromagnetically measured range: Prediction from surface weather data. *Radio Science*, 6(3):357–367.

Ifadis, I. (1986). The atmospheric delay of radio waves: Modelling the elevation dependence on a global scale. Technical Report 38L, School of Electrical and Computer Engineering, Chalmers University of Technology, Göteborg.

Jagieniak, S., Boelow, T., Haverland, M., Cannon, M., E., and Fotopoulos, G. (2000). An Integrity Concept for GPS Phase Ambiguity Software as Position Reference for Flight Inspection. In *Proceedings of the International Symposium on Precision Approach and Automatic Landing (ISPA 2000)*.

Janes, H. W. (1991). An Error Budget for GPS Relative Positioning. *Surveying and Land Information Systems*, 51(3):133 – 137.

Jarlemark, P. O. J. (1997). Analysis of Temporal and Spatial Variations in Atmospheric Water Vapor Using Microwave Radiometry. Technical Report 308, School of Electrical and Computer Engineering, Chalmers University of Technology, Göteborg, Sweden.

Jensen, A. B. O. (1999). Performance of Network RTK using Fixed and Float Ambiguities. Project Report for the course ENGO 699.59. Department of Geomatics Engineering, The University of Calgary. Unpublished.

Jensen, A. B. O. (2002). Investigations on the Use of Numerical Weather Predictions, Ray Tracing, and Tropospheric Mapping Funcions for Network RTK. In *Proceedings of the 15th International Technical Meeting of the Satellite Division of the Institute of Navigation (ION GPS-2002)*. In press.

Jensen, A. B. O. and Cannon, M. E. (2000). Performance of Network RTK Using Fixed and Float Ambiguities. In *Proceedings of the 2000 National Technical Meeting of the Satellite Division of the Institute of Navigation (ION NTM 2000)*, pages 797–805.

Jensen, A. B. O., Townsend, B. R., and Cannon, M. E. (2002). The use of Numerical Weather Predictions for Network RTK. In *Proceedings of the European Navigation Conference - GNSS2002, Copenhagen*. Nordic Institute of Navigation. In press.

Johansson, J. M. (1997). Modeling of the earth atmosphere in space geodetic applications. In *Geodetic Applications of GPS, Lecture Notes for Nordic Autumn School*, number 16 in Reports in Geodesy and Geographical Information Systems, pages 109–134. National Land Survey of Sweden. Edited by Bo Jonsson.

Johansson, J. M. and Scherneck, H. G. (1997). Three years of continuous observations in the SWEPOS network. In *Geodetic Applications of GPS, Lecture Notes for Nordic Autumn School*, number 16 in Reports in Geodesy and Geographical Information Systems, pages 181–202. National Land Survey of Sweden. Edited by Bo Jonsson.

Joosten, P. and Tiberius, C. (2000). Fixing the Ambiguities - Are You Sure They're Right? *GPS World*, 11(5):46–51.

Klobuchar, J. A. (1996). Ionospheric Effects on GPS. In *Global Positioning System: Theory and Applications, Vol. I*, Progress in Astronautics and Aeronautics, Vol. 163., pages 485–515. American Institute of Astronautics and Aeronautics. Edited by B.W. Parkinson, J. J. Spilker, P. Axelrad, and P. Enge.

Klobuchar, J. A., Dohety, P. H., and Bakry El-Arini, M. (1995). Potential Ionospheric Limitations to GPS Wide-Area Augmentation System (WASS). *Navigation: Journal of The Institute of Navigation*, 42(2):353–370.

Kunches, J. M. and Klobuchar, J. A. (2000). Eye on the Ionosphere, Some Aspects of the Variability of Geomagnetic Storms. *GPS Solutions*, 4(1):77–78.

Lachapelle, G. (1993). NAVSTAR GPS Theory and Applications. Lecture Notes for ENGO 625. Department of Geomatics Enginering, the University of Calgary.

Langley, R. (1996). Propagation of the GPS signal. In *GPS for Geodesy*, Lecture Notes in Earth Sciences, pages 103–140. Springer-Verlag. Edited by Kleusberg, A. and Teunissen P. J. G.

Langley, R. B. (2000). GPS, the Ionosphere, and the Solar Maximum. *GPS World*, 11(7):44–49.

Leick, A. (1990). *GPS Satellite Surveying*. John Wiley & Sons, first edition.

Lemoine, F. G., Smith, D., Smith, R., Kunz, L., Pavlis, E., Pavlis, N., Klosko, S., Chinn, D., Torrence, M., Williamson, R., Cox, C., Rachlin, K., Wang, Y., Kenyon, S., Salman, R., Trimmer, R., Rapp, R., and Nerem, S. (1996). The development of the NASA GSFC and DMA joint geopotential model. In *Proc. Symp. on Gravity, Geoid and Marine Geodesy, Tokyo*, pages 461–469.

Lutz, P. and Gounon, R. (2001). Thales Navigation's LRK, an Improved RTK Solution for Long Baselins: Technique, Applications and Field Results. In *Proceedings of the 14th International Technical Meeting of the Satellite Division of the Institute of Navigation (ION GPS-2001)*, pages 1027–1038.

Marel, H. v. d. (1998). Active GPS control stations. In *GPS for Geodesy*, Lecture Notes in Earth Sciences, pages 157–179. Springer-Verlag, 2 edition. Edited by Kleusberg, A. and Teunissen P. J. G.

Mendes, V. B. (1999). Modelling the neutral-atmosphere propagation delay in radiometric space techniques, Ph.D. dissertation. Technical Report 199, Department of Geodesy and Geomatics Engineering, University of New Brunswick, Fredricton, New Brunswick.

Merrigan, M. J., Swift, E. R., Wong, R. F., and Saffel, J. T. (2002). A Refinement to the World Geodetic System 1984 Reference Frame. In *Proceedings of the 15th International Technical Meeting of the Satellite Division of the Institute of Navigation (ION GPS-2002)*.

Misra, P. and Enge, P. (2001). *Global Positioning System, Signals, Measurements and Performance*. Ganga-Jamuna Press.

Nichols, J., Hansen, A., Walter, T., and Enge, P. (2000). High-Latitude Measurements of Ionospheric Scintillation Using the NSTB. *Navigation: Journal of The Institute of Navigation*, 47(2):112–120.

Niell, A. E. (1996). Global mapping functions for the atmosphere delay at radio wavelengths. *Journal of Geophysical Research*, 101(B2):3227–3246.

NIMA (2000). Department of Defence, World Geodetic System 1984. Technical Report 8350.2, 3. edition, amendment 1, January 2000, National Imagery and Mapping Agency.

Odijk, D., Teunissen, P. J. G., and Tiberius, C. C. J. M. (2002). Triple Frequency Ionosphere-free Phase Combinations for Ambiguity Resolution. In *Proceedings of the European Navigation Conference - GNSS2002, Copenhagen*. Nordic Institute of Navigation. In press.

Pany, T., Pesec, P., and Stangl, G. (2001a). Atmospheric GPS Slant Path Delays and Ray Tracing Through Numerical Weather Models, a Comparison. *Physics and Chemistry of the Earth*, 26A(3):183–188.

Pany, T., Pesec, P., and Stangl, G. (2001b). Elimination of Tropospheric Path Delays in GPS Observations with the ECMWF Numerical Weather Model. *Physics and Chemistry of the Earth*, 26A(6-8):487–492.

Papitashvili, V. O., Gromova, L. I., Popov, V. A., and Rasmussen, O. (2001). Nothern Polar Cap magnetic activity index PCN: Effective area, universal time, seasonal and solar cycle variations. Technical Report 01-01, Danish Meteorological Institute.

Penna, N. T., Dodson, A. H., and Chen, W. (2001). Assessment of EGNOS Tropospheric Correction Model. *The Journal of Navigation*, 54(1):37–55.

Poppe, B. B. (2000). New Scales Help Public, Technicians Understand Space Weather. *EOS, Transactions of the American Geophysical Union*, 81(29).

Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. (1997). *Numerical Recipes in C - The Art of Scientific Computing*. Cambridge University Press, 2 edition.

Ralston, A. and Rabinowitz, P. (1978). *A First Course in Numerical Analysis*. International Series in Pure and Applied Mathematics. McGraw-Hill, 2nd edition.

Raquet, J. (1998). *Development of a Method for Kinematic GPS Carrier-Phase Ambiguity Resolution Using Multiple Reference Receivers*. PhD thesis, Department of Geomatics Engineering, The University of Calgary.

Raquet, J. and Lachapelle, G. (2001). RTK Positioning with Multiple Reference Stations. *GPS World*, 12(4):48–53.

Raquet, J., Lachapelle, G., and Fortes, L. P. (1998a). Use of a Covariance Analysis Technique for Predicting Performance of Regional Area Differential Code and Carrier-Phase Networks. In *Proceedings of the 11th International Technical Meeting of the Satellite Division of the Institute of Navigation (ION GPS-98)*, pages 1345–1354.

Raquet, J., Lachapelle, G., and Melgaard, T. E. (1998b). Test of a 400 km x 600 km Network of Reference Receivers for Precise Kinematic Carrier-Phase Positioning in Norway. In *Proceedings of the 11th International Technical Meeting of the Satellite Division of the Institute of Navigation (ION GPS-98)*, pages 407–416.

Rasmussen, A., Sørensen, J. H., Nielsen, N. W., and Amstrup, B. (2000). Uncertainty of Meteorological Parameters from DMI-HIRLAM. Technical Report 00-07, Danish Meteorological Institute.

Rothacher, M. (1997). GPS satellite orbits, orbit determination, and the IGS. In *Geodetic Applications of GPS, Lecture Notes for Nordic Autumn School*, number 16 in Reports in Geodesy and Geographical Information Systems, pages 55–108. National Land Survey of Sweden. Edited by Bo Jonsson.

Saastamoinen, J. (1973). Contributions to the Theory of Atmospheric Refraction. *Bulletin Geodesique*, 105, 106, 107:279–298, 383–397, 13–34. Printed in three parts.

Sass, B. H., Nielsen, N. W., Jørgensen, J. U., Amstrup, B., and Kmit, M. (2000). The Operational HIRLAM System at DMI. Technical Report 00-26, Danish Meteorological Institute. Copenhagen.

Schueler, T. (2001). *On Ground-based GPS Tropospheric Delay Estimation.* PhD thesis, Universität der Bundeswehr, München.

Schueler, T., Hein, G. W., and Eissfeller, B. (2000a). Improved Tropospheric Delay Modeling Using an Integrated Approach of Numerical Weather Models and GPS. In *Proceedings of the 13th International Technical Meeting of the Satellite Division of the Institute of Navigation (ION GPS-2000). Salt Lake City, Utah.*, pages 600 – 615.

Schueler, T., Hein, G. W., and Eissfeller, B. (2000b). On the Use of Numerical Weather Fields for Troposphere Delay Estimation in Wide Area Augmentation Systems. In *Proceedings of GNSS 2000. Edinburgh*, pages 1077–1091. Royal Institue of Navigation.

Seeber, G. (1993). *Satellite Geodesy. Foundations, Methods and Applications.* Walter de Gruyter.

Shaw, M., Sandhoo, K., and Turner, D. (2000). Modernization of the Global Positioning System. *GPS World*, 11(9):36–44'.

Skone, S. (1994). Characterizing the Magnetic Field at Geostationary Orbit. Master's thesis, Department of Physics, University of Alberta, Edmonton.

Skone, S. (1998). *Wide Area Ionosphere Grid Modelling in the Auroral region.* PhD thesis, Department of Geomatics Engineering, The University of Calgary.

Skone, S., Knudsen, K., and Jong, M. (2001). Limitations in GPS Reciever Tracking Performance Under Ionospheric Scintillation Conditions. *Physics and Chemistry of the Earth*, 26A(6-8):613–621.

Solheim, F. S., Vivekanandan, J., Ware, R. H., and Rocken, C. (1999). Propagation delays induced in GPS signals by dry air, water vapor, hydrometeors, and other particulates. *Journal of Geophysical Research*, 104(D8):9663–9670.

Spilker, J. J. (1996). Tropospheric Effects on GPS. In *Global Positioning System: Theory and Applications, Vol. I*, Progress in Astronautics and Aeronautics, Vol. 163., pages 517–546. American Institute of Astronautics and Aeronautics. Edited by B.W. Parkinson, J. J. Spilker, P. Axelrad, and P. Enge.

Strang, G. and Borre, K. (1997). *Linear Algebra, Geodesy, and GPS.* Wellesley-Cambridge Press.

Sun, H., Cannon, M. E., and Melgaard, T. (1999). Real-Time GPS Reference Network Carrier Phase Ambiguity Resolution. In *Proceedings of the 1999 National Technical Meeting of the Institute of Navigation, San Diego, California.*

Teunissen, P. J. G. and Tiberius, C. C. J. M. (1994). Integer Least-square Estimation of the GPS Phase Ambiguities. In *Proceedings of the International Symposium on Kinematic Systems in Geodesy, Geomatics and Navigation, KIS'94.* Department of Geomatics Engineering, the University of Calgary.

Torge, W. (1989). *Gravimetry.* Walter de Gruyter.

Torge, W. (1991). *Geodesy.* Walter de Gruyter, 2nd edition.

Townsend, B. and Jensen, A. B. O. (2001). Using a Virtual Reference Station to Compensate for Coordinate Transformations in GPS Surveying. In *Proceedings of the International Symposium of Kinematic Systems in Geodesy, Geomatics and Navigation (KIS 2001).* Department of Geomatics Engineering, The University of Calgary.

Townsend, B., Lachapelle, G., Fortes, L. P., Melgaard, T., Nørbech, T., and Raquet, J. (1999). New Concepts for a Carrier Phase Based GPS Positioning Using a National Reference Station Network. In *Proceedings of the 1999 National Technical Meeting of the Institute of Navigation, San Diego, California.*

Trimble (1996). *GPSurvey - Software User's Guide.*

Tsujii, T., Wang, J., Dai, L. Rizos, C., Harigae, M., Inagaki, T., Fujiwara, T., and Kato, T. (2001). A Technique for Precise Positioning of High Altitude Platforms System (HAPS) Using a GPS Ground Reference Network. In *Proceedings of the 14th International Technical Meeting of the Satellite Division of the Institute of Navigation (ION GPS-2001)*, pages 1017–1026.

Vedel, H. (2000). Conversion of WGS84 geometric heights to NWP model HIRLAM geopotential heights. Technical Report 00-04, Danish Meteorological Institute. Copenhagen.

Vedel, H., Mogensen, K. S., and Huang, X.-Y. (2001). Calculation of zenith delays from meteorological data, comparison of NWP model, radiosonde and GPS delays. *Physics and Chemistry of the Earth*, 26A(6-8):497–502.

Vollath, U., Buecherl, A., Landau, H., Pagels, C., and Wagner, B. (2000). Multi-Base RTK Positioning Using Virtual Reference Stations. In *Proceedings of the 13th International Technical Meeting of the Satellite Division of the Institute of Navigation (ION GPS-2000). Salt Lake City, Utah.*, pages 123–131.

Wallace, J. M. and Hobbs, P. V. (1977). *Atmospheric Science, an Introductory Survey.* Academic Press Inc.

Walpersdorf, A., Calais, E., Haase, J., Eymard, L., Decbois, M., and Vedel, H. (2001). Atmospheric Gradients Estimated by GPS Compared to a High Resolution Numerical Weather Prediction (NWP) Model. *Physics and Chemistry of the Earth*, 26A(3):147–152.

Wanninger, L. (1993). Effects of the Equatorial Ionsphere on GPS. *GPS World*, 4(7):48–54.

Wanninger, L. (1997). Real-time Differential GPS Error Modelling in Regional Reference Station Networks. In *Advances in Positioning and Reference Frames. IAG Scientific Assembly, Rio de Janeiro.*, pages 86–92. Springer.

Weber, R., Ray, J., and Kouba, J. (2002). Review of IGS Analysis Products. Draft version of Position Paper from Network and Analysis Center Workshop, IGS, Ottawa.

Wübbena, G., Bagge, A., Seeber, G., Böder, V., and Hankemeier, P. (1996). Reducing Distance Dependent Errors for Real-Time Precise DGPS Applications by Establishing Reference Station Networks. In *Proceedings of the 9th International Technical Meeting of the Satellite Division of the Institute of Navigation (ION GPS-96). Kansas City, Missouri*, pages 1845–1852.

Yang, X., Sass, B. H., Elgered, G., Johansson, J. M., and Emardson, T. R. (1999). A Comparison of Precipitable Water Vapor Estimates by an NWP Simulation and GPS Observations. *Journal of Applied Meteorology*, 38:941–956.

Zhang, J. (1999). Precise Estimation of Residual Tropospheric Delays in a Spatial GPS Network. In *Proceedings of the 12th International Technical Meeting of the Satellite Division of the Institute of Navigation (ION GPS-99). Nashville, Tennessee*, pages 1391–1400.

Zhang, J. and Lachapelle, G. (2001). Precise estimation of residual tropospheric delays using a regional GPS network for real-time kinematic applications. *Journal of Geodesy*, 75(5-6):255–266.

A

INDICES FOR IONOSPHERIC ACTIVITY

This appendix contains a description of the most commonly used indices for ionospheric activity, supplemented with examples showing the indices and plots of activity from both calm days and for a day with a major ionospheric storm. The last part of the appendix contains a description of selected forecast services.

The indices described provide a good general description of the global or regional activity, but because of the high temporal and geographical variability of the ionospheric activity, it is always better to evaluate the activity at any given site by using local geomagnetic data (Skone, 1998).

A.1 Ionospheric Indices

Magnetic storms and substorms can be characterized by the global Kp, Ap or DST indices. These are all determined by the storm signatures that show up as variations in the geomagnetic field when observed with magnetometers.

K and Kp index

A local K-index[1] is a 3-hour index for geomagnetic activity observed with a magnetometer. The index is a measure of the observations relative to "an estimated value of the local quiet daily variation" (Campbell, 1997). Magnetic field strength is measured continuously, and the logarithmic K index is computed based on the difference between the maximum and minimum observed field strength variations during the 3-hour interval.

The K-index vary from 0 (quiet) to 9 with indication of thirds by + and -. For instance 3, 3+, 4- , 4 .

The planetary or global K index, called the Kp index, is the mean standardized K index from 13 observatories including Brorfelde in Denmark. The observatories are distributed between 44° and 60° North or South, and all observatories are thus located outside the auroral and equatorial regions. Furthermore there is a majority of observatories on the Northern hemisphere (Campbell, 1997).

The Kp index can thus give a good indication of the total global variations in geomagnetic activity throughout a day, but the Kp index cannot indicate where the activity was going on. A high Kp number is normally caused by a high global geomagnetic activity, and a low Kp number is caused by a low global activity. But a medium number, e.g. Kp=6, can be caused by a medium global activity, or by a high activity in some places and a low activity in other places.

NOAA has lately developed a new scale for the Kp-values for easier explanation of the "size" of geomagnetic storms for "the public". By the new scale a geomagnetic storm activity is rated from G1–G5 corresponding to a minor, moderate, strong,

---

[1] K stands for Kennziffer, German for a logarithmic value (Campbell, 1997).

severe, or extreme storm. A minor storm and extreme storm corresponds to the Kp values of 5 and 9 respectively. The new scale is described by Poppe (2000) who also gives the relations to the Kp values.

a and Ap index

The local a-index (denoted by a small a) ranges from 0 to 400 and represents the logarithmic K-value converted to a linear scale. If the a-value is above 30 it is an indication of a local storm activity. The a index is also a 3-hour index determined at various observatories, and the planetary ap index is a mean of the a indices from the 13 observatories mentioned above.

The combined Ap-index (denoted by capital A) is determined once a day as a mean of the eight 3-hour ap values. The Ap index is thus a daily index for geomagnetic activity, whereas the Kp index is a 3-hour index for global geomagnetic activity.

A table with corresponding logarithmic Kp and linear Ap values is found in Campbell (1997).

A “major ”storm is defined by NOAA as a storm causing an Ap index higher than 40 (Skone, 1998).

Kp and Ap indices are available on the internet from NOAA (ftp://ftp.ngdc.noaa.gov/stp/geomagnetic_data/indices/kp_ap/)

DST index

The Disturbed Storm Time (DST) index was adopted in 1964 and is based on observations from four magnetometer stations located close to the Equator and distributed in longitude (Campbell, 1997). The DST is determined similarly to the Kp index, but it is a 1 hour index and thus gives a better picture of the variability throughout a day. The purpose of the DST is to indicate magnetic storm activity, and since the observatories are all located close the Equator it is a good measure of the activities in this region. The index is not equally appropriate for high latitudes and especially not for the auroral region.

The PC index

In 1999 the International Association of Geomagnetism and Aeronomy officially adopted a polar cap index called the PC-index. It is based on observations from Qaanaq in Northern Greenland and Vostock in Antarctica. The PC-index for the Northern hemisphere is available from the Danish Meteorological Institute with a 1 minute resolution since 1975, and the index is described in detail by Papitashvili et al. (2001).

Other indices

Campbell (1997) mentions several other indices, for instance the Kn and Ks indices which are variations of the Kp index, with a better representation of the activities in the Northern and Southern hemisphere. An AE index has been developed to indicate auroral activity based on data from observatories located within, or close to, the auroral oval.

A.1.1 Some examples

This section contains some examples of how the indices can be used to indicate the level of ionospheric activity. Data from April 6–7 2000 is used as an example of an active day, and data from September 5 and 14, 2000 is used as examples for quiet days.

High ionospheric activity

A major geomagnetic storm took place around midnight UTC on April 6–7, 2000. The storm caused clear northern lights over central Europe and as far South as Florida. The event, which was caused by a coronal mass ejection from the sun on April 4, is described by Langley (2000) and on http://www.spaceweather.noaa.gov/stories/sw3.htm.

During the storm, GPS data was collected by KMS, National Survey and Cadastre - Denmark with Trimble 4000ssi receivers, which use a codeless L2 carrier tracking technique. The GPS data was collected with the purpose of positioning a number of off shore oil rigs with respect to reference stations located inland in Denmark, and the survey campaign thus included several long baselines of 2–300 km. During the storm all observation on L2 were unusable, and the lack of L2 data prevented the use of linear combinations of the L1 and L2 observables, making it impossible to solve ambiguities for the long baselines.

The Ap index for the two days were 82 and 74 respectively. The Kp indices from NOAA are shown in Figure A.1.

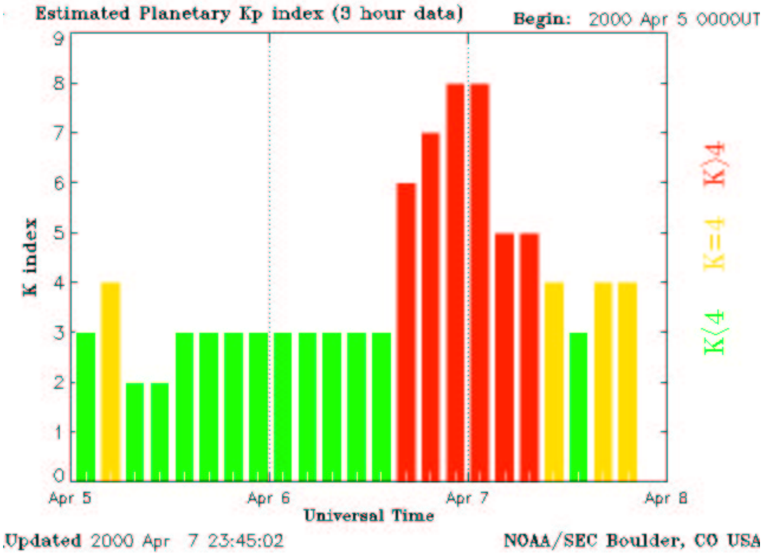

Figure A.1: Kp indices for April 5, 6, and 7, 2000. From gopher://sec.noaa.gov/.

The geomagnetic activity is shown in Figure A.2 by a magnetogram from the Brorfelde Magnetic Observatory (BFE), which is located about 50 km West of Copenhagen.

The plot shows the observed magnetic field strength in nanotesla for the three vector components (X, Y, Z) of the geomagnetic field during the two days. A detailed description of the plot can be found on http://web.dmi.dk/fsweb/soljord/projekter/geofys_obs/latest_mag_bfe.html.

The magnetogram from Brorfelde was generated using the INTERMAGNET service on http://www.intermagnet.org/, where various magnetic data can be obtained from participating observatories. The data from Brorfelde was made available by the Danish Meteorological Institute, which runs the magnetograph.

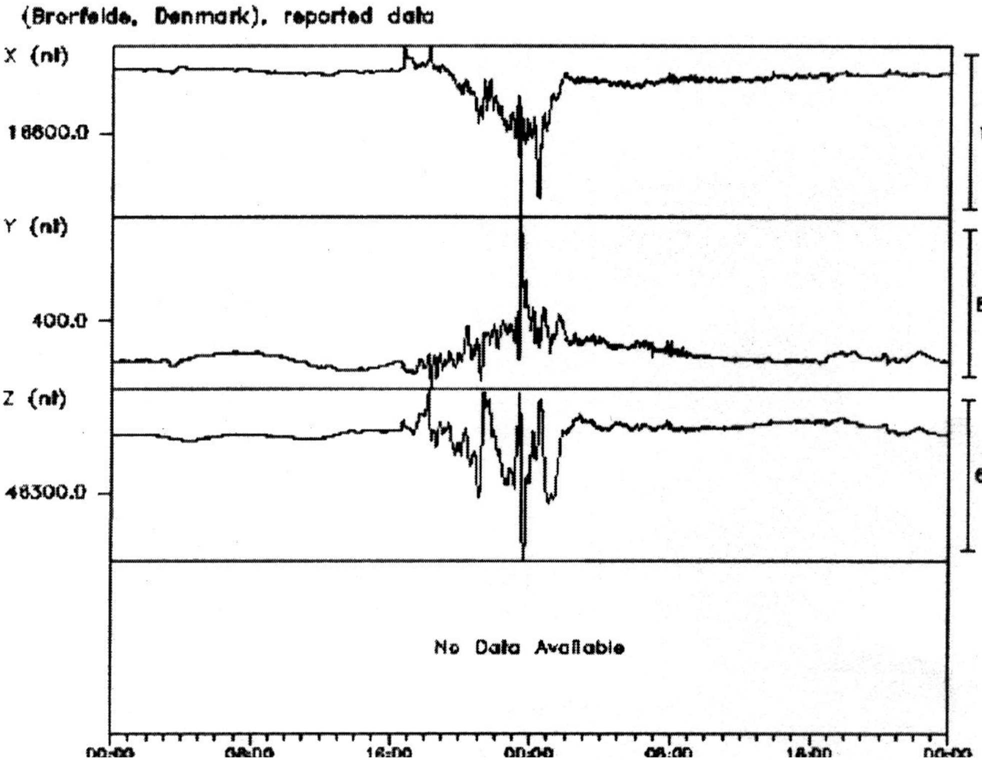

Figure A.2: Magnetogram from Brorfelde for April 6–7, 2000. Derived from http://ub.nmh.ac.uk/htbin/imagform

Figure A.3 shows the components of the IMF field strength in the solar–magnetospheric reference system, for the same two days in April. When the Z component is negative, the geomagnetic field will be disturbed as described in Chapter 2. When comparing this plot with the magnetogram from Brorfelde (Figure A.2), a correlation between the strong southward IMF and a high level of geomagnetic activity can be seen. Also note that the start of the geomagnetic disturbances in the Brorfelde plot was delayed about 1 hour compared to the start of the variations in the IMF field strength.

The plot showing the components of the IMF field was derived from data from the WIND satellite given on http://eqs1.plasma.mpe-garching.mpg.de/cdaweb/istp_public/.

Figure A.4 shows the DST index for the month of April 2000. A significant dip in the curve shows the storm on April 6 and 7.

Finally, Figure A.5 shows the PC-index for the two days, and these plots indicate that the storm indeed had a global impact. The storm is noticed by a sudden increase in the PCN-index just after 17:00 hours UTC on April 6.

Low ionospheric activity

In Chapters 4 and 5, GPS data collected on September 5 and 14, 2000 was used for testing the use of NWPs in GPS positioning. These two days were both characterized by a low ionospheric activity.

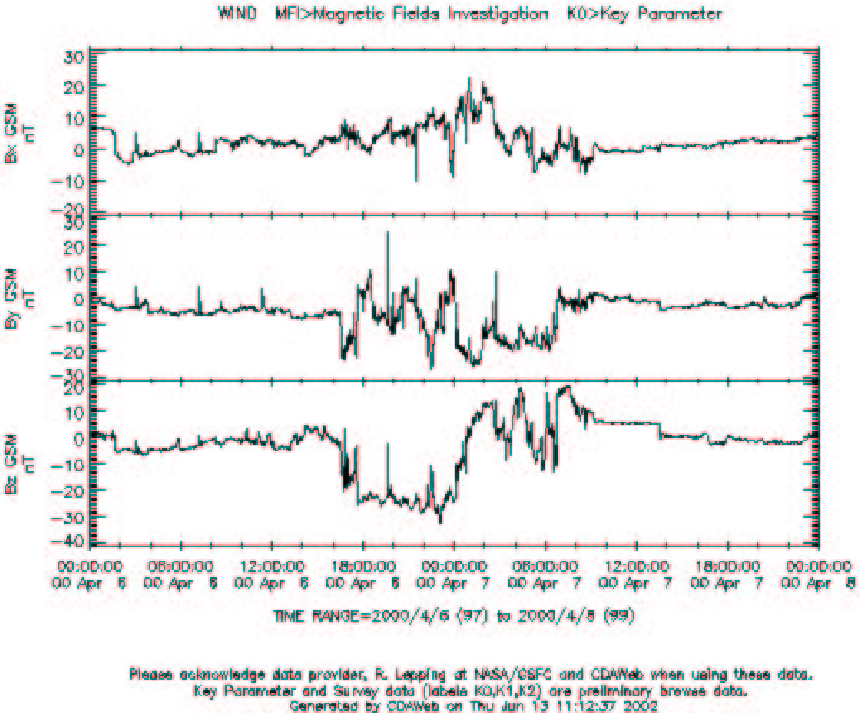

Figure A.3: IMF field strength in nanotesla, for April 6–7, 2000. From http://eqs1.plasma.mpe-garching.mpg.de/cdaweb/istp_public/

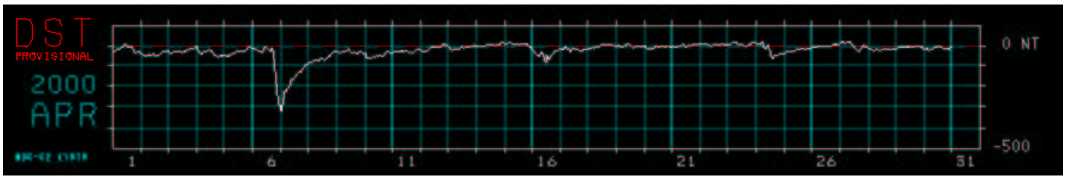

Figure A.4: DST indices for the month of April 2000. From http://swdcdb.kugi.kyoto-u.ac.jp/dstdir/

The Ap index for the two days was 7 and 4, respectively. The Kp indices from NOAA are shown in Figures A.6 and A.7, and the indices range from 1 to 3 for the two days.

The geomagnetic activity is shown by magnetograms from the Brorfelde Magnetic Observatory (BFE) in Figures A.8 and A.9. Note the difference in the scale of the plots when compared to the plot in Figure A.2. The max-min scale is given on the right hand side of the plots. Furthermore the diurnal variation of a fairly quiet geomagnetic day is seen in the plots.

Figures A.10 and A.11 show the components of the IMF field strength for the same two days in September. The Z-component of the IMF is negative for parts of the days, but the IMF field strength is very low compared to the strength shown in Figure A.3. So even though there is some activity going on, the disturbances are not powerful enough to cause a global impact on GPS signals.

The DST index for the month of September is given in Figure A.12, and it shows a DST index of basically zero for both September 5 and 14, 2000.

Finally, the PC-index for the two days are given in Figures A.13 and A.14. Again

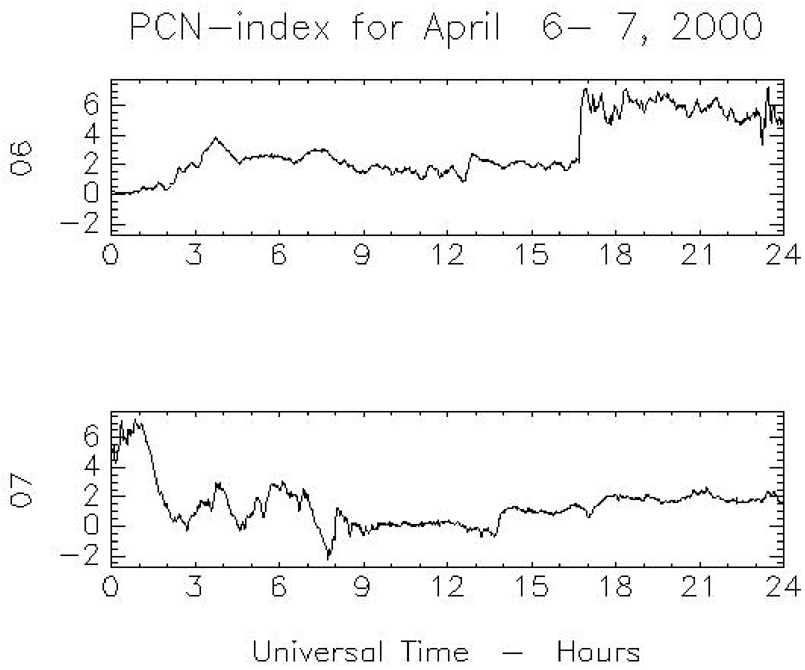

Figure A.5: Northern Polar Cap index for April 6 and 7, 2000. From http://web.dmi.dk/fsweb/projects/wdcc1/pcn/pcn.html

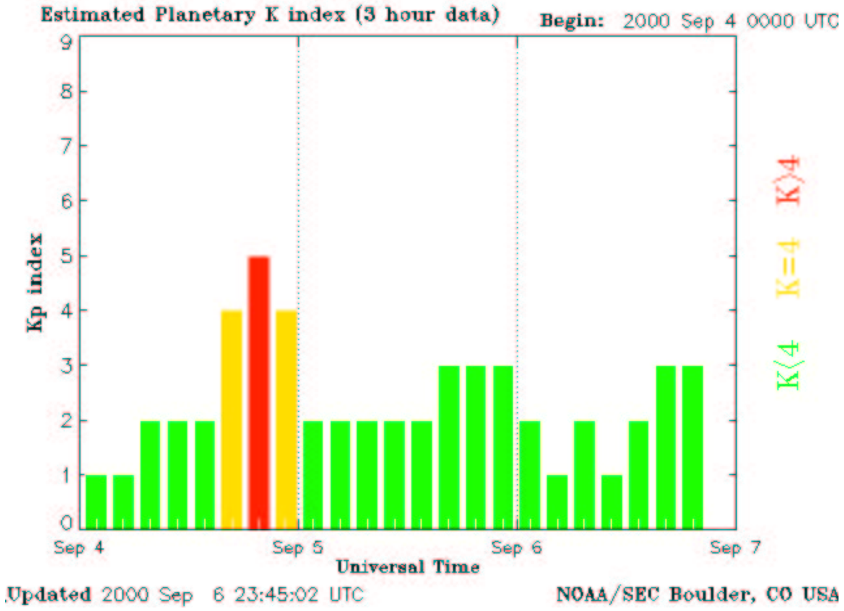

Figure A.6: Kp indices for September 5, 2000. From gopher://sec.noaa.gov/.

notice the difference in scale when compared to the plots from April 6 and 7, 2000.

The plots for the September days show that even in the polar regions some activity is taking place, indicating the active nature of the geomagnetic field.

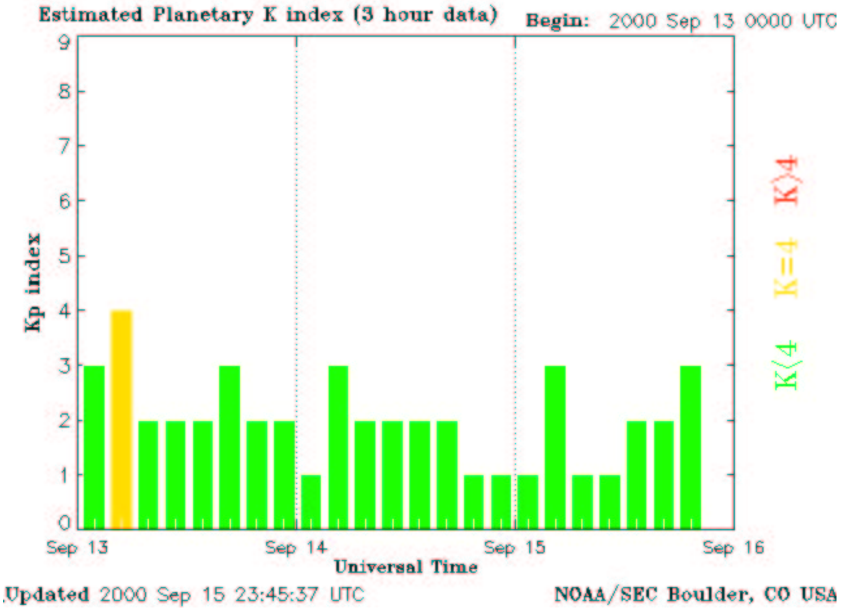

Figure A.7: Kp indices for September 14, 2000. From gopher://sec.noaa.gov/.

.

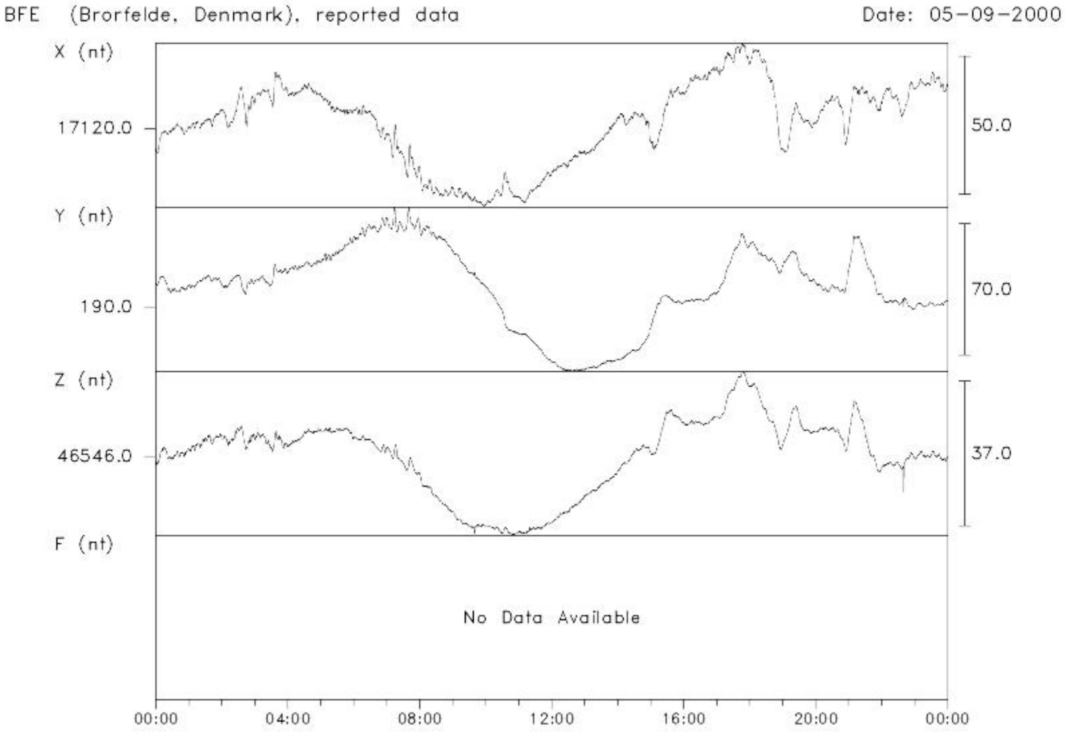

Figure A.8: Magnetogram from Brorfelde for September 5, 2000. From http://www.intermagnet.org/

A.2 Ionospheric Forecasts and Warnings

A number of forecasts and warnings of ionospheric activities are available on the internet, and they can be a very useful tool for GPS users when planning survey

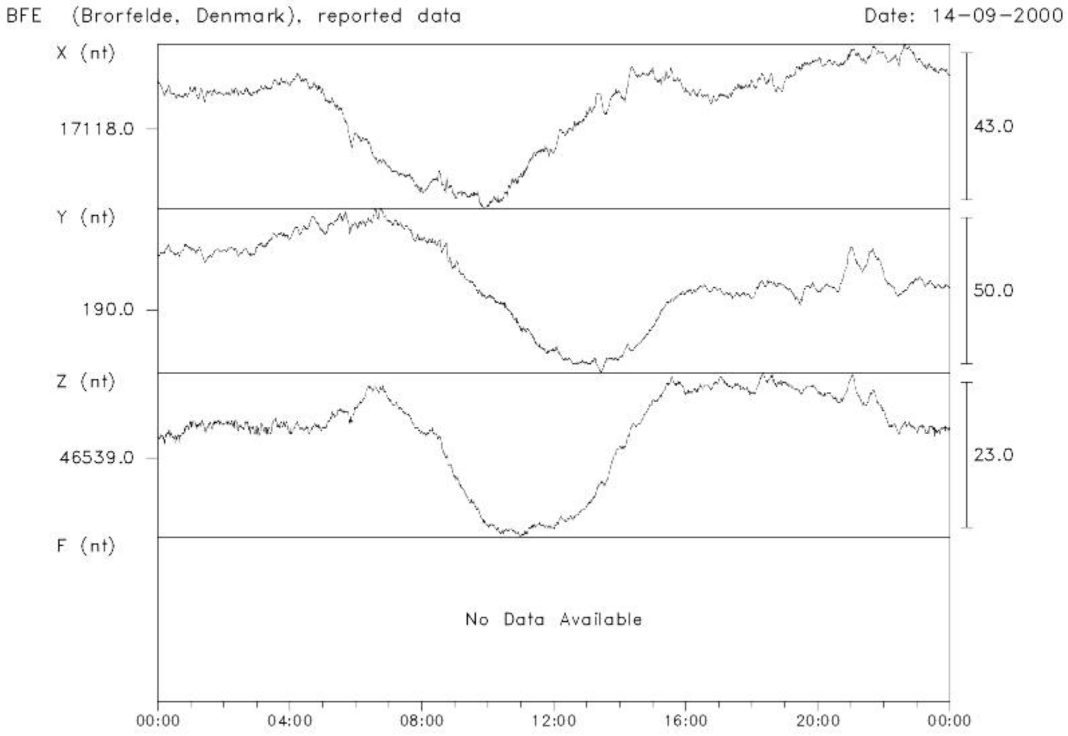

Figure A.9: Magnetogram from Brorfelde for September 14, 2000. From http://www.intermagnet.org/

activities. RTK users may especially benefit from this if the ionosphere–free linear combination of the L1 and L2 is not used by the RTK software in the equipment.

The following is a description of some of the forecasts and warnings available, but several internet sites related to the subject exist, so this is not a complete listing.

The sun rotates with a period of 27 days, and 27-day sunspot predictions can thus be generated. These are used by NOAA's Space Environment Center for predicting Ap and Kp indices, which can be found at http://www.sec.noaa.gov/nav/, along with various forecasts and warnings relevant for navigation users.

NOAA also has a general alert web site, with information about the space weather (http://www.sec.noaa.gov/alerts/index.html).

The current boundaries of the auroral zone are given on NOAA's web site at http://www.sel.noaa.gov/pmap/index.html, and an example is given in Figure A.15. These plots are real-time images, and their usefulness for planning purposes is thus limited.

Lund Space Weather Center in Sweden provides forecasts for a number of solar events. Predictions of Kp indices and auroral activity are also given on their web site at http://www.irfl.lu.se/HeliosHome/forecastpage.html

Warnings for geomagnetic activity and an "auroral alert" can be found on http://www.ips.gov.au/asfc/current/index.html, along with listings of various current conditions. The web site is maintained by IPS Radio and Space Services, a part of the Department of Industry Science and Resources under the Australian govern-

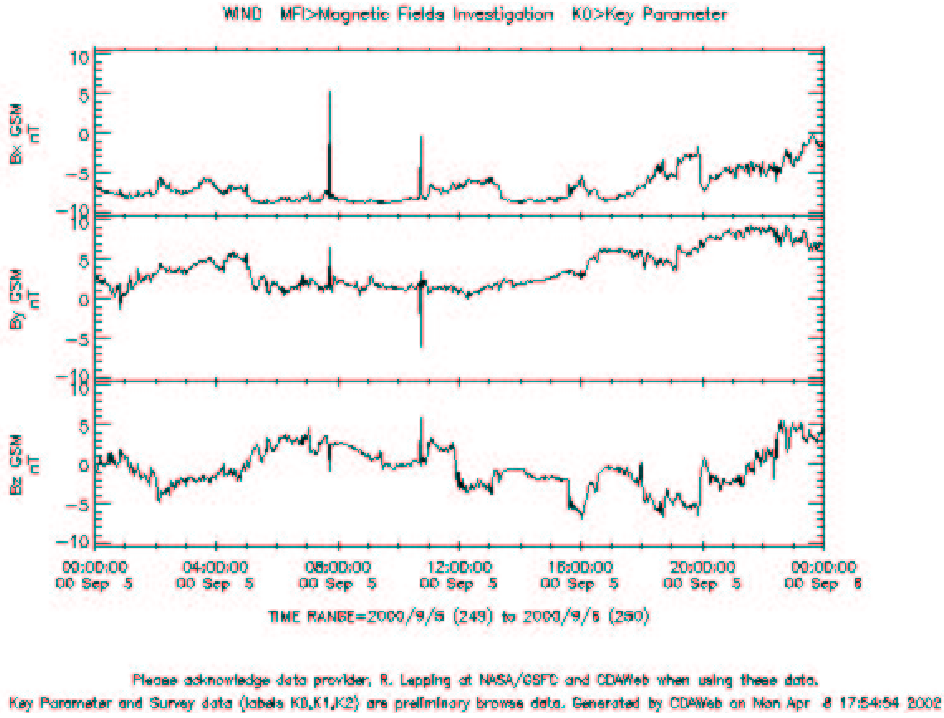

Figure A.10: IMF field strength in nanotesla, for September 5, 2000. From http://eqs1.plasma.mpe-garching.mpg.de/cdaweb/istp_public/

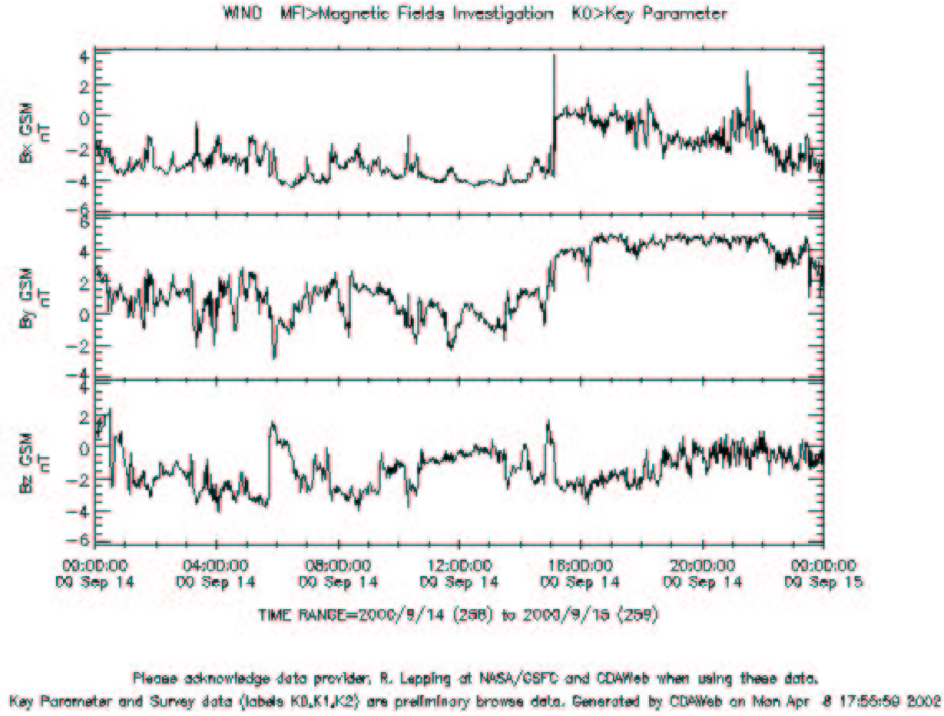

Figure A.11: IMF field strength in nanotesla, for September 14, 2000. From http://eqs1.plasma.mpe-garching.mpg.de/cdaweb/istp_public/

ment. The auroral alerts provide "notification of times when aurora are more likely to be seen at more Equatorial latitudes than is normal". This means that the alerts

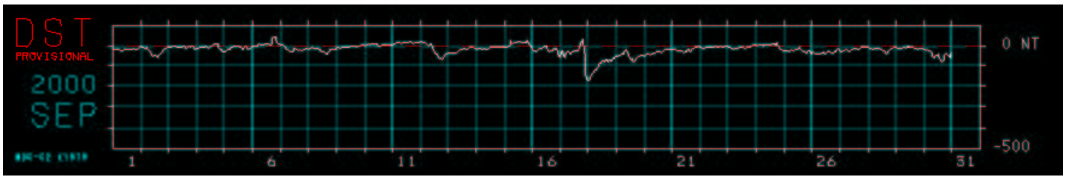

Figure A.12: DST indices for the month of September 2000. From http://swdcdb.kugi.kyoto-u.ac.jp/dstdir/

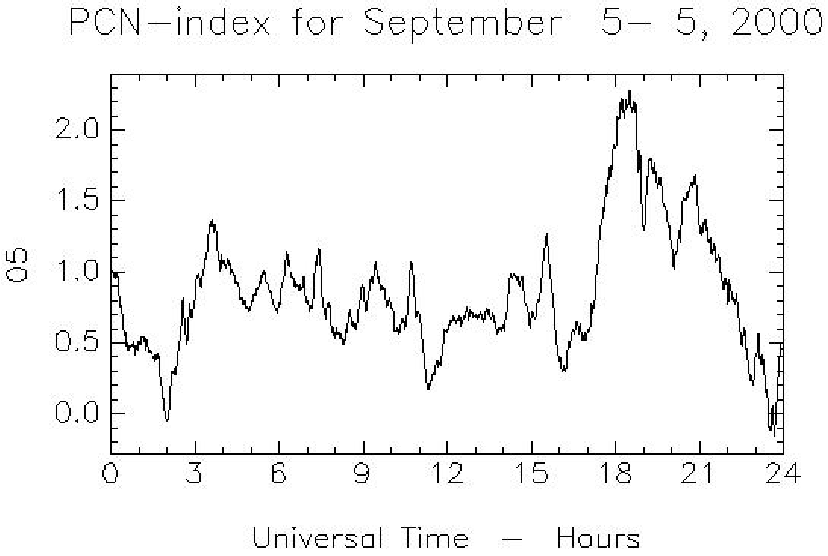

Figure A.13: Northern Polar Cap index for September 5, 2000. From http://web.dmi.dk/fsweb/projects/wdcc1/pcn/pcn.html

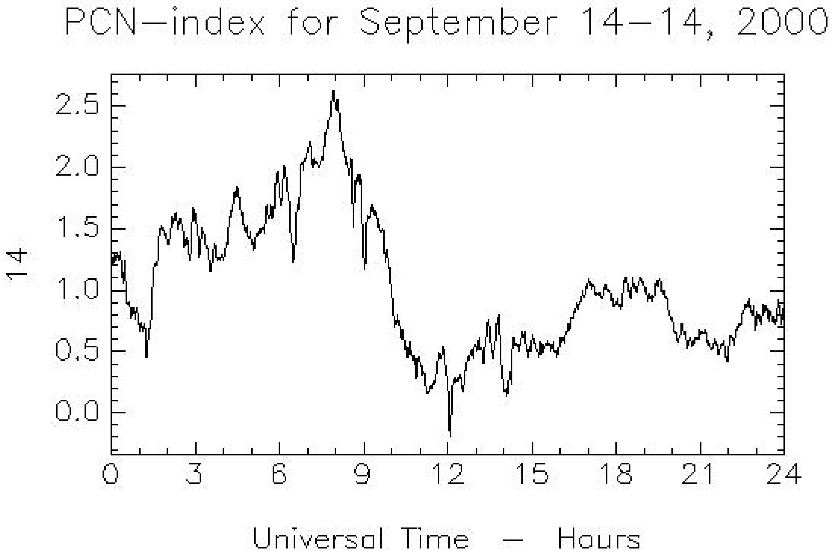

Figure A.14: Northern Polar Cap index for September 14, 2000. From http://web.dmi.dk/fsweb/projects/wdcc1/pcn/pcn.html

might actually not be useable for GPS observations within the auroral regions, since "normal" auroral activity will not trigger a warning.

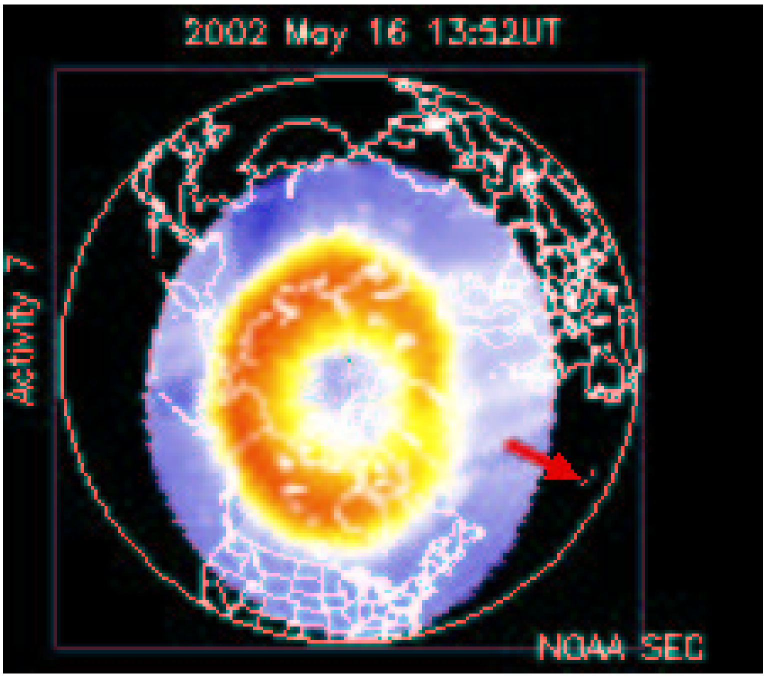

Figure A.15: Location of the auroral zone in the Northern hemisphere on May 16, 2002 at 13:52 UTC. Note that the "activity" is stronger on the night side of the Earth (the left side in the plot), and that the oval is smallest in the early afternoon hours. The red arrow points towards the noon meridian. The plot is from http://www.sel.noaa.gov/pmap/index.html

B

GPS Stations and Data Quality

This appendix contains information about the GPS stations and the quality of the GPS data.

B.1 GPS stations

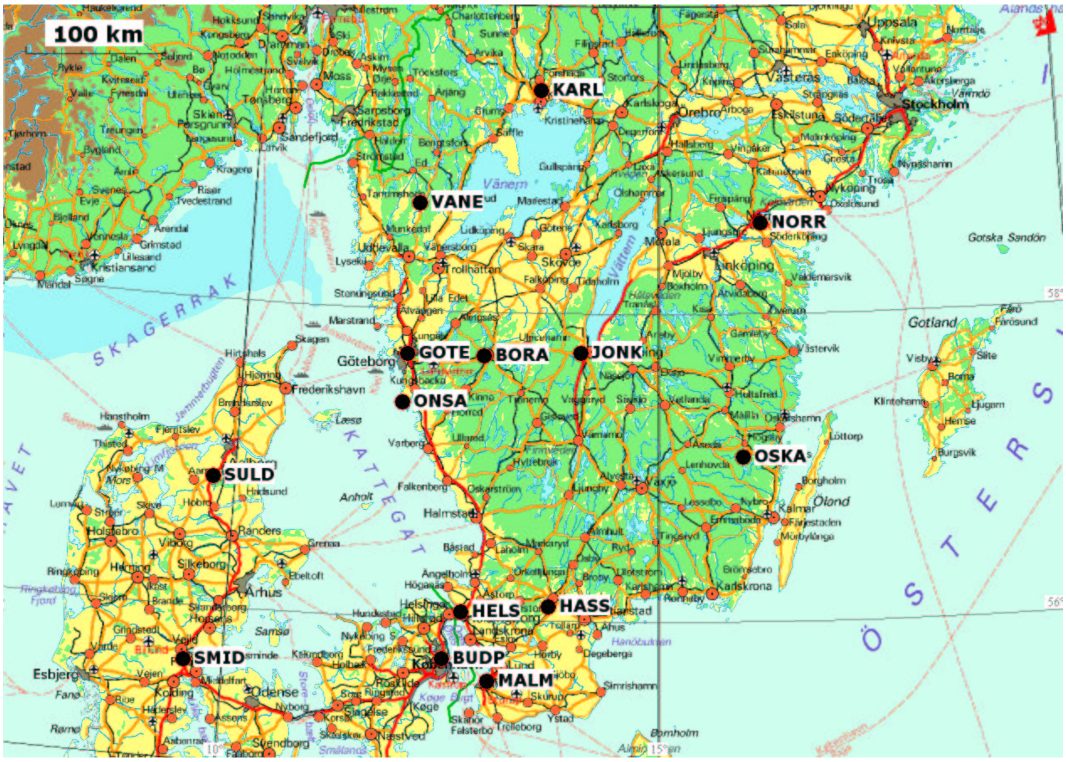

Figure B.1: Location of the 14 GPS sites.

Figure B.1 shows the location of the GPS stations, and the station positions are given in Table B.1. Both ellipsoidal and spherical latitudes are given in the table, since both types of latitudes are used: ellipsoidal latitudes in connection with the ray tracer, and spherical latitudes in connection with the zenith delays. The station height above sea level, the orthometric height, is also given in the table.

Station Abbreviation	Location Name	Ellipsoidal Latitude	Spherical Latitude	Longitude	Orthometric Height
ONSA	Onsala	57.926°	57.220°	11.926°	9.218 m
NORR	Norrköping	58.590°	58.419°	16.246°	13.005 m
GOTE	Göteborg	57.708°	57.534°	11.966°	21.053 m
MALM	Malmö	55.601°	55.422°	13.030°	46.635 m
HELS	Helsingborg	56.046°	55.868°	12.709°	55.370 m
BUDP	Buddinge	55.739°	55.560°	12.500°	57.922 m
HASS	Hässleholm	56.092°	55.914°	13.718°	78.513 m
SULD	Suldrup	56.842°	56.665°	9.742°	82.219 m
SMID	Smidstrup	55.641°	55.461°	9.559°	82.738 m
KARL	Karlstad	59.444°	59.275°	13.506°	82.956 m
OSKA	Oskarshamn	57.066°	56.890°	15.997°	119.436 m
VANE	Vänersborg	58.693°	58.522°	12.035°	134.836 m
BORA	Borås	57.715°	57.541°	12.891°	185.363 m
JONK	Jönköping	57.745°	57.572°	14.060°	227.428 m

Table B.1: Position information for the 14 Swedish and Danish GPS stations

B.2 GPS data quality

A quality assesment of the data was carried out using the TEQC software from UNAVCO[1]. The results of the analyses from both days are given in Tables B.2 and B.3. An elevation mask of 10° was used for the quality check.

In the tables, "hrs" is the observation time span, "# expt" indicates the expected number of observations based on the knowledge of the satellite orbits and the approximate station position, "# have" is the actual number of observations, and "%" is the actual number of observations in percent of what could be expected. For these stations the percentage varies between 88% and 100%, indicating a "normal" amount of data.

"mp1" and "mp2" are the mean multipath values for the L1 and L2 frequencies, respectively. A higher value indicates a higher noise level in the data. The amount of multipath is generally higher for the Swedish stations than for the Danish stations.

Finally, "o/slips" indicates the number of observations per cycle slip. Here, a higher value indicates a low number of cycle slips and thus better data quality. These values are related to the number of observations and the amount of multipath, and for a few of the Swedish stations, where both a relatively low percentage of observations are received and a high amount of multipath is observed, the number of observations per cycle slip is relatively low. Overall, the noise level is of about the same size for both days.

---

[1] UNAVCO is an American organisation that supports universities and other research institutions in their use of GPS. UNAVCO is funded by the US National Science Foundation and NASA. http://www.unavco.ucar.edu/

Station	hrs	# expt	# have	%	mp1	mp2	o/slips
BORA	24.00	45747	45184	99	0.53	0.51	2824
GOTE	24.00	45807	45419	99	0.64	0.64	622
HASS	24.00	45218	45196	100	0.49	0.50	45196
HELS	24.00	53194	50917	96	0.56	0.61	342
JONK	24.00	53936	52563	97	0.44	0.49	1460
KARL	24.00	54523	49267	90	0.75	0.72	130
MALM	24.00	52950	52135	98	0.61	0.68	1159
NORR	24.00	54242	52793	97	0.59	0.63	476
ONSA	24.00	53794	50308	94	0.64	0.69	161
OSKA	24.00	53671	49471	92	0.63	0.62	283
VANE	24.00	54275	52663	97	0.64	0.85	908
BUDP	24.00	45081	44770	99	0.13	0.14	44770
SMID	24.00	45132	45126	100	0.10	0.12	2507
SULD	24.00	45497	45151	99	0.11	0.13	5644

Table B.2: TEQC results for September 5, 2000 data

Station	hrs	# expt	# have	%	mp1	mp2	o/slips
BORA	24.00	53975	47520	88	0.66	0.60	161
GOTE	24.00	53977	50389	93	0.77	0.75	201
HASS	24.00	53260	50281	94	0.62	0.60	250
HELS	24.00	53247	51314	96	0.57	0.60	828
JONK	24.00	53974	52143	97	0.42	0.47	815
KARL	24.00	54565	49360	90	0.78	0.71	131
MALM	24.00	52998	51995	98	0.60	0.66	1333
NORR	24.00	54268	52562	97	0.58	0.62	674
ONSA	24.00	53852	49501	92	0.69	0.73	90
OSKA	24.00	53697	49415	92	0.62	0.62	348
VANE	24.00	54327	52642	97	0.63	0.82	993
BUDP	24.00	45069	44551	99	0.12	0.14	44551
SMID	24.00	45132	44944	100	0.10	0.11	6421
SULD	24.00	45494	44873	99	0.11	0.13	2805

Table B.3: TEQC results for September 14, 2000 data

C

NWP Zenith Delays - Implementational Issues

This appendix describes a number of considerations and tests carried out in relation to the implementation of Equation (3.15), which is repeated below, and which has been implemented in the programme ZTDPOINT.

$$ztd = 10^{-6}\int_0^{P_{surface}} \frac{k_1 P_d}{g} dp + 10^{-6}\int_0^{P_{surface}} \frac{R_d}{\epsilon}\left(k_2 - \epsilon k_1 + \frac{k_3}{T}\right)\frac{q}{g} dp \tag{C.1}$$

C.1 Height

Equation (C.1) is arranged as an integral with respect to pressure, which is convenient since the vertical reference used in DMI-HIRLAM-E is pressure. However, knowledge of the height is necessary in order to determine the gravity for a given layer in the model.

The height is determined from the pressure. But the pressure is not given directly in the NWP, but must be calculated from the surface pressure, which is given in the model.

The pressure, P, for any layer is determined using Equation (C.2), which was provided by M. Sørensen from the Danish Meteorological Institute:

$$P = b_i + (\sigma_i * P_s) \tag{C.2}$$

b_i is a so called vertical coordinate parameter for the layer i, σ_i is the sigma scaling factor described in Chapter 3, and P_s is the pressure at the surface of the Earth.

When P has been determined for the relevant layer, the orthometric height can be determined using expression (C.3) below as suggested by H. Vedel from the Danish Meteorological Institute. The expression is derived from the equation of state, and relates a change in the logarithmic value of pressure to a change in height:

$$\int_P^{P_1} R_d T(1 - q - q/\epsilon)\, dln(P) = \int_{h_1}^{h} g\, dh \tag{C.3}$$

R_d is the gas constant, T is temperature, q is specific humidity, g is gravity, P is pressure, and h is the height for any given layer in the NWP. P_1 and h_1 are pressure and height respectively for the next layer in the model.

When the change in height is determined, it is added to the height determined for the previous layer. With this implementation, determination of the height starts at the surface of the Earth where the height is found as follows:

According to Vedel (2000) the surface geopotential, W, given in DMI-HIRLAM-E is determined as:

$$W = h_s \cdot g_H \tag{C.4}$$

where h_s is the orthometric height at the Earths surface, given by a topographic database, and g_H is the approximate surface gravity value of 9.81 m/s^2 used for generating HIRLAM. Since the surface geopotential is given in DMI-HIRLAM-E, Equation (C.4) is used directly to calculate the orthometric height of the surface of the topographic model. The accuracy of this height is not known, but it is directly related to the (also unknown) accuracy of the topographical data that forms the basis for the NWP model development.

The pressure at the surface of the Earth is given in DMI-HIRLAM-E. Using this and Equation (C.3) makes it possible to determine the orthometric height at the lowest layer of the NWP, and thus also for the rest of the layers.

C.2 Gravity

Gravity is determined using the expression for normal gravity given by Torge (1989). The error introduced by using normal gravity instead of the gravity corresponding to the total gravity potential is about 100 mGal or 0.001 m/s^2 in the Nordic area (A. Olesen, KMS - the National Survey and Cadastre - Denmark, personal communication).

An error of this size introduces an error in the zenith total tropospheric delay of about 0.2 mm or about 0.008%, and it is thus negligible.

C.3 Geoid Model

The positions of the GPS stations are given with heights above the ellipsoid, and since the pressure given in DMI-HIRLAM-E is related to heights above sea level, a geoid model was implemented in order to convert the ellipsoidal heights to orthometric heights.

A special version of the EGM96 geoid model covering the same area as the reduced DMI-HIRLAM-E was generated by A. Olesen, from KMS. The global EGM96 model is described by Lemoine et al. (1996).

The geoid height for the GPS stations is then found by bi-linear interpolation in the geoid model, using the interpolation routine described in Section C.5. The orthometric height of the GPS station is finally determined by subtracting the geoid height from the ellipsoidal height.

C.4 Conversion of Horizontal Coordinates

DMI-HIRLAM-E is generated on a rotated regular grid of spherical latitude and longitude. The GPS station coordinates are given in ellipsoidal latitude and longitude on the GRS80 reference ellipsoid, so in order to determine the zenith delay for a given GPS station, its coordinates are converted to the rotated spherical reference. The procedure is as follows:

Initially the ellipsoidal station coordinates are converted to cartesian (X,Y,Z) coordinates, and then to spherical latitude and longitude on a sphere with the same origo as the GRS80 ellipsoid, and with a radius the same size as the sphere used for developing DMI-HIRLAM-E, i.e. a radius of 6367470.0 meters.

The spherical latitude and longitude are then rotated using a subroutine provided by Martin Sørensen, DMI, and the GPS station coordinates are now in concordance with the horizontal reference used for HIRLAM.

C.5 Horizontal Interpolation

In order to determine the zenith delay, the temperature and specific humidity must be determined for the relevant ray point by horizontal interpolation in each layer of the NWP. Both the Danish Meteorological Institute and Schueler (2001) are using a bi-linear interpolation routine for this purpose, and this was also chosen as the solution for this implementation. However, a bi-cubic spline interpolation was also tested as described below.

Bi-linear interpolation

The bi-linear interpolation routine is implemented based on Press et al. (1997). The interpolation technique is based on information from the four grid points surrounding the location of the GPS station, and the interpolation is carried out using a distance dependent weighting scheme, so that values in grid points close to the station are given a higher weight.

No test has been carried out on the accuracy of the bi-linear interpolation routine, but it was compared with the interpolation routine implemented in the GEOIP programme which is a part of the GRAVSOFT package[1]. When used with the EGM96 geoid model, the two routines gave the exact same results for a number of geoid heights determined with mm resolution.

Bi-cubic spline interpolation

In connection with implementation of the ray tracer, as described in Chapter 5, a bi-cubic interpolation algorithm was tested.

Cubic spline interpolation is described by e.g. Ralston and Rabinowitz (1978), and the function was implemented based on Press et al. (1997). When using bi-cubic spline interpolation with the ray tracer for determining the delay in zenith for station ONSA and BORA with the September 5, 01:00 UTC, DMI-HIRLAM-E field, an improvement of 0.2 mm was found in the zenith delay.

The computation time is, however, increased when using the bi-cubic spline interpolation, because the spline routine reads considerably more data in the NWP files in order to determine the second derivative of the grid point values. Computation time is a big issue in connection with the ray tracer, and the bi-linear interpolation routine was therefore implemented in both the ray tracer and the zenith delay codes, and the routine was also used for interpolation in the geoid model.

Lagrange interpolation

In connection with the process of correcting the raw GPS data with the tropospheric delay, as described in Chapter 4, knowledge about the signal elevation angle is necessary in order to map the zenith delays to lower elevation angles. For this purpose precise satellite orbits were used along with a Lagrange interpolation routine for temporal interpolation of satellite positions, as recommended by Hofmann-Wellenhof et al. (1992).

---

[1] The GRAVSOFT package is a suite of programs for physical geodesy developed through the last 20 years by the Danish National Survey and Cadastre and the University of Copenhagen. The software is being used by numerous groups all over the world, and has been throughoutly tested. The GEOIP programme is a sub programme developed for interpolation in geoid models. The GRAVSOFT software package is described for instance on http://www.gfy.ku.dk/˜cct/gravsoft.txt

Langrangian interpolation is described by Ralston and Rabinowitz (1978) and the function was implemented as recommended by Press et al. (1997). Interpolation is necessary since the precise satellite positions are only given every 15 minutes, and the data rate used here was 15 seconds.

Extrapolation

DMI-HIRLAM-E NWP covers an area that is larger than what is necessary in order to model a GPS signal received down to a 5° elevation angle. Thus, horizontal extrapolation has not been an issue in connection with this work.

C.6 Vertical Interpolation

A vertical interpolation / extrapolation routine is necessary because the height of the GPS antenna (for zenith delays) or the ray point (for ray tracing) does not correspond exactly to the height of any layer in HIRLAM. The temperature for any given height is determined assuming a constant lapse rate atmosphere, where the temperature changes as a linear function of height. The same assumption is used for specific humidity.

Neither temperature nor specific humidity are linear functions of height, as the plots in Figure C.1 show. But the approximation still works well, since it is only used for interpolation between the layers in the NWP, and between each layer the two parameters can be anticipated to change linearly.

The temperature or specific humidity lapse rate, α, is determined as:

$$\alpha_T = -\frac{\delta T}{\delta h} \qquad \alpha_q = -\frac{\delta q}{\delta h} \tag{C.5}$$

δT is the change in temperature, δh is the corresponding change in height, and δq is the change in specific humidity.

The plots of P and log(P) in Figure C.1 indicate that the pressure is decreasing exponentially with height, and by still assuming a constant lapse rate atmosphere, the following exponential formula from Mendes (1999) can be used for interpolating or extrapolating the pressure.

$$P = P_0 \Big(\frac{T}{T_0}\Big)^{\frac{g}{R_d \alpha_T}} \tag{C.6}$$

P_0 is pressure for the previous layer, T is temperature for the evaluation point, R_d is the gas constant, g is gravity for the evaluation point, and α_T is the temperature lapse rate determined by Equation (C.5).

Resolution of specific humidity

The specific humidity as plotted in Figure C.1 does not show a smooth behavior, but rather a step like decrease with altitude. This is caused by the poor resolution of the specific humidity within the GRIB format, where the specific humidity is only given with one significant digit. When generating DMI-HIRLAM-E more significant digits are used, but the resolution is lost when the data is fitted into this predefined data format, which is approved by the World Meteorological Organization.

The low resolution of the water vapor measure might affect the tropospheric delays based on the NWP, but no investigation has been made of this.

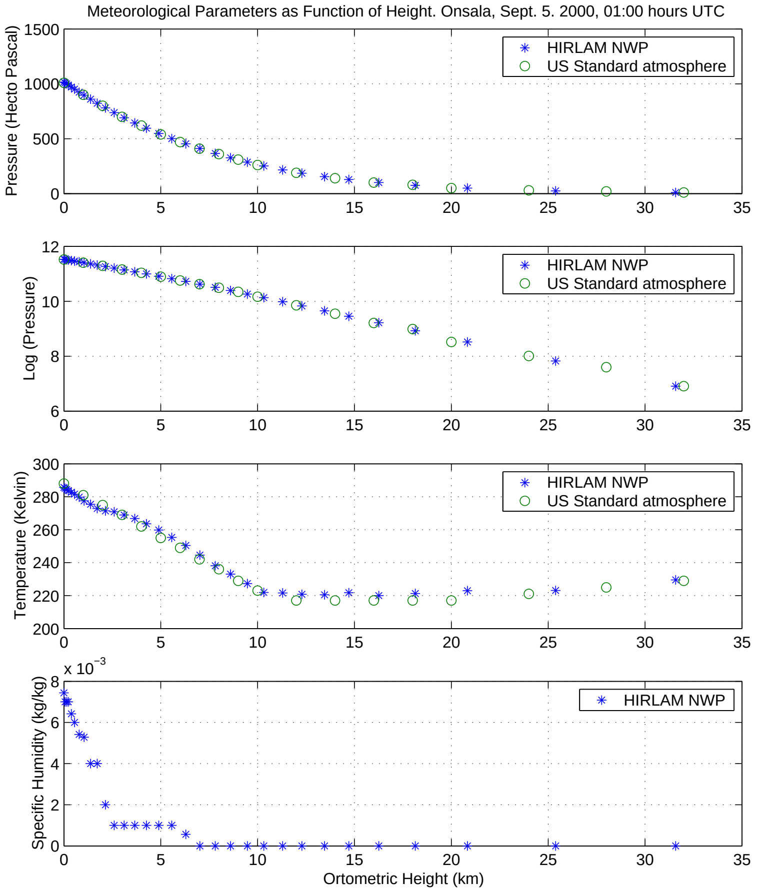

Figure C.1: Vertical profiles of pressure, log(pressure), temperature, and specific humidity for station Onsala on September 5, 2000 at 00:01 UTC. The US standard atmosphere values are added for comparison, green circles.

C.7 Contribution to Zenith Delay Above HIRLAM

The uppermost layer of DMI-HIRLAM-E has an altitude of about 30 km, but contributions to the dry or hydrostatic delay are present up to an altitude of about 80 km. The top delay is determined using the following expression given by Vedel et al. (2001):

$$Topdelay = \frac{k_1 R_d P_1}{g_1}\left[1 + 2\left(\frac{R_d T_1}{r_1 g_1}\right) + 2\left(\frac{R_d T_1}{r_1 g_1}\right)^2\right] \tag{C.7}$$

The subscript 1 indicates that the parameters are from the top layer of HIRLAM (called layer 1 within the modelling process). r_1 is the distance from the center of the Earth to the top layer.

Another solution is to use the Saastamoinen model for hydrostatic delay for determining the contribution to the delay from atop HIRLAM. This approach is for instance used by Schueler (2001). In the thesis work the approach was also tested for the GPS sites ONSA and BORA, and the Saastamoinen model showed a decrease of 0.07 mm (corresponding to 0.003%) in the hydrostatic delay when used instead of the Vedel model.

The difference between the Saastamoinen and Vedel models is thus negligible, and the Vedel approach was chosen in order to maintain consistency with the rest of the zenith delay estimation procedure.

C.8 Spherical or Ellipsoidal Pressure Shells

DMI-HIRLAM-E is generated from dynamic meteorological models based on a spherically shaped Earth with a radius of 6367470.0 meters. The different layers of the NWP are thus spherically shaped pressure shells layered above the sphere.

Since the shape of the Earth is more ellipsoidal than spherical, a test was made in which the entire NWP was converted to an ellipsoidal model, turning the pressure shells into ellipsoidally shaped shells. The GRS80 ellipsoid was used for the test, and it is characterized by (Torge, 1991):

- Semi major axis = 6378137.0 meters
- Flattening = 1 / 298.25722

Zenith delays were then determined for all 14 stations for 8 different epochs in time throughout September 5, 2000. The zenith total delay decreased with a value of 0.04–0.05 mm (or 0.002%) as compared to the situation where the spherical model was used.

Zenith delays are determined along the normal line of the sphere or the ellipsoid, and in this situation the difference between using the spherically or ellipsoidally shaped pressure shells is negligible. The shape of the shells is expected to have a more significant influence when working with signals received at lower elevation angles, as in connection with the ray tracer. This was, unfortunately, never tested in connection with the ray tracer.

C.9 Analysis Versus Prediction

For DMI-HIRLAM-E data assimilation is carried out twice daily, at midnight and noon UTC, and a new analysis field is generated based on the new data. Within the given data set, these analyses form the basis for 1 to 12 hour predictions.

The plot in Figure C.2 shows the zenith total delays at noon on September 5, 2000. The zenith delays are determined for the 14 GPS stations using both the 12-hour prediction based on the midnight data assimilation, and the new analysis based on the data assimilation at noon.

Statistics of the differences between zenith delays determined using the prediction and the analysis are given in Table C.1.

Mean	Std. Dev.	RMS
0.000 m	0.007 m	0.007 m

Table C.1: Mean, standard deviation, and RMS of the difference between DMI-HIRLAM-E analysis and 12 hour prediction at noon on September 5, 2000.

Table C.1 shows a very good quality of the 12 hour predictions, and this indicates that a 12 hour prediction can be used without problems for estimating zenith delays. This is also in line with the results given by Yang et al. (1999) as mentioned in Chapter 3.

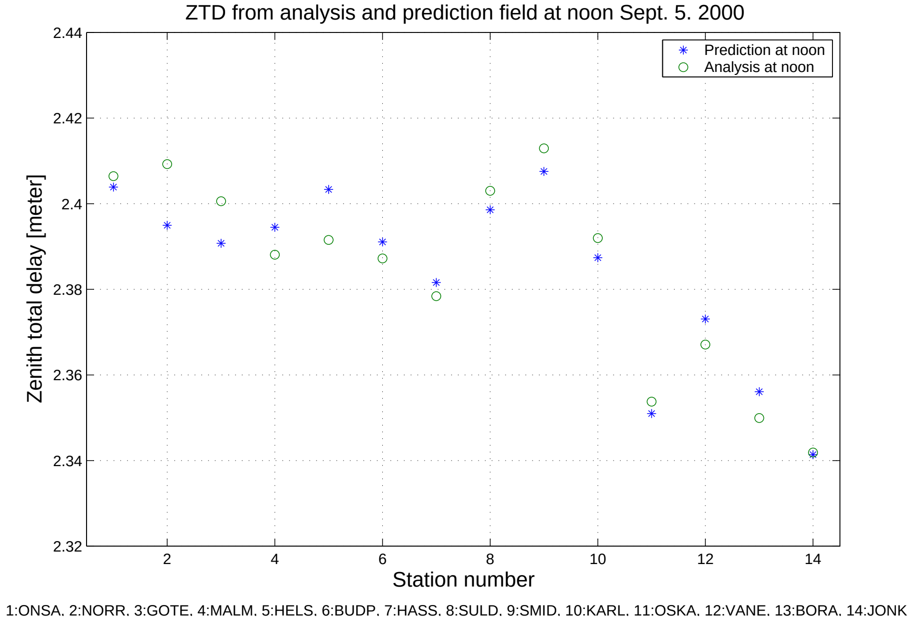

Figure C.2: Zenith total delay determined using analysis and prediction data for all stations.

.

In the plot the stations are ordered according to station height, and since the wet part of the delay is the most difficult to predict, a larger difference between prediction and analysis was expected for the stations at lower altitudes. This is not clearly seen from the plot, partly because the difference in altitude between the highest and the lowest station is relatively small, about 220 meter, and partly because the quality of the 12 hour prediction is high. For all tests described in this thesis the noon analysis has been used, rather than the 12 hour prediction.