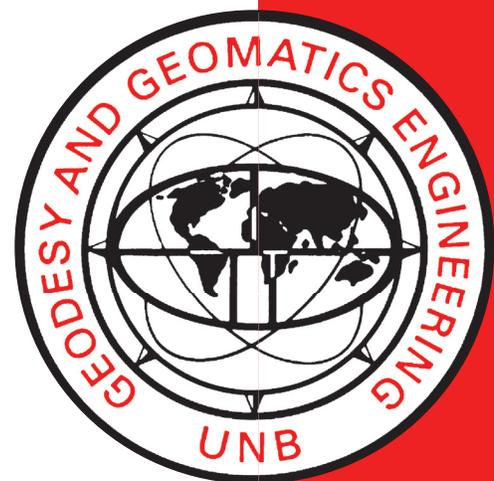


THE REDUCTION OF DIFFERENTIAL IONOSPHERIC DELAY FOR GPS CARRIER PHASE AMBIGUITY RESOLUTION

P. J. STEWART

August 1997



**TECHNICAL REPORT
NO. 185**

PREFACE

In order to make our extensive series of technical reports more readily available, we have scanned the old master copies and produced electronic versions in Portable Document Format. The quality of the images varies depending on the quality of the originals. The images have not been converted to searchable text.

**THE REDUCTION OF DIFFERENTIAL
IONOSPHERIC DELAY FOR GPS CARRIER
PHASE AMBIGUITY RESOLUTION**

Peter J. Stewart

Department of Geodesy and Geomatics Engineering
University of New Brunswick
P.O. Box 4400
Fredericton, N.B.
Canada
E3B 5A3

August 1997

© Peter J. Stewart, 1997

Abstract

To utilise the high precision of the GPS carrier phase observable it is necessary to solve for the unknown number of integer cycles between satellite and receiver. On differencing observations between pairs of satellites and receivers (the double differencing operation), any common errors will cancel. As the separation between receivers increases, so the errors induced by the charged region of the upper atmosphere (the ionosphere) on line of sight observations to the same satellite become less correlated. Thus the double differencing operation leaves a significant residual effect and ambiguity resolution may become difficult, if not impossible at station separations of greater than 15km; a value dependent on the level of solar activity. It is then perhaps apparent that modelling of this spatial change in ionospheric delay will allow for mitigation of this residual error. Such a model was produced in the course of this thesis for the real time estimation and reduction of differential ionospheric delay. Three coefficients describe a polynomial surface in latitude and longitude differences from the reference station, and the delays thus estimated are applied to carrier phase observations at the reference and rover receivers.

The success of this approach was quantified by attempting the resolution of ambiguities at successively longer baselines along the flight path used. Results show a significant increase in the distances over which ambiguities may apparently be reliably resolved. Data corrected via this model allowed the resolution of ambiguities at distances of up to 40km.

Preface

This technical report is a reproduction of a thesis submitted in partial fulfilment of the requirements for the degree of Master of Science of Engineering in the Department of Geodesy and Geomatics Engineering, February, 1997. The research was supervised by Dr. Alfred Kleusberg, and funding was provided by the Natural Sciences and Engineering Research Council of Canada.

As with any copyrighted material, permission to reprint or quote extensively from this report must be received from the author. The citation to this work should appear as follows:

Stewart, P.J. (1997). *The Reduction of Differential Ionospheric Delay for GPS Carrier Phase Ambiguity Resolution*. M.Sc.E thesis, Department of Geodesy and Geomatics Engineering Technical Report No. 185, University of New Brunswick, Fredericton, New Brunswick, Canada, 104 pp.

Acknowledgements

Thanks are due to a number of people who have provided support, be it technical, financial or personal, during the research leading to this thesis. Firstly, my supervisor, Dr. Alfred Kleusberg, who has provided a guiding influence both in the production of this thesis and in my graduate studies in general. His input and encouragement has afforded immeasurable help during my time at UNB.

Financial assistance was provided by the Natural Sciences and Engineering Research Council (NSERC). Some data used was obtained from the International GPS Service for Geodynamics (IGS). Thanks are also extended to Transport Canada and the National Research Council of Canada for the provision of kinematic GPS data.

On a more personal note, this list of those who provided support and encouragement would not be complete without mention of family and friends; my wife Melanie who has been there throughout my graduate studies as friend and confidante, my parents who have provided support and encouragement, and everyone else who has been there for me. Thanks.

Table of Contents

Abstract.....	ii
Preface	iii
Acknowledgements	iv
Table of Contents	v
List of Figures	vii
List of Tables.....	ix
1 CHAPTER ONE INTRODUCTION	1
1.1 Motivation	4
1.2 Literature Review.....	5
1.3 Thesis Outline	11
2 CHAPTER TWO BIASES AND ERRORS IN GPS	13
2.1 Troposphere.....	13
2.2 Multipath	15
2.3 Orbital Errors.....	16
2.4 Clock Errors and Selective Availability.....	16
2.5 Anti-Spoofing (AS).....	17
2.6 Summary of GPS Error Budget.....	18
3 CHAPTER THREE THE IONOSPHERE	21
3.1 Introduction	21
3.2 Structure.....	24
3.3 Temporal Variability	29
3.3.1 Diurnal Variation.....	29
3.3.2 Seasonal Variation.....	30
3.3.3 Sun Spots and the Solar Cycle.....	30
3.3.4 Atmospheric Storms.....	32
3.4 Latitudinal Variability.....	33
3.4.1 The Equatorial Ionosphere	36
3.4.2 The Polar Ionosphere	37
3.5 Summary of Ionospheric Morphology	39
3.6 Ionospheric Propagation and Refraction	40
3.6.1 Phase Refractive Index	41
3.6.2 Group Refractive Index	43
3.6.3 Ionospheric Group Delay and Phase Advance.....	44
3.7 Ionospheric Effects on GPS.....	45
3.8 The Single Layer Model of the Ionosphere	47

4 CHAPTER FOUR PROCEDURE.....	50
4.1 Mapping Functions.....	50
4.2 Hardware Differential Delays	52
4.3 Testing of VERDE using Synthesised Data	56
4.4 Construction of the Ionospheric Model.....	58
4.4.1 TECEST	58
4.4.1.1 Polynomial Surface Fitting.....	60
4.4.1.2 Kalman filtering algorithm	61
4.4.1.3 Computation of Kalman filter inputs	65
4.5 Application of the Ionospheric Model.....	70
4.5.1 ROVER	71
4.5.2 Reference Receiver Corrections in TECEST.....	74
4.6 Ambiguity Resolution.....	74
5 CHAPTER FIVE RESULTS.....	78
5.1 Multipath	78
5.2 Variability in Surface Fit Coefficients	81
5.2.1 Temporal and Spatial Changes in Estimated Ionospheric Delay	81
5.3 Preliminary Results.....	84
5.3.1 Results from Frizzle '95 data	85
5.3.1.1 Ionospheric Delay Corrections.....	85
5.3.1.2 Ambiguity Resolution Results.....	87
6 CHAPTER SIX CONCLUSIONS AND RECOMMENDATIONS	95
6.1 Summary.....	95
6.2 Conclusions.....	96
6.3 Recommendations for Future Related Research.....	97
References.....	99

List of Figures

Figure 1.1 Schematic of typical kinematic DGPS network.....	2
Figure 3.1 Recombination in the ionosphere - (a) radiative, (b) dissociative.....	21
Figure 3.2 Typical values of altitude and electron density of ionospheric layers for mid-latitude location during summer daytime.....	26
Figure 3.3 Diurnal variations in the critical frequencies of the E, F ₁ and F ₂ layers for Canberra in summer and winter and two levels of solar activity.....	28
Figure 3.4 The solar cycle.....	31
Figure 3.5 Form of the interplanetary magnetic field showing inward (-) and outward (+) sectorized structure.....	33
Figure 3.6 North-south cross section of the modified geomagnetic field showing closed and open geomagnetic field lines.....	34
Figure 3.7 Basis for the Polar Cap Absorption event.....	38
Figure 3.8 Contours of ionospheric zenith time delay in units of ns on L1 for the March equinox season for a year of low sunspot activity (sunspot number = 10).....	39
Figure 3.9 Contours of zenith ionospheric time delay in units of ns on L1 for the March equinox season for a solar maximum year (sunspot number = 153).....	40
Figure 3.10 Geometry of satellite receiver line of sight and the ionospheric shell model.....	48
Figure 4.1 Line of sight and vertical ionospheric delays with SV elevation angle.....	52
Figure 4.2 Typical latitude and longitude of Ionospheric Pierce Points (IPPs) used in surface estimation procedure. Elevation cut-off is 20 degrees, shell height is 350km. Marker shows last position of IPP.....	59
Figure 4.3 Recursive Loop of standard Kalman Filter.....	63
Figure 4.4 Constant coefficient of surface fit for day 298 of 1995 at station Albert Head. The variation of the a ₀ coefficient on application of the Kalman filter is clearly seen.....	68

Figure 4.5 Raw and filtered latitude dependent coefficient of ionospheric delay surface for day 298 of 1995 at station Albert Head.....	69
Figure 4.6 Raw and filtered longitude dependent coefficient of ionospheric delay surface fit for day 295 of 1995 at station Albert Head.....	70
Figure 4.7 Change in IPP location and elevation angle with baseline length. As the rover moves to position 2 so the distance between IPPs and the difference in observed elevation angles at reference and rover increases.....	71
Figure 4.8 Flow diagram showing the procedure followed to calculate and apply ionospheric corrections at the rover receiver.....	73
Figure 4.9 The ambiguity resolution problem.	76
Figure 5.1 Constant coefficient of ionospheric delay surface fit for three consecutive days (297 - 299) in 1995.	79
Figure 5.2 Constant coefficient of ionospheric delay surface fit for three consecutive days (297 - 299) in 1995.	81
Figure 5.3 Surface fit coefficients for day 298 of 1995 at Albert Head, B.C. Canada. The a_0 coefficient is in units of metres, a_1 and a_2 in units of metres/radian.	83
Figure 5.4 Filtered coefficients for reference station used in Frizzle '95 data collection.	86
Figure 5.5 Difference in line of sight ionospheric delay estimated with all coefficients for observations to SV#29 from reference and rover receivers. This residual delay is shown to increase with baseline length.	87
Figure 5.6 Resolved double differenced ambiguity values for SV#17.....	89
Figure 5.7 Resolved double differenced ambiguity values for SV#18.....	89
Figure 5.8 Resolved double differenced ambiguities for SV#25.....	90
Figure 5.9 Resolved double differenced ambiguity values for SV#28.....	90
Figure 5.10 Resolved double differenced ambiguity values for SV#29.....	91
Figure 5.11 Resolved double differenced ambiguity values for SV#31.....	91
Figure 5.12 R.m.s. carrier phase residuals at time of ambiguity resolution plotted against baseline length.....	93

List of Tables

Table 1	Standard error model for C/A code measurements with SA off.....	18
Table 2	Standard error model for C/A code measurements with SA on	19
Table 3	Estimated hardware delays from VERDE showing the effect of changing ionospheric shell height.	57
Table 4	List of basis functions for some bivariate polynomials.....	60

1 Chapter One

Introduction

Observations of the phase of the carrier signal emitted by a GPS satellite may be made to a fraction of a cycle. As such, this carrier phase observable provides an extremely precise method of measuring the range from that satellite to the tracking receiver. Unfortunately only the fractional part of the phase may be measured, and thus in order to utilise this potentially precise method of determining the range to that particular satellite, the unknown integer number of cycles must be solved for. This is the problem of ambiguity resolution.

If two receivers k and m are observing two satellites p and q at the same epoch [see Figure 1.1], then on differencing the respective carrier phase observations and grouping any sources of residual error together in ε for initial simplicity, the following is obtained [Leick, 1995]:

$$\begin{aligned}\phi_{km}^{pq} &= \phi_{km}^p - \phi_{km}^q \\ &= -\frac{f}{c} \left\{ \left[\rho_k^p(t) - \rho_m^p(t) \right] - \left[\rho_k^q(t) - \rho_m^q(t) \right] \right\} + N_{km}^{pq} + \varepsilon\end{aligned}\tag{1.1}$$

where

$$N_{km}^{pq} = N_{km}^p - N_{km}^q$$

and ρ_j^i is the geometric range from satellite i to receiver j . ϕ_{km}^p and ϕ_{km}^q are the single differences of phase, obtained when two receivers k and m observe the same satellite p or q at the same epoch. It is then necessary to find that integer value N_{km}^{pq} , the double differenced ambiguity, which satisfies equation (1.1) for each double difference of phase formed.

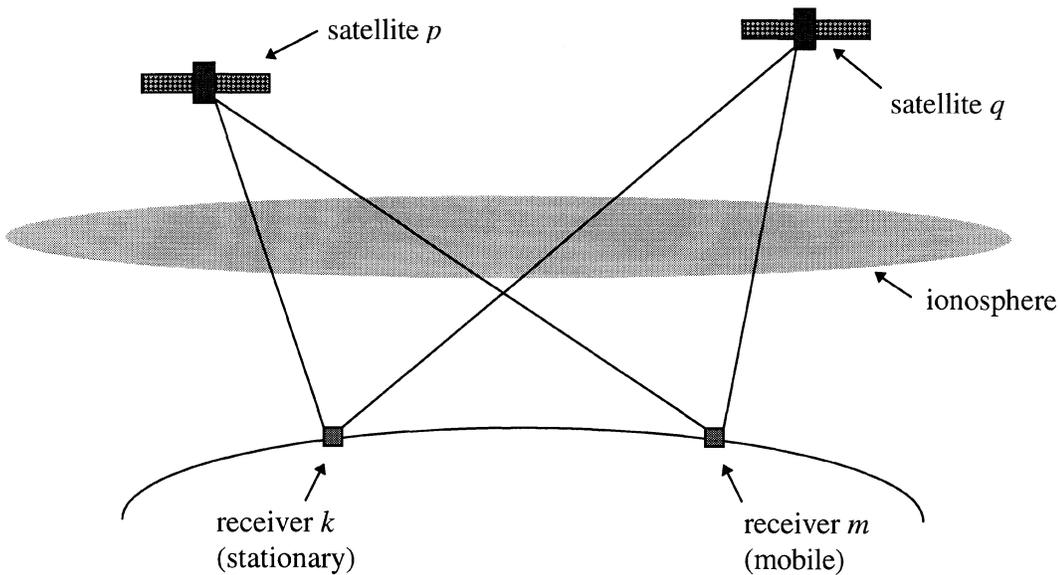


Figure 1.1 Schematic of typical kinematic DGPS network.

The value ϵ in equation (1.1) was defined to contain residual errors not accounted for in the differencing process. Thus these are the portions of any error not common between observations. If the atmosphere is considered homogeneous over the area in question and the distance between receivers is small, then the induced effect on observations will be the same for each satellite receiver pair and will cancel in the differencing operation outlined above.

As the baseline length increases, not only will the signal path for each satellite pass through a more distant part of the atmosphere, but the observed satellite elevation angle will also become increasingly different. Thus the range of baselines over which such ambiguities may be resolved is limited by the homogeneity of atmospheric conditions along lines of sight from each receiver to a particular satellite and the baseline length between receivers.

Since the atmosphere is not a homogeneous medium, then the worth of such an assumption will typically degrade with distance from the base station receiver. In particular, the charged region of the upper atmosphere (the ionosphere) is of note in this respect. The typical range of ionospheric corrections for the FAA Wide Area Augmentation System (WAAS) was computed by *Klobuchar et al.* [1995]. The results in this paper show that if a user is located approximately 800km from a base station the difference on slant ionospheric delays is about 2m [Klobuchar et al., 1995]. These figures were computed using data from a time of low solar activity, and should be scaled significantly to approximate similar errors at solar maximum. Nonetheless, significant differential ionospheric delays may be observed due to magnetic disturbances, events which cannot be predicted successfully at the present time. Therefore using a model produced from real-time data rather than based on some prediction of ionospheric delay should allow for at least partial mitigation of such spatial variations.

1.1 Motivation

Abidin [1993] states that the “...difficulties [in ambiguity resolution] arise primarily from the ephemeris errors and residual atmospheric delays...”. Therefore better modelling of these atmospheric delays should enable more reliable ambiguity resolution.

The differencing operation between satellites and receivers results in the elimination of a number of common errors, including satellite and receiver clocks, as mentioned above. If the ionosphere is considered a spatially homogeneous medium and the distance between receivers is short, then the ionospheric delay will also cancel since the area of ionosphere traversed and the satellite elevation angles will be almost equal. Unfortunately, this homogeneity is not the case, and for a reliable solution to the ambiguity resolution problem, any spatial change in ionospheric delay must be modelled. This research concentrates on estimating and removing this residual ionospheric delay for improved ambiguity resolution.

The possibility of using a real time, code based model of ionospheric delay to correct carrier phase data at both the base and rover was therefore investigated. The work carried out here provides a basis for this technique, estimating ionospheric delay, applying the ionospheric delay estimates to phase data, using post processing techniques to simulate real time data correction. Thus only data collected before the current epoch is used in the estimation of the ionospheric delay values via a Kalman Filter.

1.2 Literature Review

Modelling and mitigation of the ionospheric delay imparted on GPS observations has been attempted by many groups and individuals. A summary of some of the most recent work is provided here, with an emphasis placed on techniques used over longer baselines or areas of coverage.

Investigations by the author of the software package used for the evaluation of the influence of the ionospheric corrections showed some interesting results in the initialisation of integer ambiguities over successively longer baselines [Mader, 1995]. Mitigation of differential ionospheric effects may be attempted only if the integer ambiguities are already known from a separate source. This therefore means that an additional ambiguity resolution estimation stage must be attempted, implemented by computing ionospheric corrections based on a number of candidate sets of integers. The ambiguity function technique [Counselman and Gourevitch, 1981] provides a method of obtaining integer ambiguities independent of the phase history of the data. Using a set of dual frequency data collected at a reference station and on board an aircraft, the resolution of ambiguities over successively longer baselines was attempted. Resolution out to a baseline length of 250km was apparently possible [Mader, 1992].

Dedes and Mallett [1995] describe a system in which on-the-fly (OTF) ambiguity resolution and the estimation of ionospheric effects are undertaken simultaneously.

Modelling of the ionospheric delay is undertaken as a first order Gauss-Markov process. Preliminary results have shown that the pseudorange multipath dominates the estimation of the ionospheric delay. Thus they suggest that further work will require the proper modelling of multipath effects.

Wanninger [1995] proposes a system in which three or more reference stations are used to provide a fixed ambiguity solution and thus carrier phase based ionospheric delay corrections. This modelling of the differential ionosphere allowed the application of fast and on-the-fly (OTF) ambiguity resolution techniques to baselines of the order of 10-50 km.

A similar scheme is outlined by *St-Pierre and Santerre* [1996] where two monitor stations are used. Ambiguities are initially fixed using double differenced observations between these two receivers. This then allows the calculation of a carrier phase based local model of ionospheric delay for the improvement of ambiguity resolution.

One current topic of much interest in the GPS research field is that of the Wide Area Augmentation System (WAAS), a Wide Area DGPS (WADGPS) system initiated by the Federal Aviation Administration (FAA). The intention of the WAAS is to provide corrections to single frequency GPS on a continental scale such that particular accuracy criteria are met. Such a large undertaking has necessarily produced a significant amount of research into the spatial and temporal decorrelation of ionospheric delay values.

To correct for ionospheric delay, a network of nodes on a grid have been established, and the ionospheric delay at each of these points will be available. It is then the users task to take the delay values at the nearest grid points and interpolate the delay for each particular satellite observed. *Chao et al.* [1996] have used both an empirical and a theoretical decorrelation function to describe spatial variation in ionospheric delay. It is suggested that the spatial change will increase in magnitude as we move towards the next solar maximum. Previous research has seen the calibration of WAAS ionospheric delay estimates by producing a set of interfrequency bias values [Chao et al., 1995]. It was pointed out that the broadcast WAAS ionospheric corrections would otherwise be inherently biased by satellite and reference station hardware delays. This is an important point not only for the WAAS but also in any attempt to model the ionosphere using GPS observations.

Chavin [1996] has developed a number of theoretical ionospheric situations in which magnetic storms are affecting the ionosphere. The goal was then to assess their affect on the WAAS system. Such magnetic storms are difficult to predict, and the specifications of the WAAS must take this into account.

Doherty et al. [1994] found that the Contiguous United States (CONUS) area is free of significant short term changes in the ionosphere. These findings were produced during work looking into the required rate of update of ionospheric estimates for WAAS. Diurnal

changes were removed by passing the delay estimates through a high pass filter with a cut-off value of 15 minutes. The levelled carrier phase technique was investigated. Levelled carrier phase refers to a technique whereby the absolute value of ionospheric delay is obtained from P-code pseudorange observations, and relative changes from carrier phase measurement, reducing the relative errors associated with P-code observations.

The initial investigations done by Wilcox Electric into the WAAS ionospheric correction technique used an adaptive Kalman filter to update the ionospheric delay estimates at the grid points described earlier [Stull and Van Dierendonck, 1995]. A measure of the potential user error due to ionospheric decorrelation was also produced. A 30 second filter update rate was used since even in the more dynamic polar and equatorial regions the rate of change of ionospheric delay over this period is unlikely to exceed the state error values [Bishop and Klobuchar, 1990].

Wu et al. [1996] set out to provide real time ionospheric corrections, their intention being to reduce the User Equivalent Range Error (UERE) by 30%. A reduction of the ionospheric range error to less than 0.5m is claimed.

Research at the NASA Jet Propulsion Laboratory (JPL) has gravitated toward the use of levelled carrier phase for the global modelling of the ionosphere. *Mannucci et al.* [1995] produced a triangular grid of ionospheric delay estimates in a sun fixed coordinate system, proposing an application of this technique to the WAAS system rather than the

rectangular earth fixed network currently in preparation. The proposed method here was based on the world wide GYPSY/TRIN method developed at JPL [Mannucci et al., 1993] and was shown to be capable of accuracies at the 0.25 to 0.5 metre level, when compared to the input test data set computed from the Bent model of the ionosphere.

Fraile-Ordóñez [1995] also used a combination of carrier phase and code observations to provide an estimate of ionospheric delay, using a least squares based algorithm. Good results are claimed for the application of this method to the computation of corrections for a WADGPS reference station. The improvement on utilising the inherently lower noise carrier phase data is clearly stated. One major disadvantage of this levelled carrier phase approach recognised here is the requirement to reinitialise the ionospheric delay estimate after each cycle slip.

Daniel et al. [1996a, 1996b] have taken the approach of updating the current broadcast single frequency model developed by *Klobuchar* [1987]. The intention is to provide a more robust set of parameters to describe ionospheric delay. Specifically, the concerns over the current Broadcast Model are its inability to represent the well known equatorial anomaly, and the inadequacy of the mapping function used to convert vertical delays from the model to line of sight values which may be applied by the single frequency user.

Håkegård [1994] produced a real time ionospheric model for use in a European wide area DGPS system. Cross correlation of the incoming encrypted Y-codes is used to determine Total Electron Content (TEC) measurements. Four reference stations around Europe were used. One minute averaged values of TEC are used, and a 15 minute model update value proposed. Data from ionosondes was also available. Distances between reference stations were up to 4000 km, beyond the normal ionospheric correlation distances. The technique used was to provide updates to a chosen ionospheric model. A grid network of TEC values is then available to DGPS users, similar to the approach taken in the WAAS system.

The model provided good estimates of the TEC values when compared to those measured directly by a test receiver, with the model estimated TEC within 1σ value of the measurement for much of the 24 hour period. It was noted that no major ionospheric disturbances occurred during the test period, and the underlying model without real time updates also represented the measured TEC well.

Webster [1993] created a regional ionospheric model for use in the correction of single frequency GPS data. Carrier phase observations were used to create the model and a TEC surface was created. Corrections are provided for satellite-receiver double differences of phase and are applied as an input to a positioning software package in a post-processed mode.

It then appears from the current state of the art in carrier phase ambiguity resolution that a major limiting factor is the differential ionosphere. That is, the spatial change in the ionospheric delay is significant enough that ambiguity resolution becomes difficult, if not impossible beyond a certain baseline threshold, dependent on ionospheric conditions at that time. Indeed, a study of ambiguity resolution techniques in precise aircraft landing situations intimated that some modelling of ionospheric delay should be undertaken should the baseline increase beyond 15km [Blomenhofer et al., 1993].

1.3 Thesis Outline

Chapter Two provides a summary of the important error sources and biases which affect GPS.

Chapter Three gives an introduction to the structure and variability of the ionosphere. Variation over both time and space is outlined, setting the scene for the model produced during this research. The effect of the ionosphere on electromagnetic wave propagation is summarised, with a derivation of equations to evaluate ionospheric delay given.

Chapter Four is an outline of the procedure followed for the purposes of this thesis. Documentation of the software developed here is given with development of the equations used. An introduction to the theories of surface approximation and Kalman filtering is

provided, along with an outline of how these were utilised for the modelling of the ionosphere.

Chapter Five gives a description of the results obtained at various stages of this work. Analysis of the influence of the surface estimation approach taken is provided, along with quantification of the success of the ionospheric corrections so produced.

Chapter Six provides a summary of the procedures and results produced here, drawing a number of conclusions and outlining recommended improvements to the technique.

2 Chapter Two

Biases and Errors in GPS

Aside from the effects of the ionosphere, to be discussed in detail in the following chapter, other significant error sources are present in GPS. The most important of these are outlined below along with methods used to overcome them.

2.1 Troposphere

The lowest region of the atmosphere, extending to a maximum of about 16km above the earth's surface, is known as the troposphere. The neutral atmosphere may be considered to include the stratosphere, extending the altitude of the area considered to approximately 50km. However, since the majority of the neutral atoms and molecules occur below 16km it is usual to refer to the delay induced as the tropospheric delay.

This area of the atmosphere is electrically neutral and is non-dispersive at GPS frequencies. Thus the delay induced is equal for both L1 and L2 and will cancel on differencing the two. This phenomenon is beneficial in dual frequency observations, but counter productive if modelling of the tropospheric effect on GPS is to be done. The effect on code and carrier is also equal.

Vertical tropospheric delay results in an apparent range increase of 2.4m, but can reach 9.3m at a 15 ° elevation angle [Brunner and Welsch, 1993]. The tropospheric delay may be split into a wet and dry component since the refractivity of a parcel of air, N , can be defined as the sum of two terms:

$$N = N_A + N_B \quad (2.1)$$

where N_A is proportional to the total density of air and N_B is mainly determined by the total density of the water vapour contained in the air parcel [Brunner and Welsch, 1993]. The dry component accounts for around 90% of the tropospheric delay and is more easily modelled than the remaining wet component. Significant changes in water vapour content may occur over tens of km and hours of time [Spilker, 1996]. Fortunately, the component of the total tropospheric delay caused by water vapour content is considerably smaller than that caused by the dry atmosphere.

Tropospheric error is much more prevalent in measurements taken to satellites at low elevation angles. It is at such points that the modelling of the tropospheric delay becomes more difficult.

Modelling of the tropospheric delay at GPS frequencies has been undertaken by a number of groups. *Janes et al.* [1991] provide a comparison of various models used or proposed for use in GPS. Accuracies at the centimetre level when compared to ray tracing “ground truth” values were shown for the estimation of the dry delay.

2.2 Multipath

Multipath may be described as “...the phenomenon whereby a signal arrives at a receiver via multiple paths attributable to reflection and diffraction” [Braasch, 1996]. On arrival at the receiver the reflected and diffracted signal will interfere with the direct signal. Multipath effects are caused at both the space and control segments of the GPS due to reflection from surrounding surfaces. The ‘multipath signature’ of a particular satellite is noticeable since the pattern of reflection will repeat each sidereal day as the satellite traces the same path across the sky, and thus reflects from the same surfaces, for a stationary receiver at the same sight. The multipath effect is proportional to the frequency of the signal, and thus the effect on the P-code modulation is 120 times that of the L1 carrier [Klobuchar, 1996]. In addition, multipath effects are generally stronger for low elevation satellites, and this should be taken into account in any processing of GPS observations.

Reference stations such as that used in the approach outlined in this thesis should be carefully placed to avoid nearby reflecting surfaces. Errors of 15m or more may be induced in a static receiver in extreme cases [Parkinson, 1996].

2.3 Orbital Errors

Any error in the ephemeris broadcast by a particular satellite will cause a ranging error between that satellite and receiver. Ephemeris errors tend to grow with time from the last control segment upload since the ephemeris parameters reflect a prediction in time.

On dissecting the error in the broadcast ephemeris it is usual that the tangential and cross track errors are an order of magnitude larger than those induced in the radial component [Parkinson, 1996]. Fortunately the induced ranging error is governed more by radial error than the larger components.

Precise orbital parameters are available for post processing of GPS data, allowing the minimisation of ephemeris related errors. Since the synthesis of a real time approach was to be attempted, this approach was not taken here.

2.4 Clock Errors and Selective Availability

The fractional stability of the clocks in a GPS satellite is of fundamental importance in the one way ranging of the system. Disregarding for a moment the effects of Selective Availability, the caesium and rubidium oscillators in a satellite clock are reported as having stabilities of the order of 1 part in 10^{13} over the course of a day [Parkinson, 1996].

Selective Availability (SA) refers to an intentional degradation of the measurement accuracy possible with GPS. This is carried out by the United States Department of Defense (DoD), justified in the interests of national security. Since the theoretical accuracy of C/A code measurements was found to be much higher than initially expected, then some deterioration of the system was deemed necessary. SA is implemented as the dual process of dithering of the satellite clock, designed to be unpredictable over periods longer than approximately 10 minutes (the δ process) [Spilker, 1996], and manipulation of the broadcast ephemeris data (the ϵ process) [van Graas and Braasch, 1996]. Since the satellite clock is affected, then changes to both the C/A and P codes, and to the carrier measurements will be noticed. Selective Availability has been in permanent usage since November 15th 1991 after a short period during the Gulf War in which it had been turned off, and is currently the largest single error source in GPS positioning. GPS accuracy specifications with SA on are given as 100m (2drms) in horizontal position and 140m in vertical position. [van Graas and Braasch, 1996].

2.5 Anti-Spoofing (AS)

Anti-Spoofing refers to an encryption of the P-code on L1 and L2 such that only authorised users are able to track the code. Many receivers are able to circumvent this encryption to some extent by using the C/A carrier phase loop to guide the other observables tracking loops [Gourevitch and Nolan, 1993]. Indeed, the Ashtech Z-XII

receivers used to obtain much of the data used in this work use a process of Z-tracking to provide P-code pseudoranges.

2.6 Summary of GPS Error Budget

If SA were to be turned off, the dominant error source for GPS is usually the ionosphere. The following tables provide an outline of the error characteristics of C/A code measurements of GPS both with and without the effects of Selective Availability.

Error source	One-sigma error, m		
	Bias	Random	Total
Ephemeris data	2.1	0.0	2.1
Satellite clock	2.0	0.7	2.1
Ionosphere	4.0	0.5	4.0
Troposphere	0.5	0.5	0.7
Multipath	1.0	1.0	1.4
Receiver Measurement	<u>0.5</u>	<u>0.2</u>	<u>0.5</u>
User Equivalent Range Error (UERE), rms.	5.1	1.4	5.3
Filtered UERE	5.1	0.4	5.1
Vertical one-sigma errors - VDOP = 2.5	12.8		
Horizontal one-sigma errors - HDOP = 2.0	10.2		

Table 1 Standard error model for C/A code measurements with SA off [from Parkinson, 1996, p. 481]

Error source	One-sigma error, m		
	Bias	Random	Total
Ephemeris data	2.1	0.0	2.1
Satellite clock	20.0	0.7	2.1
Ionosphere	4.0	0.5	4.0
Troposphere	0.5	0.5	0.7
Multipath	1.0	1.0	1.4
Receiver Measurement	<u>0.5</u>	<u>0.2</u>	<u>0.5</u>
User Equivalent Range Error (UERE), rms.	20.5	1.4	20.6
Filtered UERE	20.5	0.4	20.5
Vertical one-sigma errors - VDOP = 2.5	51.4		
Horizontal one-sigma errors - HDOP = 2.0	41.1		

Table 2 Standard error model for C/A code measurements with SA on [from Parkinson, 1996, p. 482].

It can be seen that the influence of Selective Availability imparts an approximately four-fold increase in horizontal and vertical position error. A similar effect will be noticeable in the P-code pseudorange observations used in this thesis.

3 Chapter Three

The Ionosphere

The previous chapter has outlined a number of potential error sources in GPS observations. The intention here is to provide an introduction to the ionosphere and its effect on such measurements.

3.1 Introduction

This area of charged particles surrounding the Earth is a direct consequence of the interaction of solar processes with the upper atmosphere. A parcel of high energy Extreme Ultra Violet (EUV) light ($\lambda = 17-175\text{nm}$), known as a photon, incident on a molecule or atom energises an outer electron of that particle, giving it enough energy to escape the attraction of the positively charged nucleus. Thus, a positively charged ion is created, and the negatively charged electron becomes free.

Conversely, ions and electrons will tend to be lost by the process of recombination. Two types of recombination take place, although their relative efficiencies are widely different [Hargreaves, 1992].

Figure 3.1a shows relatively slow radiative recombination, where energy is emitted as the electron directly recombines with a positive ion. The process of dissociative recombination, however is 10^5 times faster, despite the two stage process shown in figure 3.1b.

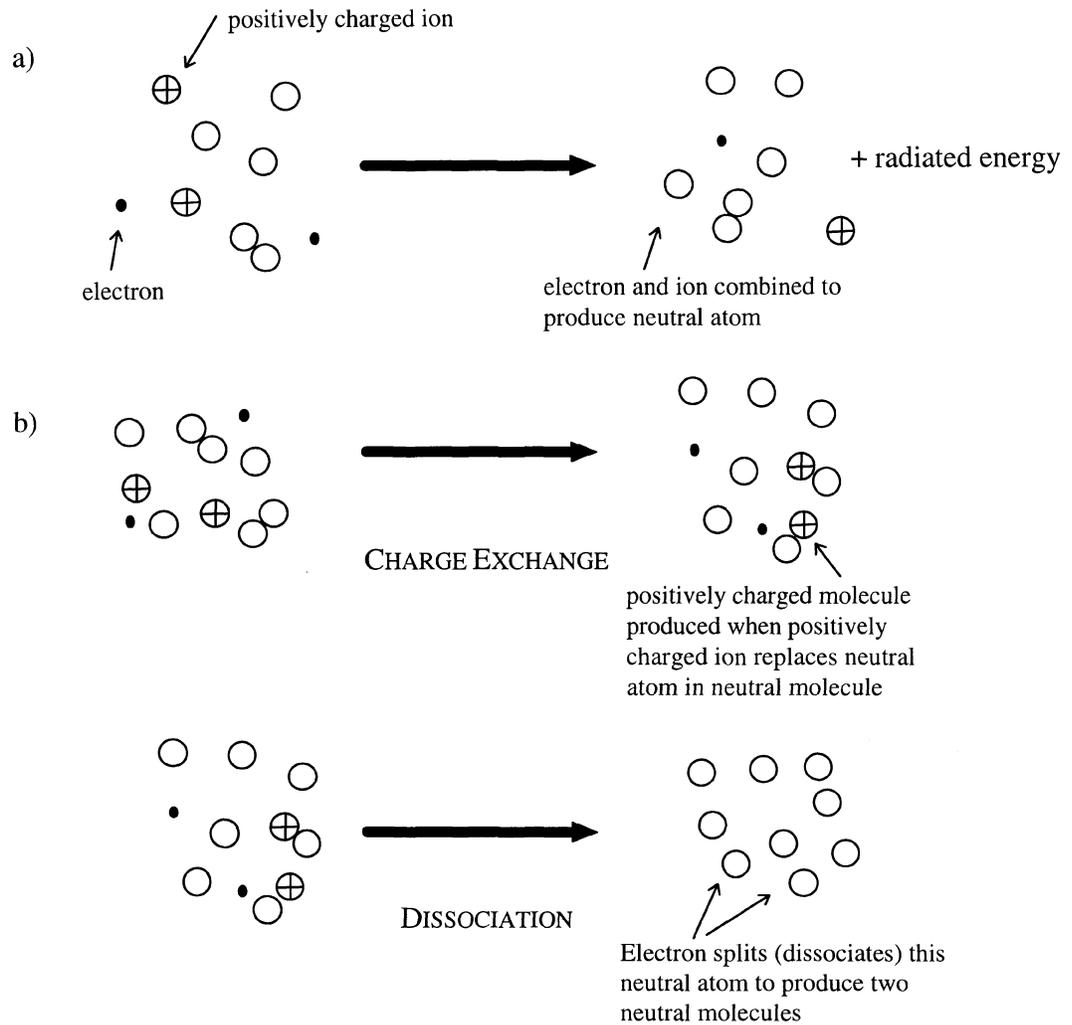


Figure 3.1 Recombination in the ionosphere - (a) radiative, (b) dissociative [from McNamara, 1991, p. 19].

Electrons may also be lost by attachment to neutral atoms, the loss of electrons being more important than the gain of a negatively charged ion for microwave propagation,

since the relatively heavy ion has little influence in this respect. The ionization of the upper atmosphere is therefore a function of the relative rate of production and loss of electrons.

To quantify the process of ionospheric production by ionizing radiation, the Chapman production function may be applied. This leads to the theory of Chapman layers, a model for the ionosphere. In order for the function to be applied, certain assumptions must be made regarding the incident radiation and atmospheric content. These are [Hargreaves, 1992] :

- A single species of atom or molecule populates the entire atmosphere, the density of which decreases exponentially with height.
- The atmosphere is plane stratified.
- Absorption of solar radiation is proportional to concentration of gas particles.
- The coefficient of absorption is constant, i.e. the radiation is monochromatic.

Although these assumptions tend to be far from the real world situation, the Chapman function still serves as a good approximation for the ionization process. Thus the function may be used as a tool in the modelling of the ionosphere [Bassiri and Hajj, 1994].

The rate of production, q , of ion-electron pairs is then given as;

$$q = \eta \sigma n I \quad (3.1)$$

where I is the intensity of ionizing radiation, n is the number of particles capable of being ionized by that radiation, σ is the absorption cross section (the amount of radiation absorbed by an atom or molecule), and η is the ionization efficiency.

The Chapman production function is written [Hargreaves, 1992]:

$$q = q_{m0} \exp(1 - z - \sec \chi \cdot e^{-z}) \quad (3.2)$$

where z is the reduced height for the neutral gas, $z = (h - h_{m0}) / H$, where H is the scale height, kT/mg . Here k is Boltzmann's constant, T is absolute temperature, m is the mass of the atom or molecule type present and g is gravity. χ is the solar zenith angle, h_{m0} is the height of maximum production rate with the Sun overhead and q_{m0} is the production rate at h_{m0} .

The level of ionization will thus vary with solar zenith angle and optical depth, defined as the absorption coefficient multiplied by the number of absorbing atoms in a unit column down to the level being considered. The production rate is defined as a maximum when the optical depth is unity [Hargreaves, 1992].

The relative importance of recombination and attachment in a Chapman layer are signified by the Greek letters α and β respectively, e.g. an α -Chapman layer is one in which recombination is the dominant effect in the loss of electrons, which tends to be the case at lower levels of the ionosphere [Hargreaves, 1992].

Ionization is also caused by other solar emissions. Energetic protons, associated with the largest of solar flares, are capable of producing an ionizing effect, most notably at high latitudes. The arrival time of the stream of protons is a function of their velocity, taking

from ten minutes to a few days to arrive in the vicinity of the Earth. [McNamara, 1991]. Their interaction with the earth's geomagnetic field determines their effect on the ionosphere.

X-rays associated with the solar flare production can penetrate far into the atmosphere, perhaps as low as 30-40 km [Hargreaves, 1992]. Ionization occurs at the level of the atmosphere at which the x-rays are stopped, by the twin processes of photoelectron absorption, where the incident energy is totally absorbed by one or more electrons, enabling escape, and by the Compton effect, whereby the incident photon is absorbed by an outer electron of a particle, and a photon of longer wavelength produced [Hargreaves, 1992]. The electron density of the lower ionosphere may be increased by a factor of 10 following a large flare.

3.2 Structure

The Earth's atmosphere is governed by differing forces at different altitudes. If the atmosphere is classified according to the state of molecular mixing, then it may be split into two major components. The lower reaches of the atmosphere (below about 100km) are well mixed by turbulence, and this region may be termed the homosphere. Above this altitude however, temperature is increased by solar heating, and the positive temperature gradient means that molecules and atoms of different molecular weight tend to separate by

gravity. In this region, the rate of molecular diffusion tends to be greater than the mixing effect of turbulence, and hence atmospheric content varies with altitude.

This process may be qualified by examining the following equation [Hargreaves, 1992];

$$D = kT/mv \quad (3.3)$$

where k is Boltzmann's constant, T is the absolute temperature, m is the mass of the atom being considered, and v is the collision frequency; essentially the number of collisions a particle makes with others in one second. D is the diffusion coefficient, a measure of the diffusion taking place, which obviously increases with temperature, and decreases with atomic mass. The collision frequency will also tend to decrease with height as the atmosphere becomes more and more rarefied.

Thus particles of different mass tend to separate to different altitudes, producing the so-called heterosphere. The processes of ionization outlined above therefore interact with these layers, and a stratified ionosphere is produced. Since the density of particles decreases with altitude, whilst the incident solar radiation decreases as it propagates through the atmosphere, then it may be apparent that the level of ionization will pass through a maximum at some point. This is an important point with relation to the thin shell model introduced later in the thesis.

The ionosphere has been classified into a number of layers according to the molecular structure of each layer. Different species will have differing values of ionization potential and absorption at different wavelengths of EUV. Then applying the Chapman function

from above, the production of free electrons is a function of the response of the local molecular make up to the incident radiation.

The ionosphere may be split into a maximum of four layers, dependent on the time of day and other solar dependent factors, as will be seen in section 3.3. These are denoted D, E, F₁ and F₂, and their relative altitudes and electron densities are shown in Figure 3.2. The D, E and F₁ layers may all be explained to some extent by the Chapman production function. The existence of the F₂ region however depends on diffusion, which this theory omits [Hargreaves, 1992].

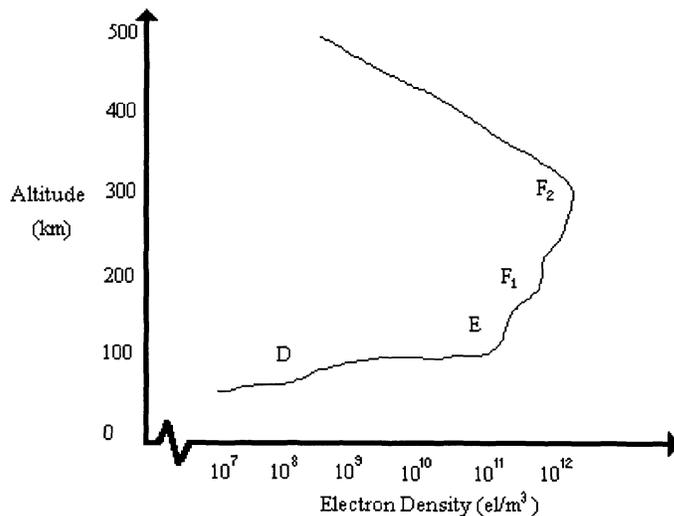


Figure 3.2 Typical values of altitude and electron density of ionospheric layers for mid-latitude location during summer daytime [from Webster, 1993, p. 25].

An important parameter in terms of both ground based and satellite communications is the critical frequency of each layer, denoted, for example: f_0F_1 . This may be explained in

terms of the plasma frequency [Hargreaves, 1992]. An approximate empirical formula for this plasma frequency is given as:

$$f_N = (80.5 N)^{1/2} \quad (3.4)$$

where N is the number of electrons per m^3 , and f_N , the plasma frequency, is in Hertz. The plasma frequency thus defined is the natural oscillation frequency of a specific plasma, and is the frequency a signal must exceed in order to propagate through the medium. Therefore the term critical frequency is used.

Values for the critical frequency vary according to time and layer, typical values for f_oF_2 being between 4.8 MHz and 11.2 MHz, dependent on season and solar activity [see section 3.3][McNamara, 1991]. Obviously these values of f_N are much smaller than the frequency of the GPS carrier, but the effect is still felt, and this will be further discussed in section 3.7.

Empirical formulae for the values of f_oE and f_oF_1 exist in terms of the solar zenith angle χ and the sunspot number R [defined below];

$$f_oE = 0.9[(180 + 1.44R)\cos \chi]^{1/4} \text{ MHz},$$

$$f_oF_1 = (4.3 + 0.01R)\cos^{0.2}\chi \text{ MHz}$$

which are reasonable approximations to the truth [McNamara, 1991]. It may be seen from the above that different layers of the ionosphere will have differing responses to ambient conditions and hence different critical frequencies. Typical values of the critical frequencies of the E, F₁ and F₂ layers are shown in Figure 3.3. A definite diurnal variation is noticeable, this being the dominant variation in ionospheric effect, an important point in

the initial validation of the model produced here. The following sections serve to explain some of these variations in critical frequency.

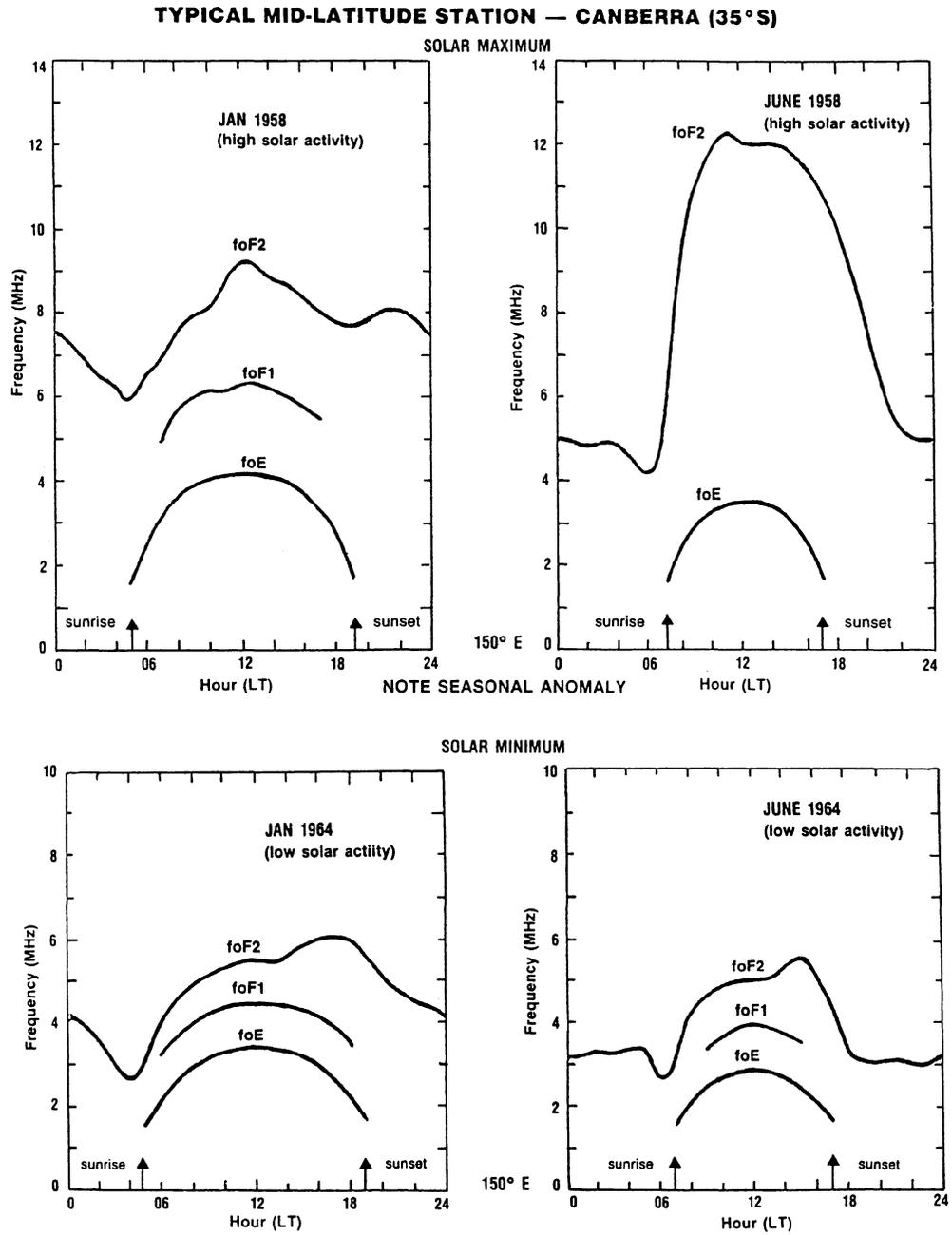


Figure 3.3 Diurnal variations in the critical frequencies of the E, F₁ and F₂ layers for Canberra in summer and winter and two levels of solar activity [from McNamara, 1991, p. 25].

3.3 Temporal Variability

The temporal variability of the ionosphere may be thought of on a variety of different time scales, all of these variations being direct or indirect functions of the influence of the Sun.

3.3.1 Diurnal Variation

On a day to day basis, the ionospheric layers encountered in the previous section tend to come and go with earth rotation i.e. they are related to whether an area is on the day or night side of the Earth. Maximum ionization will occur at approximately 2 pm local time as the influence of incident solar radiation becomes a maximum.

At night, recombination is the dominant process since photoionization cannot take place. Therefore the D, E and F₁ layers tend to largely disappear. The existence of the F₂ layer is not so closely tied to incident radiation and therefore is present at night, although with lower density and maximum density at a greater height. The explanation of this phenomenon is best left to texts such as *Hargreaves* [1992]. Night time electron content also gains an important contribution from the upper level ionization of the protonosphere, an area extending out to the altitude of the GPS satellite itself. The protonosphere is a region of mainly ionized hydrogen, and may contribute up to 50% of the electron content during the night-time hours [Klobuchar, 1996].

3.3.2 Seasonal Variation

There is also seasonal variation of relative electron densities, a general statement being that the ionization is greatest in summer. This is in part a consequence of variations in atmospheric chemistry over the seasons [Hargreaves, 1992]. The solar zenith angle also has a seasonal as well as diurnal variation - in winter the zenith angle at noon is always greater than the corresponding angle in summer. Thus the critical frequency for the D, E and F₁ layers is greatest in summer, although this is not the case for the F₂ region at mid latitudes. This anomaly is caused by seasonal changes in the atmospheric chemical make-up [McNamara, 1991].

3.3.3 Sun Spots and the Solar Cycle

Emissions from the Sun do not tend to vary by more than 1%, but even this seemingly small variation is extremely important in terms of its effect on the ionosphere [Hargreaves, 1992].

Sunspots appear as dark areas in a white light image of the Sun due to their relatively low temperature when compared to the rest of the photosphere. The Wolf Sunspot Number, R , defined as

$$R = k(f + 10g) \quad (3.5)$$

where f is the total number of spots seen, g is the number of disturbed regions (either single spots or groups) and k is a constant for the particular observatory, a function of the sensitivity of the instrument, has been used since 1848 as the predominate method of

quantifying solar activity [Hargreaves, 1992]. By studying the temporal variations of data such as the observed sunspots, some definite trends may be noticed. Figure 3.4 is a representation of the sunspot number since 1610.

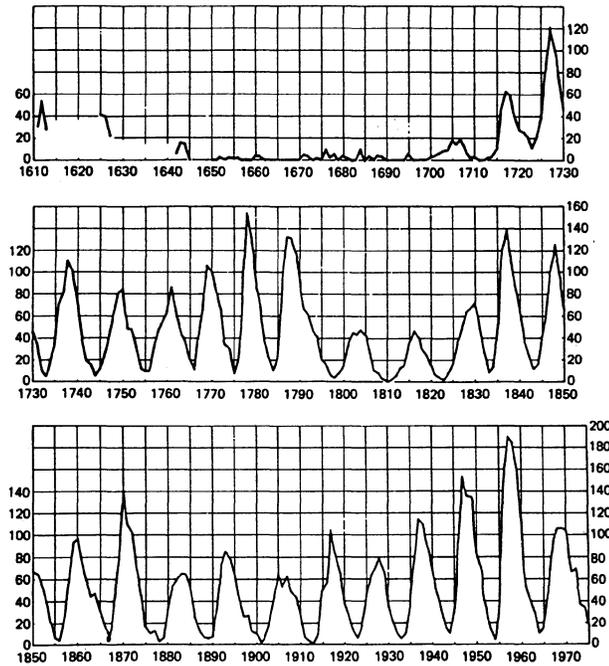


Figure 3.4 The solar cycle
[from McNamara, 1991, p. 11].

From this figure it can be seen that there exists an approximate repetition every 11 years. This phenomenon is known as the solar cycle, and has been shown to have a definite correlation with errors in trans-ionospheric signal propagation and consequently on satellite based positioning systems [Tscherning and Goad, 1985; Weigel, 1989].

Analysis of observations collected since 1729 has shown an average period for this cycle of 11.1 years. Significant deviations from this value do occur however, the shortest

cycle being of 7 years duration, the longest 17 years [Friedman, 1985]. Consequently, the prediction of solar activity is not as easy as may have been at first thought. It must also be remembered that in terms of geological time the period for which data is available is rather short, and this 11 year cycle may be a relatively short term phenomenon. The solar cycle is still however a useful parameter in ionospheric delay prediction.

3.3.4 Atmospheric Storms

Storms in the upper atmosphere lead to temporal variation in the ionospheric content, and hence in the critical frequency and refractive index of the plasma. The cause of such storms is not fully understood, but some contribution from the constant solar wind is thought [Hargreaves, 1992]. If the electron content along a signal path changes rapidly enough it may become difficult or impossible to track trans-ionospheric signals such as those from the GPS satellites [Langley, 1996].

The F-region ionospheric storm is of importance here, since the layer of maximum critical frequency occurs in the F₂ layer. The storm is a three stage process, with an initial increase in ionization lasting a few hours, a subsequent decrease and a long (one to several days) recovery period [McNamara, 1991].

3.4 Latitudinal Variability

The Interplanetary Magnetic Field (IMF) is a magnetic field frozen into the solar wind; a constant outflow of material from the sun. Due to solar rotation the radial ejection of matter and consequently the IMF is given a spiral form, much as if the water from a hose pipe is rotated in a horizontal plane. This IMF has a sectored structure, the direction being inward and outward in adjoining sectors [see Figure 3.5]. Thus as the earth orbits the sun, so it passes into different sectors of the IMF.

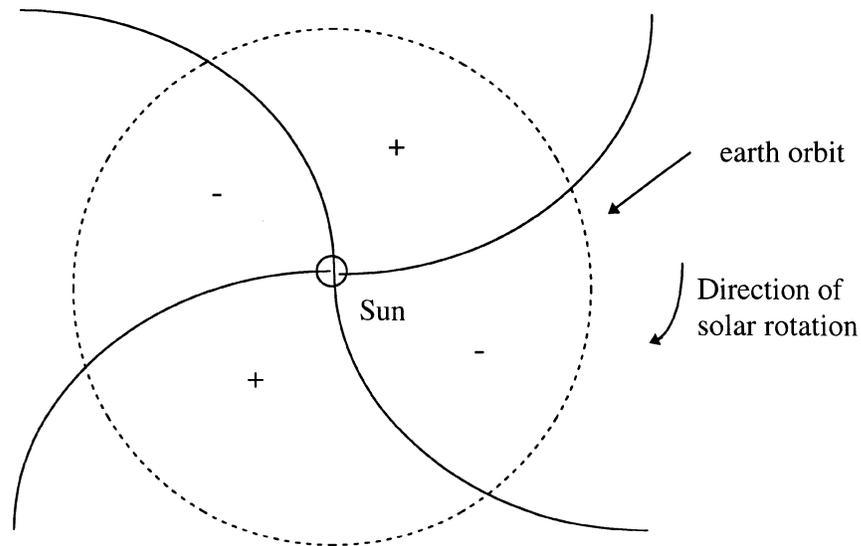


Figure 3.5 Form of the interplanetary magnetic field showing inward (-) and outward (+) sectored structure [after Hargreaves, 1992, p. 147].

The IMF interacts with the terrestrial magnetosphere (that region of the upper atmosphere where behaviour of charged particles is dominated by the geomagnetic field), and so affects circulation patterns within the ionosphere. This therefore leads to

differences in the driving forces of motions within the ionosphere at different latitudes [Hargreaves, 1992]. Although the cause is not fully understood, it is the form of such effects in which we are interested, these having been observed by the use of magnetometers at medium and high latitudes. The interaction of the IMF and the magnetosphere then leads to significant modification of the geomagnetic field lines and hence the ionosphere. Figure 3.6 shows clearly the major effect, including the formation of the magnetotail on the night side of the earth and the two neutral points on the magnetopause (the outward termination of the geomagnetic field).

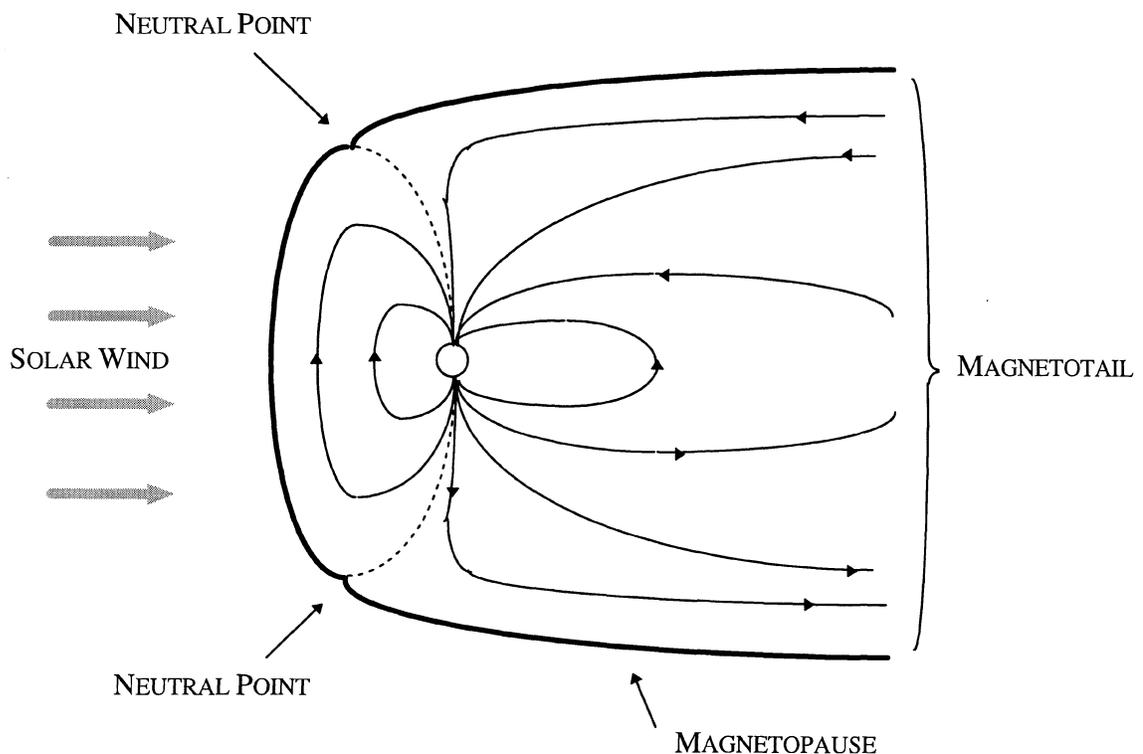


Figure 3.6 North-south cross section of the modified geomagnetic field showing closed and open geomagnetic field lines. [from Hargreaves, 1992, p. 152].

At these points the flux density of the field line falls to zero, and particles from the sun are allowed to enter into the lower reaches of the atmosphere above the poles without crossing magnetic field lines. The general area around this point is known as a polar cusp.

The sectored structure of the IMF also appears to have some relevance to the magnetospheric circulation, the north-south component being of more importance than the east-west. Whether the IMF is flowing northward or southward has important connotations for the level of interaction of the solar wind and the magnetosphere. It appears from observations of the magnetosphere that a southward flowing IMF will induce a much larger effect on the terrestrial magnetosphere. Major effects associated with such a situation include an inward movement of the magnetopause, and an equatorward displacement of the dayside cusps [Hargreaves, 1992].

Thus the behaviour of the ionosphere may be split into two broad latitudinal zones on this basis - the polar zone, and the zone including middle and low latitudes. (N.B. it should be noted that it is to the geomagnetic latitude that reference is given). Many exceptions exist to this classification, most of which are quite beyond the scope of this thesis. Anomalies to this classification in the equatorial ionosphere are given in section 3.4.1.

On examining the formulae given for f_0E and f_0F_1 given in section 3.2, it is obvious that there is some degree of dependence of the values of critical frequency on the solar zenith angle χ . This then suggests that the level of ionization will vary with latitude as this angle

changes. However, the most dramatic latitudinal variations are as a result not of χ but of the interaction of solar radiation and the magnetic field.

At middle and low latitudes, the behaviour of the ionosphere is more regular than at polar latitudes. There are still many variations attributed not only to solar effects, but also to other influences such as high altitude winds [Hargreaves, 1992]. **Sporadic-E** is one such variation, its production not entirely understood. Areas of higher than usual critical frequency exist, giving rise to rapid fluctuations in amplitude, phase and angle of arrival, known as scintillations, on trans-ionospheric signals. The typical critical frequency of a sporadic E cloud is approximately 10 MHz, again much less than the GPS frequencies, but great enough to produce a measurable effect.

3.4.1 The Equatorial Ionosphere

As an electromagnetic wave passes through an ionized plasma, it may be considered to be split into two components which are circularly polarised in opposite directions and propagate independently. These are known as the ordinary and extraordinary waves and their relative phase will gradually increase as they pass through the ionosphere. On leaving the plasma the waves recombine and a rotation of the plane of polarisation can be observed. This is the phenomenon of Faraday Rotation. The amount of this rotation is dependent on the electron content along the signal path and is thus used as a measure of the ionosphere [McNamara, 1991].

Scintillation (rapid fluctuations in amplitude and phase of the received signal) and Faraday Rotation patterns observed at equatorial latitudes point towards an equatorial anomaly, an increase in ionization at such latitudes. Due to the combined effect of electrical and magnetic fields, the electrons in these regions tend to drift upwards and then are propagated along the horizontal lines of force of the geomagnetic field. This phenomenon is known as the fountain effect, which results in peaks of electron density 10° to 20° either side of the equator. This effect may be attributed to $\mathbf{E} \times \mathbf{B}$ drift, where a motion is introduced into the electrons orthogonal to both the geomagnetic field and the ring current, formed when particles of opposite charge flow in opposite directions around the equatorial zone [Hargreaves, 1992]. Here \mathbf{E} denotes the electric field, and \mathbf{B} the magnetic field. $\mathbf{E} \times \mathbf{B}$ drift may be explained in terms of Lenz's law, whereby a motion is induced perpendicular to both the magnetic and electric field.

3.4.2 The Polar Ionosphere

In polar regions, behaviour of the ionosphere is greatly modified, in large part due to the fact that the geomagnetic field lines are almost vertical, and hence charged particles will tend to interact with the ionosphere.

The **gyrofrequency** of a charged particle in relation to a magnetic field \mathbf{B} , is the rate of gyration of the particle around lines of force of \mathbf{B} . With relation to the ionosphere, the charged particles concerned are the ions and electrons formed by solar terrestrial interactions, and the magnetic field is the geomagnetic field. The angular frequency ω_B of

this gyration is dependent on the mass of the particle and the magnetic field, as is the radius of the circle described. The gyrofrequency itself, however is independent of the particle velocity [Hargreaves, 1992]. If there is no component of velocity along the direction of the geomagnetic field, then any charged particles arriving will simply gyrate around the lines of force and will not be able to cross them. This is the situation found in equatorial regions where the geomagnetic field is close to horizontal. Moving away from the area of the geomagnetic equator will mean that some component of the velocity of charged particles of solar origin will be along the lines of force. Therefore the particle concerned will describe a spiral path, being brought down in to the atmosphere in polar regions where increased ionization will thus occur. This is especially true for high energy protons released by a large solar flare, which may cause a **Polar Cap Absorption** event or PCA, substantially increasing the level of ionization in the D-region of the polar ionosphere [McNamara, 1991]. The PCA is illustrated in Figure 3.7.

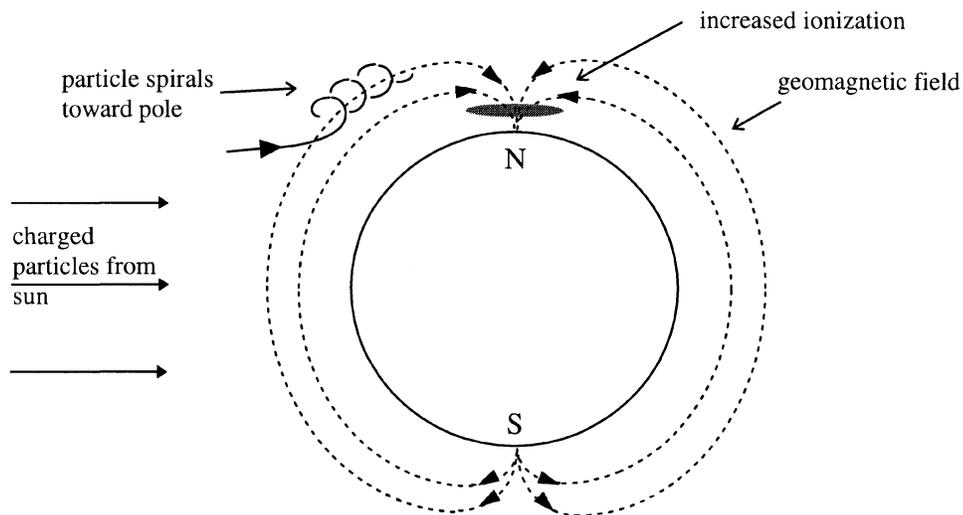


Figure 3.7 Basis for the Polar Cap Absorption event [from McNamara, 1991, p. 113].

3.5 Summary of Ionospheric Morphology

It will be beneficial to summarise the variability of the ionospheric make-up, paying special attention to effects which are apparent in the model produced here. The level of ionization is most stable in the mid-latitude regions where the influence of the geomagnetic field is felt less. As one moves toward the pole or equator then both absolute electron content and variability increases and the ionosphere becomes more difficult to model. A typical contour map of ionospheric delay for 1995, a time of low solar activity, is shown in Figure 3.8. The amount of delay [defined in section 3.6] induced by the ionosphere on an electromagnetic signal travelling through it is therefore an indication of the level of ionization.

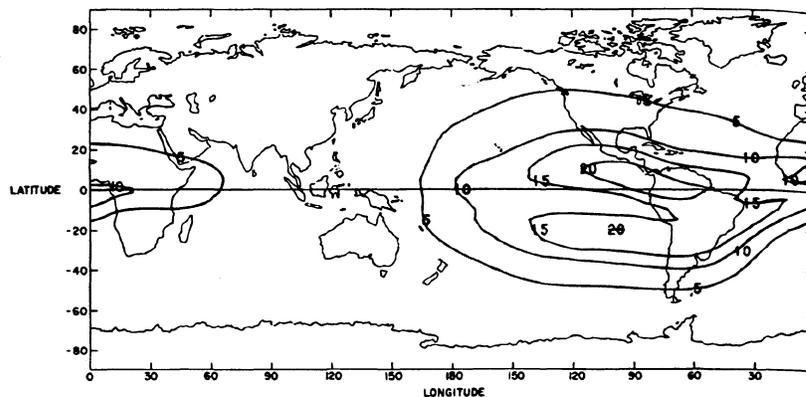


Figure 3.8 Contours of ionospheric zenith time delay in units of ns on L1 for the March equinox season for a year of low sunspot activity (sunspot number = 10) [from Klobuchar, 1996, p. 504]

As solar activity increases, governed by the 11 year solar cycle introduced earlier, so will the extent of ionization of the upper atmosphere. This is illustrated well by comparing

Figure 3.9 above to the following map of ionospheric delays at a time of heightened solar activity [Figure 3.9].

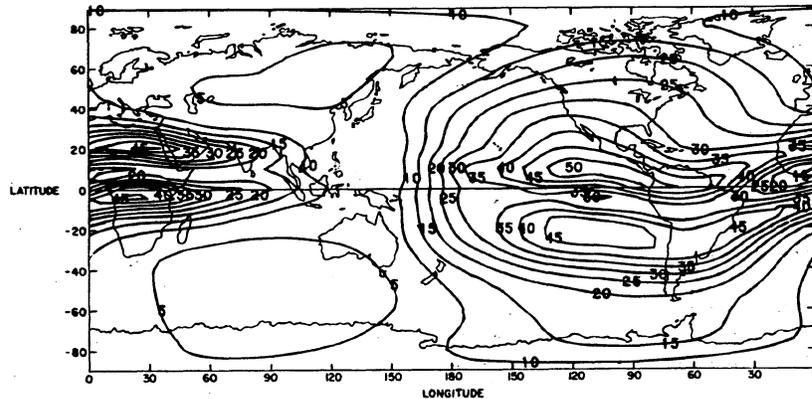


Figure 3.9 Contours of zenith ionospheric time delay in units of ns on L1 for the March equinox season for a solar maximum year (sunspot number = 153) [from Klobuchar, 1996, p. 503]

3.6 Ionospheric Propagation and Refraction

The propagation of microwave frequency electromagnetic waves through the ionosphere is a function of the signal frequency and the electron content of a unit cylinder along the propagation path [Georgiadou, 1994]. The ion content is not of direct consequence for electromagnetic wave propagation since the relative size of the ion means that it will not be excited by passing energy of this level to any significant extent. The number of electrons encountered by a signal is therefore of paramount importance, and some measure of this integrated electron density, the **Total Electron Content (TEC)**, is then required.

3.6.1 Phase Refractive Index

At microwave frequencies, the ionosphere may be regarded as a **dispersive** medium, i.e. the refractive effect of the ionosphere is proportional to the frequency of the signal. To enable the refractive effect of the ionosphere to be established, use is made of a truncated version of the **Appleton-Hartree formula**, the full version of which is given below:

$$n^2 = 1 - \frac{X}{1 - jZ - \left[\frac{Y_T^2}{2(1 - X - jZ)} \right] \pm \left[\frac{Y_T^4}{4(1 - X - jZ)^2} + Y_L^2 \right]^{1/2}} \quad (3.6)$$

where n is the complex refractive index of the ionosphere, $n = \mu - j\chi$. The values of X , Y and Z are dimensionless, being defined as follows:

$$X = \omega_N^2 / \omega^2$$

$$Y = \omega_B^2 / \omega^2$$

$$Y_L = \omega_L^2 / \omega^2$$

$$Y_T = \omega_T^2 / \omega^2$$

$$Z = \nu / \omega$$

where ω_N is the angular plasma frequency [equation (3.4) multiplied by 2π], ω_B is the electron gyrofrequency, and ω_L and ω_T are the longitudinal and transverse components of ω_B with respect to the direction of propagation. The electron collision frequency is ν , and the angular wave frequency ω .

This equation may be dramatically simplified by making some assumptions about the signal being propagated and the relative effects of some of the parameters described above

[Hargreaves, 1992]. Considering that for GPS the signal frequency is much greater than the plasma frequency, then if components of the geomagnetic field are omitted, and absorption is considered small (as will be the case for microwave frequencies being propagated in the rarefied ionosphere, cf. Z above, where if the collision frequency is small and the angular wave frequency large, Z will be close to zero), then the Appleton-Hartree equation may be written [Langley, 1996]:

$$n = 1 - \frac{X}{2(1 \pm Y_L)} \quad (3.7)$$

Now continuing the approximation process by ignoring the effects of longitudinal components of the geomagnetic field we obtain the following:

$$n = 1 - \frac{1}{2} \frac{f_N^2}{f^2} \quad (3.8)$$

From section 3.2 the plasma frequency, f_N is approximated as:

$$f_N = (80.5 N)^{1/2} \quad (3.9)$$

and therefore an approximation for the phase refractive index of the ionosphere may be written:

$$n = 1 - \frac{40.3 N}{f^2} \quad (3.10)$$

where N is the electron density.

The effect of such an approximation may be seen if the full form of the Appleton-Hartree formula is expanded to an infinite series and truncated at fourth order terms [Hartmann and Leitinger, 1984]:

$$n = 1 - \frac{1}{2} \left(\frac{f_N}{f} \right)^2 \pm \frac{f_N^2 f_g |\cos \theta|}{f^3} - \frac{1}{8} \frac{f_N^4}{f^4} - \frac{f_N^2 f_g^2}{4 f^4} (\sin^2 \theta + 2 \cos^2 \theta) \pm \dots (3.11)$$

where the signal frequency, f is related to its angular frequency by $f = \omega / 2\pi$, f_g is the electron gyrofrequency, and f_N is the plasma frequency. θ is the angle between the geomagnetic field and the direction of propagation of the wavefront in question. The magnitude of each of these terms has been estimated by *Bassiri and Hajj* [1993]. Typical values of 0 - 2cm and 0 - 2 mm for third and fourth order terms respectively are given. Omitting third and fourth order terms in this expansion is then justified for signal frequencies much greater than the plasma frequency since these terms will tend to zero for large f . Further justification for this approximation is given by *Hartmann and Leitinger* [1984]. They state that for signals above 100 MHz (i.e., in the range of interest here) since the electron gyrofrequency is everywhere less than 1.4 MHz, then for signal frequencies of this magnitude, neglecting terms in ω_b is valid.

3.6.2 Group Refractive Index

The group refractive index is defined as [Langley, 1996]:

$$n_g = n + f \frac{dn}{df} \quad (3.12)$$

where f is the frequency of the signal being propagated. Using the value of phase refractive index, n , obtained above we have:

$$\begin{aligned} \frac{dn}{df} &= \frac{d}{df} \left(1 - \frac{40.3N}{f^2} \right) \\ &= -40.3N \frac{d}{df} \left(\frac{1}{f^2} \right) \\ &= -2.(40.3N) \frac{1}{f^3} \end{aligned}$$

and therefore the group refractive index is given by:

$$\begin{aligned}
 n_g &= 1 - \frac{40.3N}{f^2} + 2 \frac{40.3N}{f^2} \\
 &= 1 + \frac{40.3N}{f^2}
 \end{aligned}
 \tag{3.13}$$

3.6.3 Ionospheric Group Delay and Phase Advance

For code measurements, the observed satellite receiver range is the integral of the group refractive index along the signal path:

$$\begin{aligned}
 S_c &= \int_{path} n_g ds \\
 &= \int_{path} \left(1 + \frac{40.3N}{f^2} \right) ds \\
 &= S + \frac{40.3TEC}{f^2}
 \end{aligned}
 \tag{3.14}$$

where S is the true range and TEC is the electron content integrated along the signal path. The modulated code range measurement shows a delay in its time of reception and hence an increase in the apparent range. This is termed the **group delay**. Similarly for phase observations:

$$\begin{aligned}
 S_p &= \int_{path} n ds \\
 &= \int_{path} \left(1 - \frac{40.3N}{f^2} \right) ds \\
 &= S - \frac{40.3TEC}{f^2}
 \end{aligned}
 \tag{3.15}$$

The phase measurement is then shorter than the true range and thus is advanced by an amount controlled by the signal frequency and electron density. This is the **phase advance**.

By comparing equations (3.14) and (3.15) it can be seen that the observed ranges from code and carrier measurements differ from the true value by an amount equal in magnitude but opposite in sign. Thus a measurement of the ionospheric delay from pseudorange observations may be used to correct the equivalent carrier phase observation.

3.7 Ionospheric Effects on GPS

The L-band signals transmitted from GPS satellites are classed as microwave. As such, the ionosphere acts as a **dispersive** medium for the GPS. The effect of the ionosphere is inversely proportional (in a first approximation) to the square of the frequency of the signal, as shown in the previous section. This effect can therefore be used to correct for much of the ionospheric effect if two signals of different microwave frequency are available. Indeed this is the case with GPS, where the L1 and L2 signals are transmitted at 1575.42 MHz and 1227.60 MHz respectively.

Consider for a moment the well-known observation equations for GPS pseudorange measurements, with receiver i observing satellite j on L1 [Kleusberg and Teunissen, 1996]:

$$P_{1,i}^j = \rho_{1,i}^j + I_{1,i}^j + T_i^j + c[dt_{1,i} - dt^{1,j}] + c[d_{1,i} + d^{1,j}] + dm_{1,i}^j + e_{1,i}^j \quad (3.16)$$

and similarly on L2

$$P_{2,i}^j(t) = \rho_{2,i}^j + I_{2,i}^j + T_i^j + c[dt_{2,i} - dt^{2,j}] + c[d_{2,i} + d^{2,j}] + dm_{2,i}^j + e_{2,i}^j \quad (3.17)$$

where the right hand side of each contains terms involving the geometric distance between satellite and receiver, ionospheric and tropospheric delays, satellite and receiver clock errors, satellite and receiver equipment delays, multipath, and unmodeled errors.

Differencing equations (3.16) and (3.17) results in:

$$P_{2,i}^j - P_{1,i}^j \approx \{I_{2,i}^j - I_{1,i}^j\} + c\{[d_{2,i} - d^{2,j}] - [d_{1,i} - d^{1,j}]\} + \{dm_{2,i}^j - dm_{1,i}^j\} + e_{2,i}^j - e_{1,i}^j \quad (3.18)$$

where the approximation sign is used to indicate that certain negligible terms have been omitted. Equation (3.18) is then the inter-frequency observable for pseudorange observations on L1 and L2. Terms involving the geometric distance between the satellite and receiver very nearly cancel and are therefore not shown, and the tropospheric effect may also be omitted on account of this delay being equal on both frequencies. Receiver clock errors cancel completely since the measurements are performed simultaneously on both frequencies. The satellite clock error interfrequency difference is non-zero due to slightly different time arguments. This residual is however negligibly small and has been excluded. Since the ionospheric effect changes for different frequencies (as per equation (3.10)), then this differencing operation will leave the difference in ionospheric effect between frequencies. Thus, a measure of the **Total Electron Content** (TEC) along the

line of sight to a particular satellite is may be obtained, subject to one major condition which is described in section 4.2.

3.8 The Single Layer Model of the Ionosphere

Since the ionosphere described above is obviously a rather complex region of the atmosphere, then some simplification in the modelling process is desirable. The approach taken here is to assume that the ionosphere may be represented locally by a homogeneous thin shell at a geocentric distance $R + h$, with R the earth radius and h a height chosen to represent the height of maximum ionization.

In order to then apply this model to the measurement of ionospheric delay described above, the delay along the signal path from receiver i to satellite j is obtained from the equivalent vertical delay by dividing by the cosine of the zenith angle measured at the ionospheric shell (see Figure 3.10) [Georgiadou and Kleusberg, 1988]:

$$TEC_i^j = VEC / \cos z_i^j \quad (3.19)$$

with VEC the vertical electron content and therefore *thickness* of the ionospheric layer. In order to utilise the assumption that the measured (line of sight) TEC may be mapped to the vertical using the zenith angle, it is necessary for local horizontal ionospheric density gradients to be small.

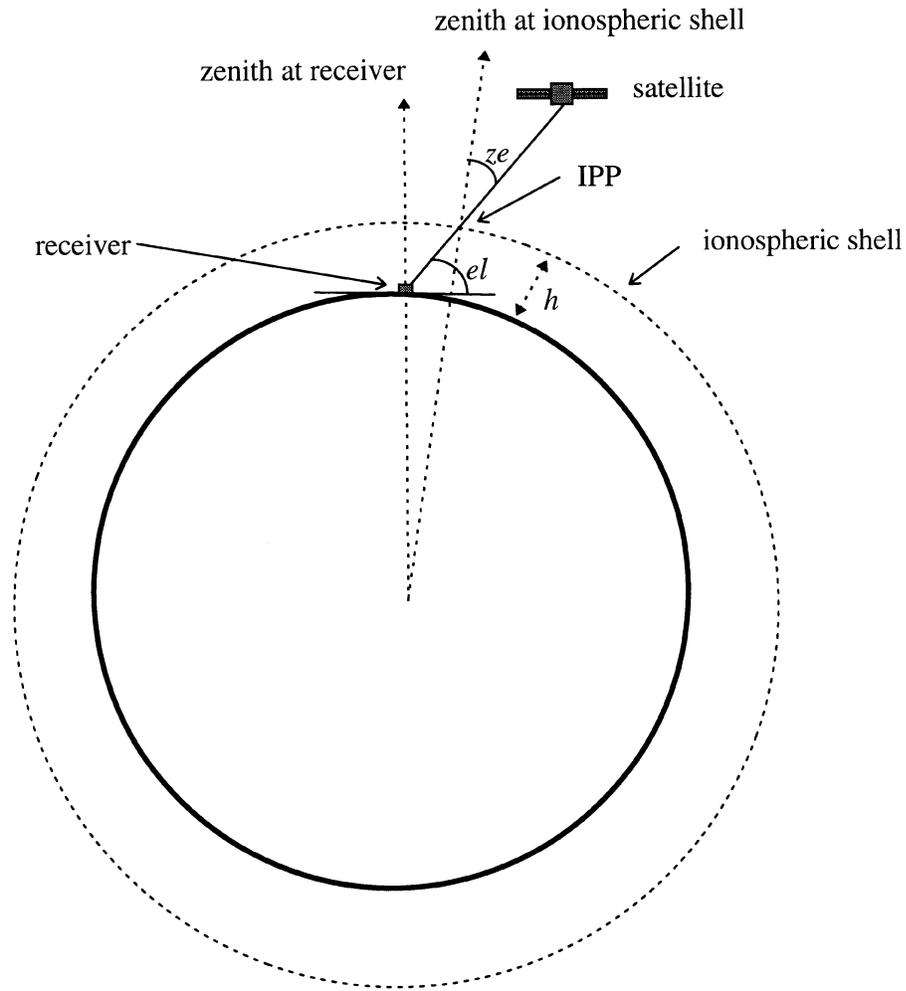


Figure 3.10 Geometry of satellite receiver line of sight and the ionospheric shell model.

To apply this mapping function, the zenith angle at the ionospheric shell is obtained from:

$$ze = \sin^{-1} \left(\cos(el) \cdot \frac{R}{R+h} \right) \quad (3.20)$$

where R is the earth radius and h is the assumed height of the ionospheric shell.

An attempt to model the local spatial variation of the slab thickness of this shell via a series of coefficients describing latitudinal and longitudinal variation in *TEC* from a base station was undertaken for this thesis. Thus the values of *VEC* will change dependent on the location on the shell intercepted by a particular satellite receiver line of sight. This location is termed the Ionospheric Pierce Point (*IPP*) (see Figure 3.10).

4 Chapter Four

Procedure

The procedure followed to correct for ionospheric delay at both the reference and rover receiver is outlined below, following step by step through the required stages.

4.1 Mapping Functions

In order for the ionospheric delay values obtained through the process described in equation (3.18) to be used in an ionospheric model, it is necessary to map these values to the vertical. Thus a vertical ionospheric delay estimate is produced, the ionosphere modelled as an infinitesimally thin shell at a particular height as described in section 3.8. It was mentioned in the introduction that the value chosen for this height is critical, and indeed this will be further expanded upon later in this chapter

This section looks at the effect that the function used to map the line of sight delay values to the vertical has on the vertical delay estimates. A search of the literature on ionospheric modelling showed a very few different mapping functions which have been investigated. Mapping functions used by various authors and agencies, such as those referenced in the literature review section of this thesis, tend to be simple functions of the

elevation angle of the satellite (el in Figure 3.10) and the height of the assumed ionospheric shell (h in Figure 3.10). The value used for the height of this shell is usually set to 350 - 400 km, chosen to reflect the height of maximum ionization. Some work has been done to further enhance the mapping function concept by using a model of the ionosphere to estimate this height [Komjathy and Langley, 1996].

Tests were undertaken as part of the work described here in order to quantify the effect any mismodelling might have. This was done by varying the assumed shell height and hence the mapping function. The results of these tests are outlined in section 4.3.

Figure 4.1 gives an indication of the transition from vertical to line of sight delays for a typical satellite, observed at station Albert Head in British Columbia. The line of sight delay value (LOS-DEL) is governed by the elevation angle, and is at a maximum when the elevation angle is small. This is a consequence of the trans-ionospheric signal travelling an oblique path through the ionosphere, spending a longer period of time within the ionized region. The ionospheric delay values are also noticeably more noisy at these low elevations, producing the “bow-tie” pattern in the vertical delay (VDEL) estimates. This is a consequence of the increased measurement noise caused by both multipath and by the lower received signal strength at these low elevation angles. It is for this reason that a relatively high elevation mask of 20 degrees was chosen in the surface estimation procedure to be introduced in section 4.4.1.1.

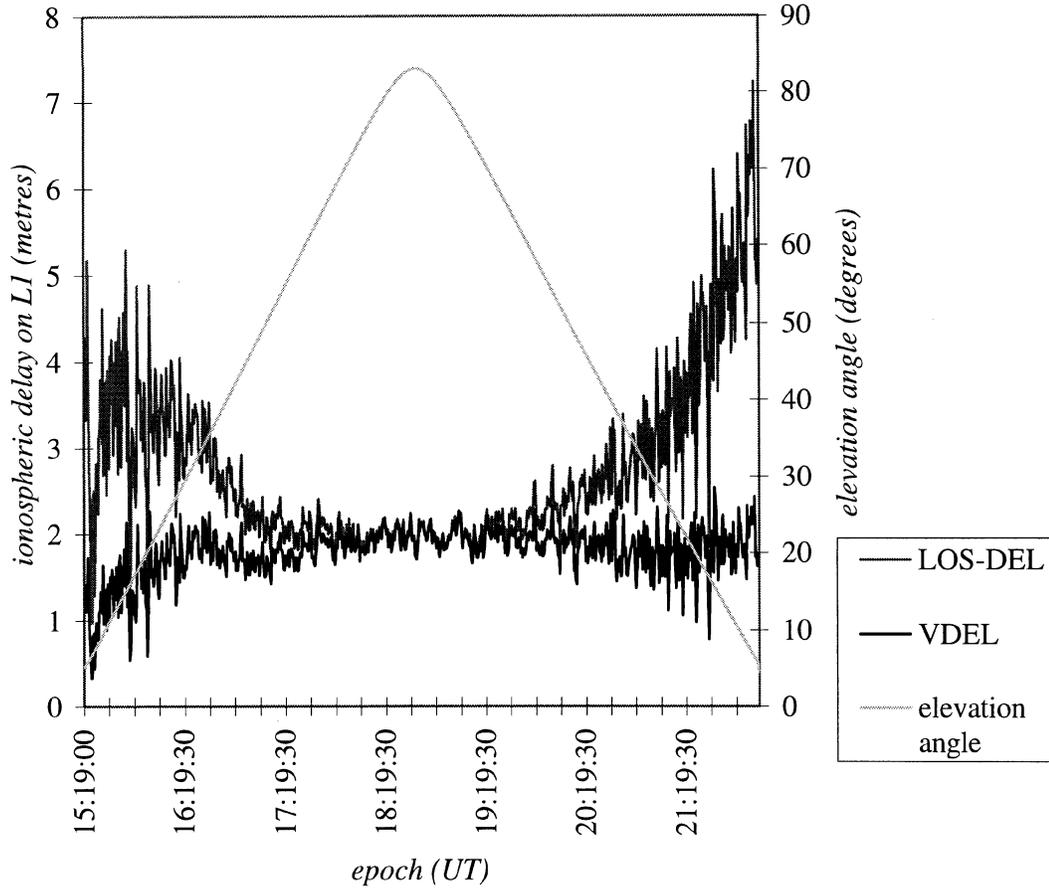


Figure 4.1 Line of sight and vertical ionospheric delays with SV elevation angle

4.2 Hardware Differential Delays

A vital part of ionospheric delay estimation is the calculation of the differential hardware delays associated with the particular satellite and receiver in question. Analysis of equation (3.18) shows that contained in the final expression for $P2 - P1$ are the delay values produced as a result of the different paths taken by the L1 and L2 signals through the satellite and receiver hardware. Therefore, in order for the geometry free linear

combination to yield a “true” estimate of the ionospheric delay, these equipment delays must be obtained. Ignoring the satellite (receiver) biases when computing TEC estimates will result in an error of ± 9 (± 30) TECU (1 TECU = 1×10^{16} e/m²) [Wilson and Mannucci, 1993]. Any error in estimating the receiver biases will manifest itself in the final VTEC values as an elevation dependent bias.

Since the P2 - P1 observable contains both the hardware and ionospheric delays, separation of the two becomes the problem in hand. A fortunate difference between the values is their response to elevation angle. The ionospheric delay is a function of elevation angle [cf. Figure 4.1] whilst the hardware delays are not. This then allows the ionospheric and hardware delays to be separated. The differential delays estimated for code measurements are assumed constant for each satellite and receiver pair and between observation periods for the same pair. The situation for carrier phase observations is a little more complicated since a new value must be estimated for each satellite pass, and after any loss of lock since the integer ambiguity value in the carrier phase observable will change. The combined hardware delays estimated from pseudorange observations also have a certain stability from day to day and thus values obtained from one day of observations may be used to correct the ionospheric observable from the following day within a certain accuracy. Based on these factors, it was decided that the model should be produced from code measurements despite the larger noise values associated with these.

In combination, the satellite and receiver biases can amount to delay values of the order of a few metres. Thus, in times of low to medium solar activity, the hardware delays may be larger than the TEC induced delay. Indeed, this is often apparent on inspection of raw pseudorange data from L1 and L2, where the L1 pseudorange may be greater than that from L2. This is the antithesis of the expected result, since with reference to equation (3.10), repeated here for convenience as equation (4.1), the ionospheric group delay on the higher frequency L1 signal would be expected to be smaller than that on the L2 pseudorange.

$$I_d = \frac{40.3 \text{ TEC}}{f^2} \quad (4.1)$$

For the purpose of this research, these differential equipment delays are produced as a by-product of the Vertical Delay (VERDE) software [Georgiadou, 1994]. This package requires an input of dual frequency GPS data, and solves for the equipment delays in a least squares adjustment.

Taking equation (3.18) and grouping a number of terms together for simplification, the following expression for the interfrequency pseudorange combination may be obtained [Georgiadou, 1994]:

$$F[P_2^j - P_1^j] = I_1^j + F[(S_{P2}^j - S_{P1}^j) + (R_{P2} - R_{P1})] + \varepsilon \quad (4.2)$$

where superscripts and subscripts indicate the transmitting satellite and frequency respectively and

$$F = \frac{f_2^2}{(f_1^2 - f_2^2)} \quad (4.3)$$

with f_1 and f_2 the frequency of the L1 and L2 carrier respectively. Satellite and receiver biases are grouped in the terms S and R . Thus an expression for the ionospheric delay referenced to L1 is obtained, along with the differential hardware delays.

As stated above, the basic premise behind the ability to separate the hardware and ionospheric delays lies with their differing response to changing elevation angle. Based on previous research by the authors of VERDE, the program models the ionospheric delay via a series of fifteen coefficients described by difference in latitude of the observing receiver and the sub-ionospheric point (the projection of the particular IPP to the earth's surface) and the local time. Writing equation (4.2) for each satellite observed and each observation epoch then produces a set of $n + 15$ equations for the fifteen coefficients of the parameterized ionospheric delay and the n hardware delays. Solving these equations by least squares gives an estimate of the required values.

VERDE was tested during this research to confirm that the equipment delays produced were indeed correct within the limits of the ionospheric model used. This testing phase also allowed a brief insight into the effect of the mapping function used to map line of sight delays to the vertical.

4.3 Testing of VERDE using Synthesised Data

By producing RINEX format input data for VERDE based on a known TEC value and shell height (350km) and computing differential delays, testing of the software was possible. The ionospheric delay was assumed to amount to 1.5m of delay on L1 and hardware delays were assumed zero. Since no hardware differential delay was input, it was expected that the delays estimated by VERDE would be zero if the same shell height is used in manufacturing the test data and in VERDE. Although this did indeed prove to be the case, further analysis with similar manufactured data brought up some interesting results.

It was possible to see the effect of using an incorrect shell height in the thin shell approximation by changing the value used in VERDE. Table 3 shows the resulting equipment delays obtained on varying the estimated shell height used. A 50km misrepresentation of this value resulted in an approximately 10cm error in the estimation of the differential delays. This amounts to approximately 0.6 TECU of ionospheric delay and is a significant proportion of the input delay value of 1.5m. Increasing the input ionospheric delay would lead to a proportional increase in the error in the hardware delay estimates.

Shell Height (km)	300	350	400	500
SV	<i>Estimated Hardware Delay (m)</i>			
1	0.107	0.000	-0.101	-0.304
2	0.111	0.000	-0.099	-0.303
4	0.115	0.000	-0.103	-0.313
5	0.112	0.000	-0.101	-0.307
6	0.111	0.000	-0.102	-0.308
7	0.112	0.000	-0.099	-0.305
9	0.117	0.000	-0.099	-0.307
12	0.112	0.000	-0.099	-0.303
14	0.111	0.000	-0.104	-0.314
15	0.137	0.000	-0.079	-0.288
16	0.121	0.000	-0.108	-0.328
17	0.112	0.000	-0.101	-0.307
18	0.109	0.000	-0.101	-0.305
19	0.111	0.000	-0.101	-0.305
20	0.112	0.000	-0.100	-0.310
21	0.108	0.000	-0.102	-0.309
22	0.123	0.000	-0.104	-0.329
23	0.112	0.000	-0.108	-0.307
24	0.112	0.000	-0.101	-0.307
25	0.108	0.000	-0.103	-0.306
26	0.124	0.000	-0.106	-0.326
27	0.112	0.000	-0.100	-0.305
28	0.121	0.000	-0.104	-0.319
29	0.107	0.000	-0.099	-0.300
31	0.124	0.000	-0.106	-0.326

Table 3 Estimated hardware delays from VERDE showing the effect of changing ionospheric shell height.

It can be seen that the hardware delays are correctly estimated as zero when the correct shell height (350km) is used in the ionospheric delay estimation. It is also apparent from equation (4.2) that any error in calculating these differential delays will directly affect the TEC estimation itself by an equivalent amount. A major reservation therefore is that the

separation of such hardware delays from the ionospheric delays is dependent on the data being correctly modelled by the thin shell approximation used. Since the synthetic data was created based on this assumption, then any errors associated with this type of mismodelling were not detected in this testing phase. This is obviously not the case with “real” data however.

4.4 Construction of the Ionospheric Model

In a real world, real time environment the hardware delays from VERDE would be estimated from a previous days observations. Therefore, some day to day stability of these values must be assumed. The hardware delays estimated in this manner are then applied to the P2 - P1 observable and a measure of ionospheric delay is thus produced.

4.4.1 TECEST

As outlined above, forming the geometry free linear combination of the P (Y) code pseudorange observables provides a measure of the ionospheric delay along the line of sight to a satellite, providing the hardware differential delays are known.

The software written for the purpose of this research takes a RINEX format file from a stationary receiver and processes the P (Y) code pseudorange data sequentially, assuming that only data up to the current epoch is available. Thus a real time approach is synthesised. The geometry free linear combination between frequencies is taken, and the

resulting ionospheric delay estimates along the line of sight to each satellite in view converted to the equivalent vertical delay by applying a mapping function as described in section 4.1. Figure 4.2 shows the relative position of the IPPs and reference station for a shell height of 350km. It is apparent that the distance between that point directly overhead the reference station and the point on the ionospheric shell intersected by the satellite receiver line of sight is relatively small in global terms. It was therefore assumed that a simple spatial model of the ionospheric shell thickness would be appropriate. The vertical delay values measured at each IPP are then passed through a least squares surface fit routine, the output of which is a series of coefficients which describe an ionospheric delay surface in latitude and longitude differences from the reference station.

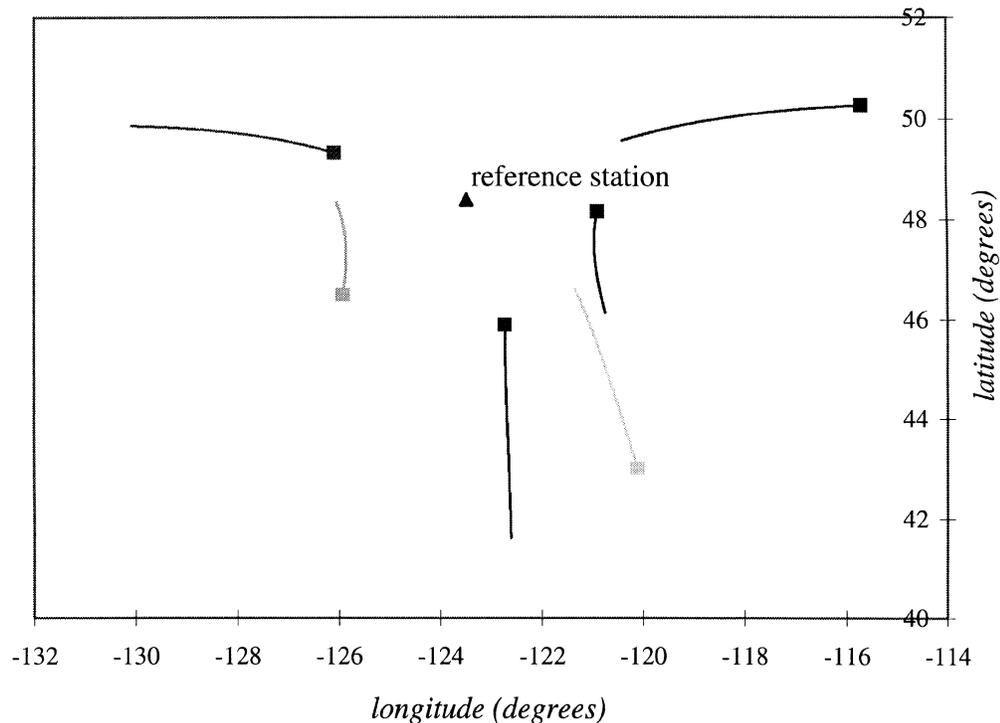


Figure 4.2 Typical latitude and longitude of Ionospheric Pierce Points (IPPs) used in surface estimation procedure. Elevation cut-off is 20 degrees, shell height is 350km. Marker shows last position of IPP.

4.4.1.1 Polynomial Surface Fitting

To describe a surface, a polynomial in two variables is required. The basis functions for four classes of such polynomials are given in Table 4 below.

Class	Basis Functions
P_0	1
P_1	1 x y
P_2	1 x y x^2 xy y^2
P_3	1 x y x^2 xy y^2 x^3 x^2y xy^2 y^3

Table 4 List of basis functions for some bivariate polynomials [from Lancaster and Šalkauskas, 1986, p. 133].

It is apparent that any function in class P_0 will provide a constant surface, parallel to the xy plane. An increased number of coefficients in the polynomial will give a more complicated surface more able to follow the data points. The choice of a suitable polynomial to describe the surface in question is then a trade-off between computational and programming efficiency and the ability to correctly represent the data points. Given that the ionospheric delay values used as data in the surface fit are estimates of the true picture, then increasing the number of coefficients in the polynomial fit may be counterproductive. An increased class of polynomial will decrease the residuals of the surface fit but also assumes a greater accuracy in the ionospheric delay estimates than may actually be the case.

The number of data points is also a factor which had to be taken into account in choosing the basis function to be used. Since a least squares fit of the surface was to be attempted then some redundancy in the measurement process is required. Thus the number of measurements should be greater than the number of coefficients to be estimated.

Because of these constraints it was decided to base initial investigations around a class P_1 polynomial. The surface computed by this routine is therefore described by:

$$\begin{aligned} \text{VTEC}_s &= a_0 + a_1(\phi_0 - \phi) + a_2(\lambda_0 - \lambda) \\ &= a_0 + a_1\Delta\phi + a_2\Delta\lambda \end{aligned} \quad (4.4)$$

where VTEC_s is the estimated vertical TEC value at the Ionospheric Pierce Point (IPP) at latitude ϕ and longitude λ . Latitude and longitude values are in units of radians. The IPP at (ϕ_0, λ_0) refers to the position directly above the reference station. The coefficients a_0 , a_1 and a_2 thus estimated are then used as the initial state for implementation in the Kalman filter described in the next section.

The surface described by these coefficients should provide a model of the temporal and spatial processes described in Chapter Three. It was therefore deemed necessary to constrict the variation of the coefficients such that the underlying physical processes are adequately modelled.

4.4.1.2 Kalman filtering algorithm

To update the ionospheric model used in this work, a Kalman filter was implemented. Advantages of this approach include the ability to take previous measurements into

account, and to better cope with the possibility of unavailable observations at a particular epoch.

The values estimated as the filter state are the three coefficients of the TEC surface fit; a_0 , a_1 and a_2 . The measurement update model was assumed to contain no underlying forcing functions, i.e., the state update matrix is simply the (3 x 3) identity matrix. Thus the state update is governed solely by the process noise, chosen so as to allow the Kalman filtered state to follow the general trend of the real data as closely as possible whilst suppressing measurement noise.

The measurement noise value input to the filter gives an indication of the accuracy of each new measurement. The third set of required input parameters are the variances of the initial state estimate. These are computed from the residuals of the initial TEC fit for the first epoch. The initial covariance matrix of the state is then a diagonal matrix (assuming zero correlation between the initial state parameters).

The formation of the Kalman filter is based on that given by *Brown and Hwang* [1992]. Initial input consists of an *a priori* estimate of the state (in this case provided by the least squares fit of the TEC measurements from the first epoch, described in the previous section) and its associated covariance matrix described above. Also required are values for the process and measurement noise as described above. An important step in the

implementation of a Kalman filter is then the choice of measurement and process noise values to be used.

Following the concept of Figure 4.3, the first task in the implementation of the Kalman filter is to compute the Kalman Gain (\mathbf{K}_k) to be used to assimilate the current state estimate (projected from the previous epoch) and the current measurement. This gain is computed from:

$$\mathbf{K}_k = \mathbf{P}_k^- \mathbf{H}_k^T (\mathbf{H}_k \mathbf{P}_k^- \mathbf{H}_k^T + \mathbf{R}_k)^{-1} \quad (4.5)$$

where \mathbf{P}_k^- is the estimated state error, \mathbf{H}_k^T is the design matrix describing the connection between the measurement and state vector at time t_k , and \mathbf{R}_k is the measurement noise.

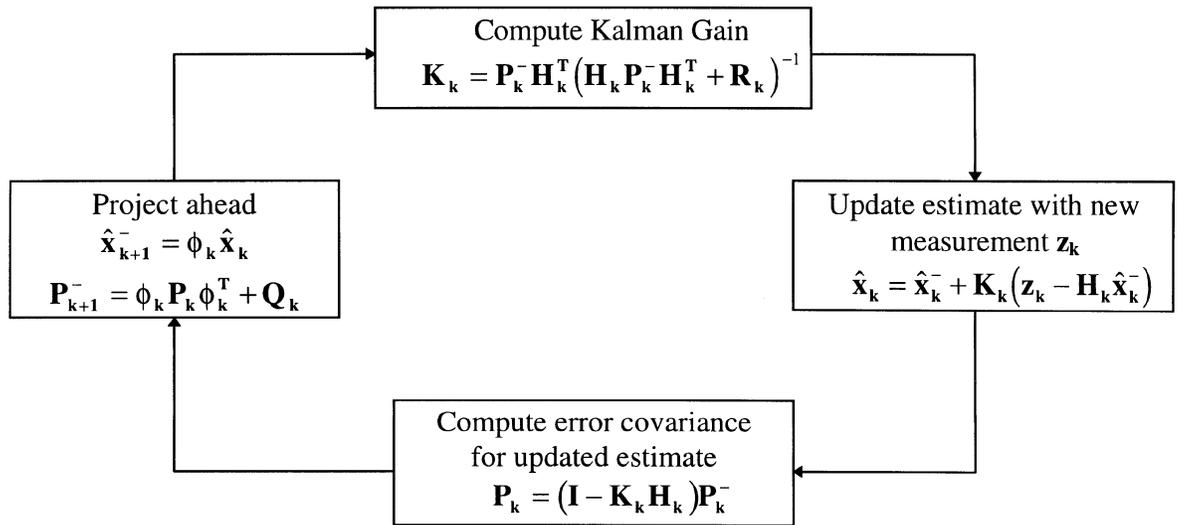


Figure 4.3 Recursive Loop of standard Kalman Filter [from Brown and Hwang, 1992, p. 235].

The state is then updated based on a weighted formulation of the prior estimate and current measurement, following equation 4.6 below:

$$\hat{\mathbf{x}}_k = \hat{\mathbf{x}}_k^- + \mathbf{K}_k (\mathbf{z}_k - \mathbf{H}_k \hat{\mathbf{x}}_k^-) \quad (4.6)$$

where $\hat{\mathbf{x}}_k$ is the new state estimate, updated according to a certain combination of the prior estimate $\hat{\mathbf{x}}_k^-$ and the new measurement \mathbf{z}_k , governed by the filter gain \mathbf{K}_k .

The next stage is to update the state covariance matrix to take account of the influence of the new measurement on the confidence given to the updated state:

$$\mathbf{P}_k = (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k) \mathbf{P}_k^- \quad (4.7)$$

where \mathbf{P}_k is the updated state covariance, \mathbf{I} is the identity matrix, and \mathbf{K}_k , \mathbf{H}_k , and \mathbf{P}_k^- are as before.

The state and state covariance matrix are then projected forward to the next epoch as shown in equation (4.8), and the process begun again.

$$\begin{aligned} \hat{\mathbf{x}}_{k+1}^- &= \phi_k \hat{\mathbf{x}}_k \\ \mathbf{P}_{k+1}^- &= \phi_k \mathbf{P}_k \phi_k^T + \mathbf{Q}_k \end{aligned} \quad (4.8)$$

where $\hat{\mathbf{x}}_{k+1}^-$ and \mathbf{P}_{k+1}^- are the projected state estimate and its associated covariance for time t_{k+1} , ϕ_k is the state transition matrix (in this case the identity matrix), and \mathbf{Q}_k is the process noise.

It can therefore be seen that the Kalman filter allows the state estimate to be based not only on the current epoch, but on all past data too, without the need to store large amounts of data.

4.4.1.3 Computation of Kalman filter inputs

Computing initial values for the process and measurement noise is an important part of the implementation of the Kalman filtering algorithm described above. It will be noticed that the process noise \mathbf{Q} controls the change in the projected state estimate in the absence of any forcing function, as is the case here. This is a far from simple task, and is glossed over in many texts and papers on the implementation of Kalman filters.

The approach considered here was to apply a number of potential values of the process and measurement noise, and evaluate their worth based on the closeness of the fit to the original data. This is a methodological, if not entirely mathematically rigorous approach. It is suggested that future work in this area would include a mathematical analysis of the measurement and process noise associated with an implementation of a Kalman filter to ionospheric estimation.

The balance required in the implementation of a Kalman filter is to allow the state estimate to take account of the input data enough to enable the state to follow the general trend without being effected by the noise of the measurement itself.

The purpose of any filter is to separate one thing from another. In the Kalman filter, the rejection of one filter state estimate over another is based upon the probability of the occurrence of a particular estimate. The applicability of a Kalman filtering algorithm is particularly useful when both the process being modelled and the unwanted parts of the

signal are random in character. The filter must thus give some best compromise of allowing the underlying process to pass and suppressing the unwanted background noise, without constricting the filter state so much that filter divergence takes place.

Filter divergence exhibits itself by the gain becoming so small that new measurements have no effect on the output state. This is often caused by having an incorrect ratio between the measurement and process noise [Schwarz, 1987]. Filter divergence may also be a result of an oversimplified model, where an error will be caused between the actual system noise and the estimate given by \mathbf{Q} . Filter divergence may be remedied by increasing the filter process noise artificially. This is however a dangerous procedure, since a direct and first order effect will be apparent in the state covariance matrix should an incorrect process noise be input (see equation (4.8)).

The output covariance estimates given by the Kalman filter are obviously directly affected by the input weight estimates given in \mathbf{R} and \mathbf{Q} . Thus, wrongly estimating these matrices will result in false estimates of the accuracy of the filter state. It is then important that the input matrices are representative of a large enough actual data set to be reliable [Schwarz, 1987].

In this particular case, data analysis of the fit of epoch by epoch calculated ionospheric parameters were calculated, and the errors so obtained used as the input for the initial state error covariance \mathbf{P}_k^- . Values for the measurement error \mathbf{R} are estimated based on the

known performance of the receiver in question and the modelling process used to obtain the ionospheric delay measurements. Typical values of P-code pseudorange error are of the order of 2m [Parkinson, 1996]. Rewriting the simplified interfrequency pseudorange linear combination and omitting the hardware biases since these are now assumed known we have:

$$I_d = P_2 - P_1 \quad (4.9)$$

Now applying the technique of error propagation to equation (4.9):

$$\Sigma_I = A \Sigma_{P_1, P_2} A^T \quad (4.10)$$

a value of 2.82m is obtained as the *a priori* standard deviation of the ionospheric delay estimates (assuming zero correlation between P_1 and P_2).

The process noise \mathbf{Q} is more difficult to obtain, and a trial and error procedure was followed, with the criteria a best fit to the general trend of the epoch by epoch calculated data. *Doherty et al.* [1994] conducted an investigation into time rate of change of ionospheric delay for a data set spanning the time period from December 1993 to July 1994, well down the solar cycle decreasing limb. Statistics of the time rate of change for intervals of 1, 2, 5 and 10 minutes were computed, the results showing a maximum variation for a 1 minute span of approximately 0.3m. This maximum deviation was observed at a high latitude station, more affected by the auroral influence. At times of high solar activity, a three fold increase in variation is anticipated, the induced delay not exceeding 1 metre of variation over a 1 minute observation period.

The data used in the research outlined here was collected in March of 1995, near the minimum of the solar cycle. Thus it may be anticipated that variations will be smaller yet, and therefore an approximate process noise value may be estimated. The filter update rate was initially set to the collection rate of the original observation data (i.e. a smoothing process is carried out) and therefore using a value of 1 metre per minute of delay rate of change, the process noise was initially set to 0.25m^2 for a 30 second filter update rate.

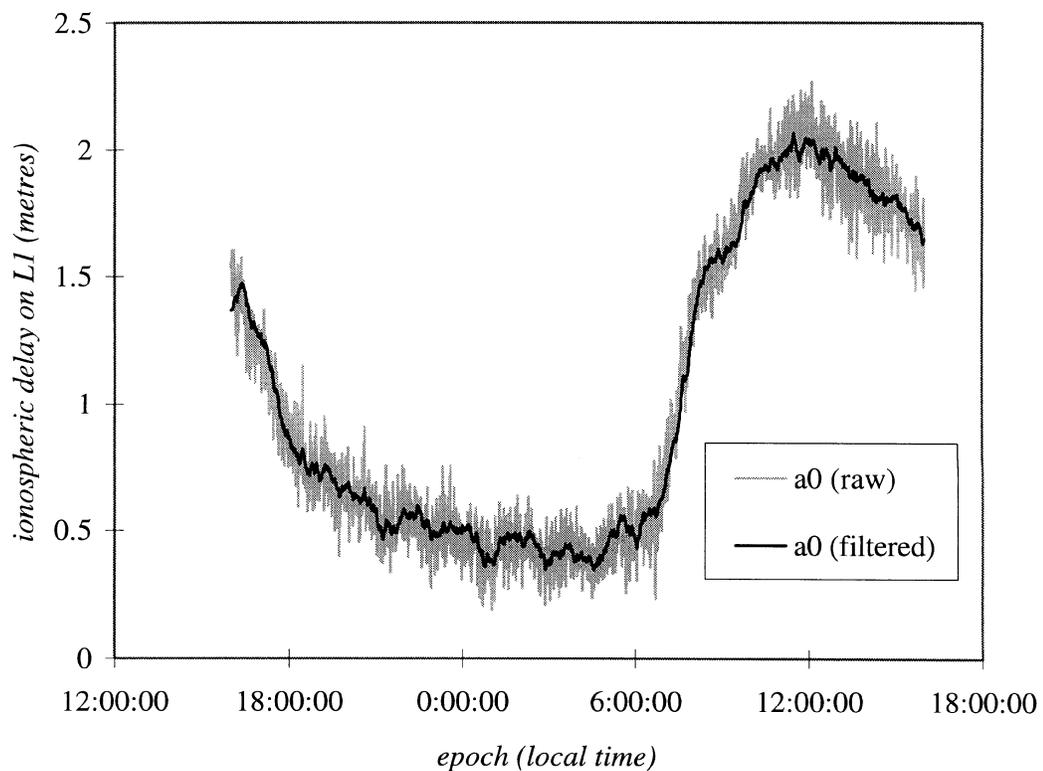


Figure 4.4 Constant coefficient of surface fit for day 298 of 1995 at station Albert Head. The variation of the a_0 coefficient on application of the Kalman filter is clearly seen.

Figure 4.4 clearly shows the impact of the filtering process on the estimated constant ionospheric surface coefficient. The filtered a_0 coefficient follows its raw counterpart's

general trend, much of the high frequency noise having been removed without any noticeable filter lag. Similar outcomes for the latitude and longitude coefficients are well illustrated in figures 4.5 and 4.6 below.

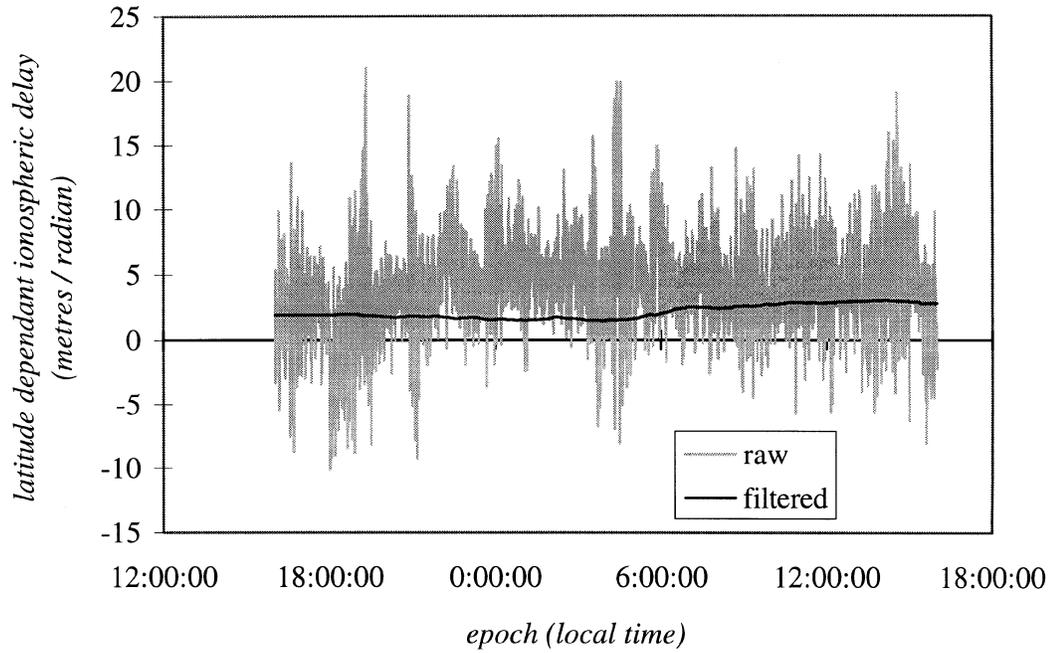


Figure 4.5 Raw and filtered latitude dependent coefficient of ionospheric delay surface for day 298 of 1995 at station Albert Head.

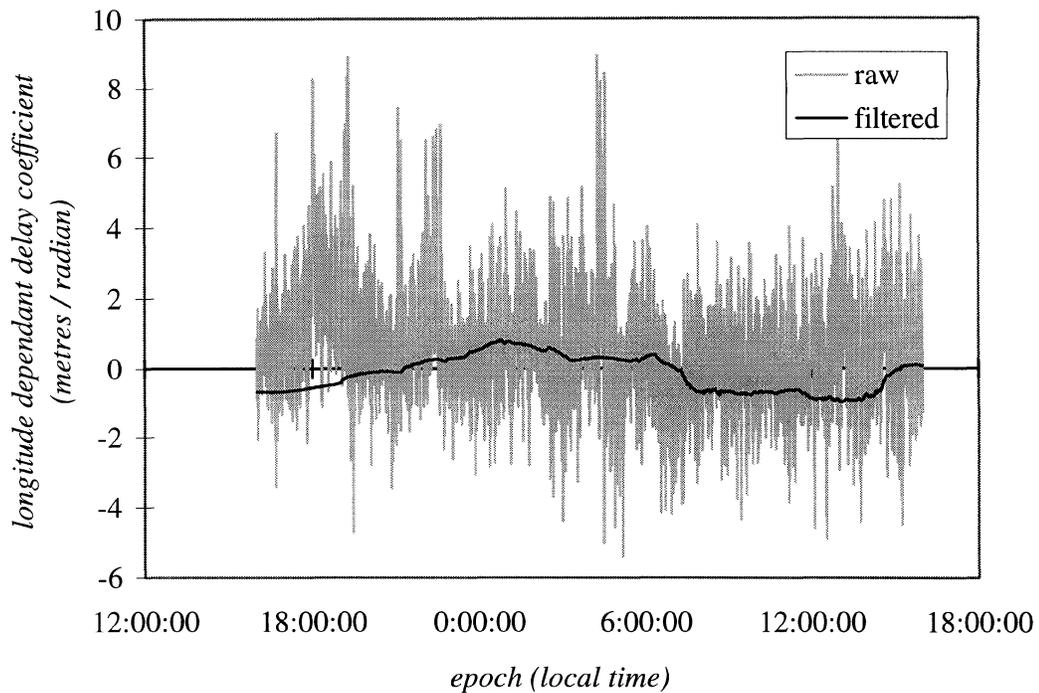


Figure 4.6 Raw and filtered longitude dependent coefficient of ionospheric delay surface fit for day 295 of 1995 at station Albert Head.

4.5 Application of the Ionospheric Model

Having produced an ionospheric delay model which describes local spatial and temporal variations, it is now required to correct the GPS carrier phase observations such that the estimated ionospheric delay is removed. Corrections are computed and applied at both reference and remote stations according to the position and time of observation of each IPP following equation (4.4).

With reference to equation (4.4), the correction applied to a particular satellite observed at both reference and rover stations is a function of two variables. At a particular

epoch, the correction computed is controlled by the latitude and longitude of each IPP. As the station separation increases so the positions intercepted on the ionospheric shell from reference and rover diverge as illustrated in the highly stylised Figure 4.7.

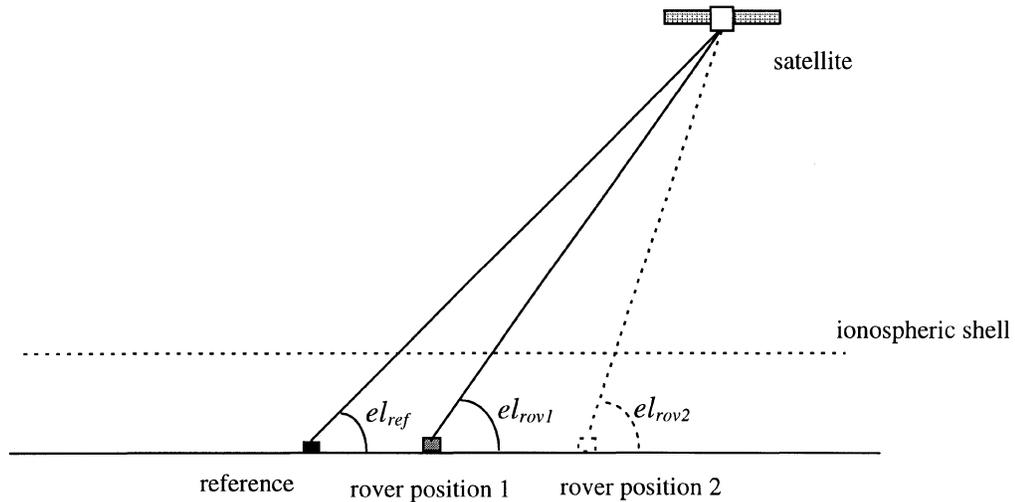


Figure 4.7 Change in IPP location and elevation angle with baseline length. As the rover moves to position 2 so the distance between IPPs and the difference in observed elevation angles at reference and rover increases.

The line of sight delay calculated for each satellite-receiver combination is calculated from its equivalent vertical delay by applying the inverse of the mapping function used in the surface estimation procedure (equation (3.19)). The elevation angle calculated is also a function of receiver position and hence will be a second controlling factor in the calculation of ionospheric delay estimates (see Figure 4.7).

4.5.1 ROVER

In a real world environment it is anticipated that the corrections calculated from the geometry free linear combination of the L1 and L2 codes at the base station would be

transmitted to the roving station and applied. The approach taken in this initial investigation was to simulate the real time application of ionospheric corrections by modifying the carrier phase data with the ionospheric delays given from the epoch by epoch ionospheric delay surface estimation.

The surface described by the coefficients output by TECEST are in units of metres of ionospheric delay on L1. On multiplying by suitable factors, the delay induced on L1 and L2 may be obtained in units of cycles. Considering also that the value produced from the surface estimation refers to a vertical delay, it is then necessary to map the delay to the line of sight to each satellite as outlined in the previous section.

The input required for this mapping function is the zenith angle, as shown in Figure 3.10. Thus, in order calculate the line of sight delays, it was required to compute an approximate position of the rover receiver since the zenith angle is a function of satellite elevation angle measured at the receiver. This was done in a simplistic fashion by solving for the position via a system of pseudorange observation equations as described in e.g. *Kleusberg and Teunissen [1996]*. The data used in the initial development of this software contained P-code pseudoranges and these were therefore used to calculate position.

Figure 4.8 gives a concise outline of the procedure followed to apply the “transmitted” corrections to the L1 and L2 carrier phase data.

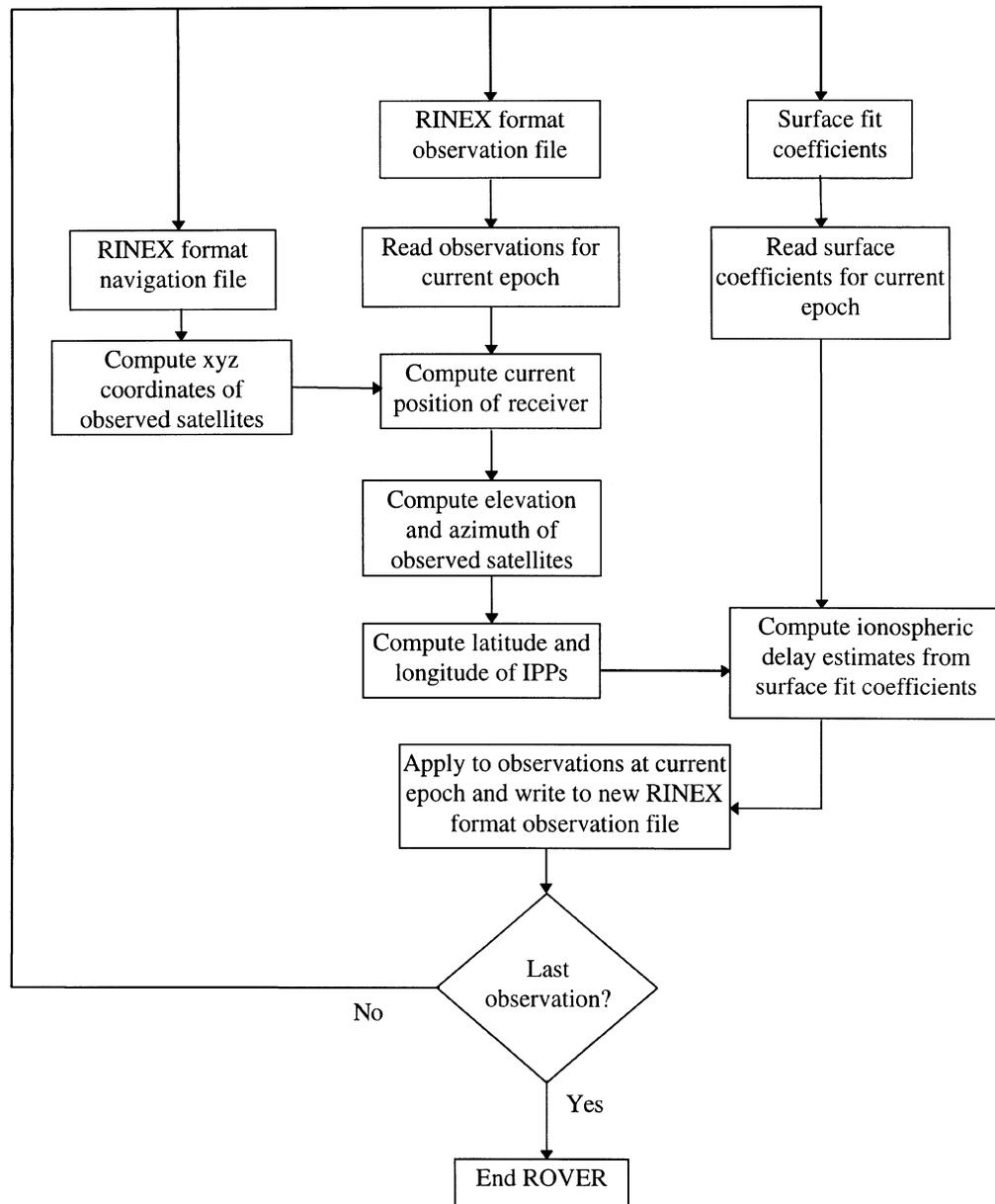


Figure 4.8 Flow diagram showing the procedure followed to calculate and apply ionospheric corrections at the rover receiver.

The first stage is to compute the required input for the calculation of the vertical ionospheric delay at each IPP. This involves calculation of ECEF coordinates for the

satellites in view and consequently the latitude and longitude of each intersection with the ionospheric shell. This information is then used along with the three coefficients to compute a vertical ionospheric delay value at each IPP. Then by mapping these values to line of sight estimates and converting to a delay in units of cycles on L1 and L2, the corrections may be applied.

A latency check is carried out to ensure that the corrections used for a particular epoch are referenced to the same time tag at both reference and rover receiver.

4.5.2 Reference Receiver Corrections in TECEST

A similar procedure is followed for the base station, controlled by the TECEST software outlined in the previous section. The correction process here is slightly simpler however since the satellite orbital positions and related IPP latitude and longitude are already available from the ionospheric delay estimation procedure.

Applying the corrections calculated from the surface fit provides continuity between the corrections applied at base and rover, an important point in any DGPS system.

4.6 Ambiguity Resolution

As previously intimated, the validation of the corrections produced by TECEST and applied by ROVER was carried out by attempting the resolution of ambiguities both

before and after these corrections have been applied. Many texts and papers [e.g. Abidin, 1993; Teunissen, 1996] exist on the theory of OTF ambiguity resolution and consequently it does not seem necessary to provide an exhaustive treatment of the subject. Nonetheless it will be instructive to provide a brief introduction to the subject.

Looking at the carrier phase observation equations provided below, an additional unknown is present when compared to the equivalent pseudorange formulae [equations 3.16 and 3.17]. Since creation of the carrier phase observable involves the measurement of only the fractional phase of the passing signal, the integer number of wavelengths between the receiver and satellite antennae at the initial epoch is unknown. This is the integer ambiguity N in equation (4.11) and Figure 4.9.

$$\Phi_{ij}^{kl} = \rho_{ij}^{kl} - I_{ij}^{kl} + T_{ij}^{kl} + \delta m_{ij}^{kl} + \lambda N_{ij}^{kl} + \varepsilon_{ij}^{kl} \quad (4.11)$$

The linear combination Φ_{ij}^{kl} is the double differenced phase observable introduced earlier.

Therefore in order to take advantage of the inherently lower noise carrier phase measurement it is necessary to solve for this new unknown. In a kinematic observation environment, one must often be able to solve for this ambiguity “on-the-fly”.

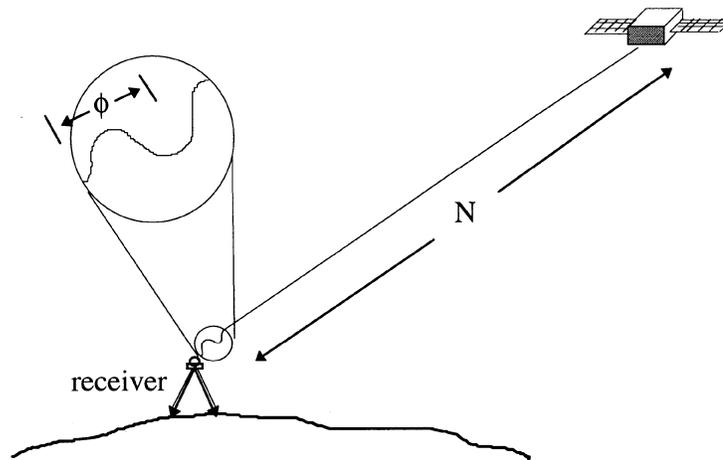


Figure 4.9 The ambiguity resolution problem.

The initial intention had been to use the Ashtech post-processing package PRISM for this task, in particular the PNAV OTF ambiguity resolution module. This had unfortunately to be abandoned since it was discovered that PNAV will not allow the user to input anything other than original Ashtech data. Despite the use of the RINEX to Ashtech format file conversion routine contained in PRISM it was found impossible to use the corrected data from TECEST and ROVER.

An alternative solution proposed was to use the Kinematic And Rapid Static (KARS) ambiguity resolution package available from the University NAVstar Consortium (UNAVCO). KARS uses the ambiguity function method to resolve ambiguities from an input of generic RINEX format data [Mader, 1992, 1995]. The ambiguity function method has the advantage that the function value is unaffected by cycle slips over the summation period.

This ambiguity function may be written [Mader, 1992]:

$$A(x, y, z) = \sum_{k=1}^K \sum_{j=1}^J \sum_{L=1}^2 \cos\left(2\pi\left[\phi_{obs}^{kjL}(x_o, y_o, z_o) - \phi_{calc}^{kjL}(x, y, z)\right]\right) \quad (4.12)$$

where ϕ_{obs}^{kjL} is the double differenced phase in units of cycles measured at the rover antenna whose correct position is (x_o, y_o, z_o) . Superscript L refers to the observed frequency, j to the satellite and k to the epoch of observation. ϕ_{calc}^{kjL} refers to the calculated double differenced phase at a trial position (x, y, z) . The difference between the observed and calculated phase observation is the *a priori* phase residual ignoring the integer biases. These may be safely omitted since they will cause an integer number of rotations to the ambiguity function leaving its value unchanged. The maximum value of the normalised ambiguity function is 1, obtained by dividing equation (4.12) by the sum of the frequency, satellite and epoch indices. This maximum occurs when $(x, y, z) = (x_o, y_o, z_o)$ and at all other points at which the difference in calculated distance between this satellite and the reference satellite changes by an integer number of wavelengths. By observing at different frequencies, epochs and satellites, a surface of maxima and minima is built up, interfering constructively at the most likely position.

The volume within which integer set candidates are searched is also set by the user, reflecting the confidence in the initial position of the centre of this search space. In this case, the centre of the initial search space was set using the same position produced in ROVER for the use in the computation of ionospheric delays [section 4.5.1]. Thus the search space volume was relatively large due to uncertainties in this position.

5 Chapter Five

Results

Once the software described in the previous chapter to compute and apply the ionospheric corrections required at both rover and reference had been implemented, analysis of the output of both TECEST and ROVER was carried out. Evaluation of the influence of these corrections was also done by examining their effect on carrier phase ambiguity resolution.

5.1 Multipath

As the GPS signal arrives at the antenna, not only the direct signal but that reflected from surrounding objects is received [see section 2.2]. Since a longer path is taken by the reflected signals, they arrive with code and carrier phase offsets. This is the phenomenon of multipath, which is evident as noise corrupting the observables. As the geometry between a stationary receiver and the satellite constellation repeats every sidereal day, so the multipath signature will repeat. A caveat here is that this is of course dependent on the local reflective environment remaining constant. This may not be the case in transient areas such as near parking lots or airfields [Klobuchar , 1996].

The multipath effect on ionospheric delay measurements is evident in the results shown in Figure 5.1 for successive days of observation from the same site (IGS station Albert Head in British Columbia, Canada). There is a clear correlation between the high frequency components of the estimated constant coefficient a_0 of the surface fit pattern for successive days. It can therefore be assumed that this observed pattern has an influence on the ionospheric delay estimation process. The more obvious long period variation is a result of the diurnal change in ionospheric delay discussed in section 3.3.1.

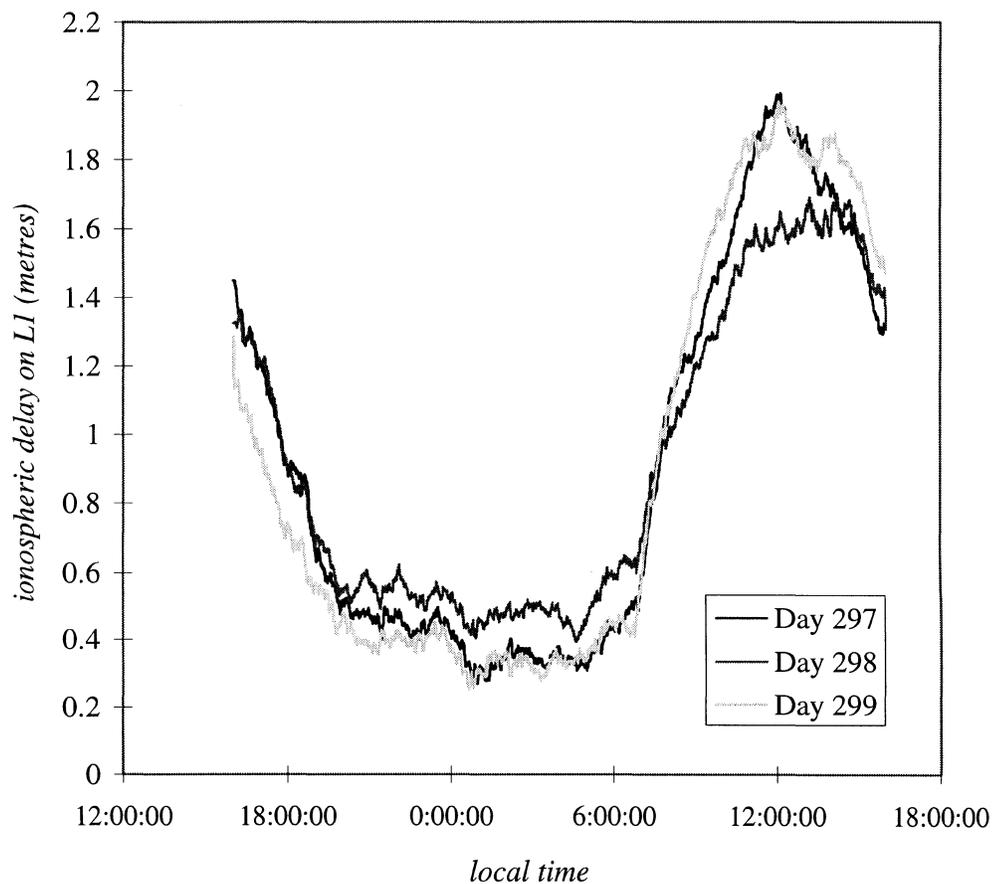


Figure 5.1 Constant coefficient of ionospheric delay surface fit for three consecutive days (297 - 299) in 1995.

Leick [1995] issues a warning regarding the modelling of ionospheric delay in the presence of strong multipath. The multipath effects cannot be ignored, and may be such that ambiguities cannot be resolved over even short baselines. Since the resolution of ambiguities is the main purpose of this thesis, this is an important point.

As previously mentioned, the data used in the initial testing of the software developed was collected from IGS stations, the sites of which should be chosen for optimal avoidance of such interference factors. Thus it was assumed that multipath effects induced are small enough to be safely ignored in initial tests. In real world environments however, multipath is less controllable, especially if the base station location is dictated by other factors, and more implicit modelling of this error source may be required.

The second stage of testing was carried out using data obtained from a less permanent base station site, the local environment of which is unknown in terms of multipath. Analysis of the results obtained on processing this second set of data for the effects of multipath was not as simple since successive days of observation data were unavailable. It may be presumed however that multipath is indeed a source of noise, the major effects of which are hopefully removed in the filtering of the surface coefficients. Looking at the epoch to epoch variability of the estimated surface fit coefficients and comparing between data sets gives an indication of the relative level of multipath induced noise.

5.2 Variability in Surface Fit Coefficients

The three coefficients of the ionospheric delay surface fit are governed by various spatial and temporal factors which were introduced in Chapter Three. Analysis of the filter output was undertaken, the results being summarised below.

5.2.1 Temporal and Spatial Changes in Estimated Ionospheric Delay

Again looking at plots of the constant coefficient of the surface fit, here realised as a time series over the same three days at Albert Head, the expected long period variation due to diurnal changes in the ionosphere is readily evident [see Figure 5.2]. The peak value of maximum ionization occurs just after noon local time as predicted [cf. section 3.3.1].

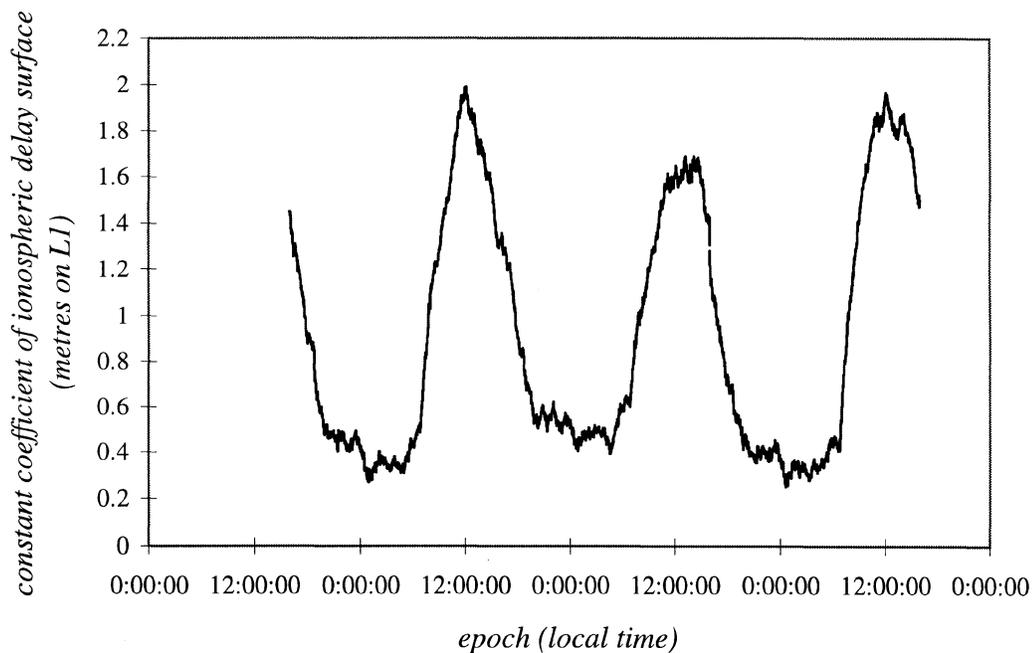


Figure 5.2 Constant coefficient of ionospheric delay surface fit for three consecutive days (297 - 299) in 1995.

Shorter term variations in the estimated ionospheric delay are caused by multipath, as discussed earlier, and small scale fluctuations in the TEC. Such TEC fluctuations will most often occur at night [Doherty et al., 1996]. These short term irregularities or scintillations are most prevalent in the equatorial and auroral regions of the ionosphere, auroral scintillations being both larger and longer lasting. Since the data used in these initial investigations was for mid-latitude stations, the effect of these scintillations is not readily apparent.

Temporal changes in the a_1 and a_2 coefficients are more subtle, and few conclusions were drawn from this initial investigation. It is suggested that further work in this area would include an investigation into the variation of the longitude and latitude based coefficients with time. This point is further expanded upon in the final chapter of this thesis.

Referring to the description given earlier (section 3.4) of latitudinal variation in the ionosphere, it may be predicted that the a_1 coefficient describing latitudinal variability in the estimated surface will be greater than the a_2 coefficient. Figure 5.3 shows a typical set of coefficients for day 298 of 1995 at Albert Head in which this predicted behaviour may be seen. The ionospheric delay values so given are then consistent with predictions in Figure 3.8 [Klobuchar, 1996].

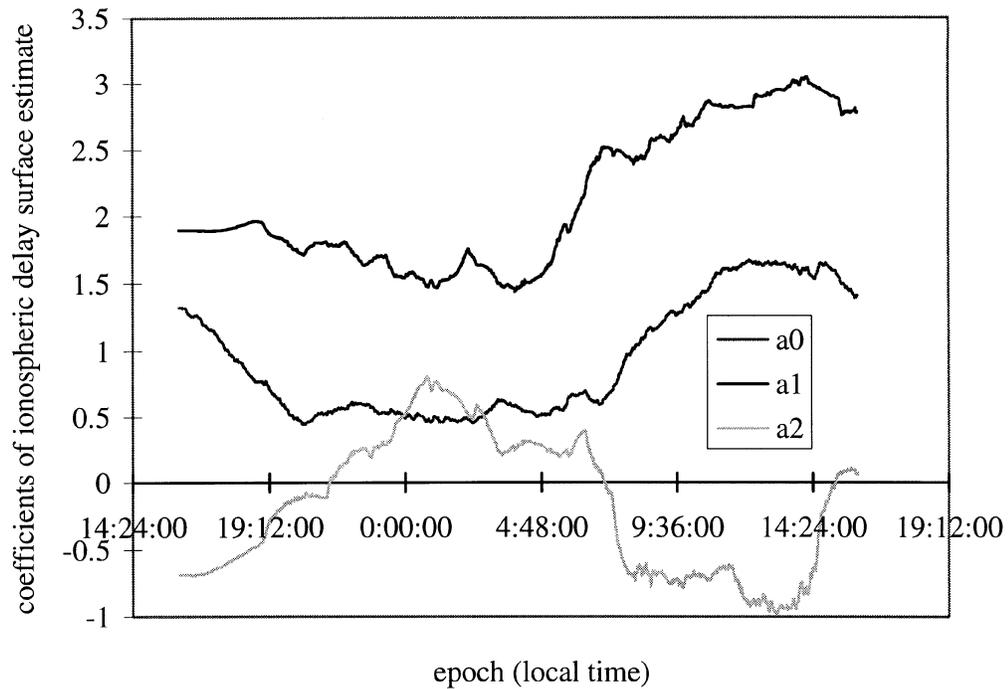


Figure 5.3 Surface fit coefficients for day 298 of 1995 at Albert Head, B.C. Canada. The a_0 coefficient is in units of metres, a_1 and a_2 in units of metres/radian.

The diurnal variation of the a_0 coefficient has been discussed in the previous section. The latitude dependent a_1 coefficient also has a significant diurnal variation suggesting that the modelled spatial change increases toward midday, the slab thickness increasing toward the equator more rapidly at the time of maximum ionization. The value of the a_1 coefficient is consistently positive, creating an ionospheric slab thickness at any one epoch which increases toward the equator.

The a_2 coefficient describing longitudinal change in slab thickness is a function of earth rotation. A component of the diurnal change in ionospheric delay is noticeable in the

estimate; the a_2 coefficient being most positive at local midnight and most negative at the time of maximum ionization, passing through zero at local dawn and dusk.

The corrections applied to reference and rover observations are a function of receiver relative position. With reference to equation (4.4), typical values of the latitude and longitude dependent influence are computed from the relevant coefficient multiplied by the difference in latitude or longitude of the IPP with respect to the reference receiver position. Thus if the IPP is 0.15 radians distant in latitude (a typical limiting value) then the latitude dependent effect will be approximately 0.3m for $a_1 = 2$ m/rad.

5.3 Preliminary Results

The data used to assess the success of the ionospheric modelling technique described here was collected in March of 1995 as part of an investigation into the effects of ice build up on aeroplanes, named Frizzle '95 [Mendes et al., 1995]. A number of missions were flown off the coast of Newfoundland, the particular data set used here chosen since the plane plots a near constant course away from the reference station, providing ideal conditions for the analysis of the impact of the ionospheric corrections for increasing baseline lengths. Receivers used both on the aircraft and at the reference station were Ashtech Z-XIIs, providing dual frequency pseudorange and carrier phase data.

5.3.1 Results from Frizzle '95 data

The strategy employed involved attempting to initialise ambiguity resolution at successively longer baselines, noting the differences in results from corrected and uncorrected data..

5.3.1.1 Ionospheric Delay Corrections

The thickness of the ionospheric shell is estimated via the three coefficients described in the previous chapter. Figure 5.4 is a representation of the filtered surface fit coefficients estimated for this particular flight. Spatial variation in slab thickness is at a maximum in the latitude direction, a typical difference being of the order of 50cm of vertical delay for an IPP located 700km distant from the reference station. As intimated earlier, variability in longitude is not so great over the distances observed here. Ionospheric pierce points observed ranged to approximately 700km from that point directly overhead the base station. Absolute values of ionospheric delay inferred by the surface fit procedure are in the 1 - 2m range, consistent with predictions of ionospheric delay for this time of relatively low solar activity, as shown in Figure 3.8 [Klobuchar, 1996].

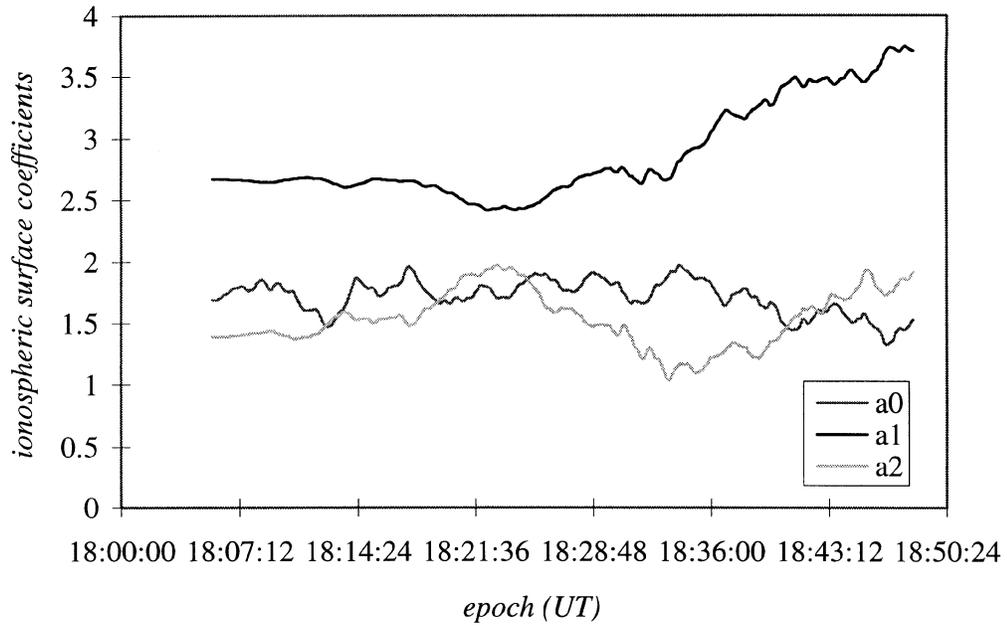


Figure 5.4 Filtered coefficients for reference station used in Frizzle '95 data collection.

Figure 5.5 below shows clearly how the *difference* in estimated ionospheric delay between stations is a function of the station separation. As the baseline increases so the difference in observed elevation angle of a particular satellite and the location of the respective IPPs will diverge. Thus the line of sight ionospheric delay estimated by the process outlined in Chapter Four is shown to be approximately proportional to baseline length.

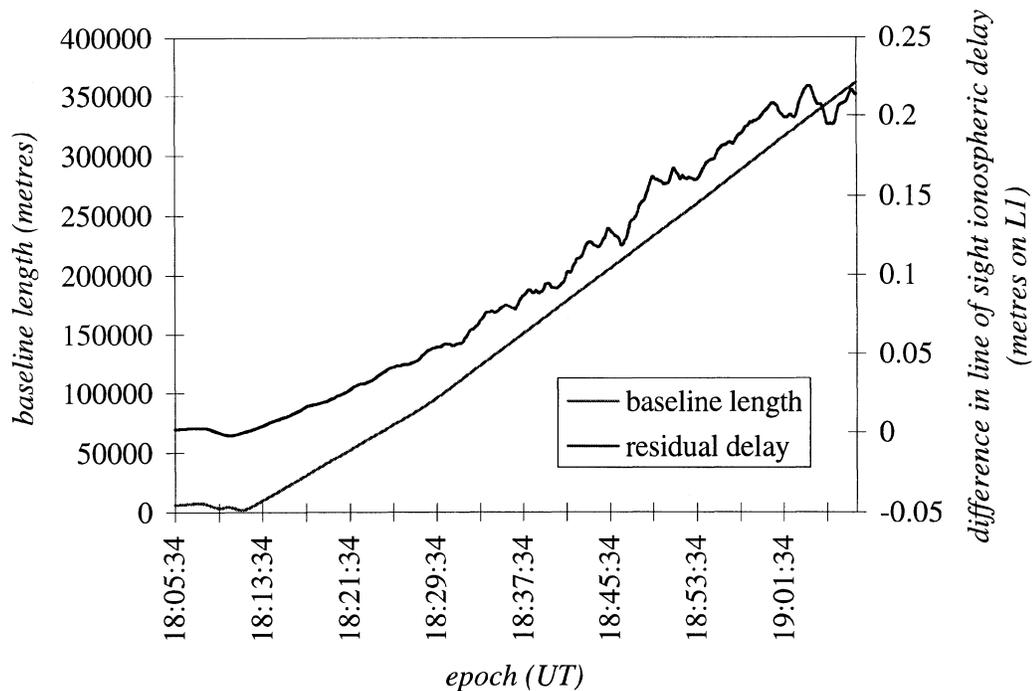


Figure 5.5 Difference in line of sight ionospheric delay estimated with all coefficients for observations to SV#29 from reference and rover receivers. This residual delay is shown to increase with baseline length.

It is then perhaps predictable that the effect on the ambiguity resolution process will be negligible for short baselines, and will increase in significance as the station separation increases.

5.3.1.2 Ambiguity Resolution Results

Comparison of the results obtained on attempting ambiguity resolution for corrected and uncorrected data produced some interesting results. The following figures show the double differenced ambiguity values resolved for a number of different processing options for each satellite observed. Both the original and corrected data were run through the KARS software [see section 4.6], and a set of resolved ambiguities provided on

attempting initiation at one minute intervals. Thus a time series of values was produced, the results of which are summarised below.

Comparing the original (*uncorrected*) with the corrected data shows that the ambiguities resolved at short (<15km) baselines are identical for both. This is consistent with the results predicted in the previous section where the *difference* in ionospheric delay correction applied at base and rover was shown to be proportional to baseline length. Thus, since this difference is small for short baselines, then on performing the double differencing operation in the ambiguity resolution process the correction will all but cancel out. If the ambiguity values resolved at such short baselines are considered correct, a reasonable assumption given the near ideal conditions over these distances, then obtaining these same integer values for each satellite at longer baselines would suggest that such values continue to provide a reliable solution. It was on this premise that the validity of solutions obtained over long distances was evaluated.

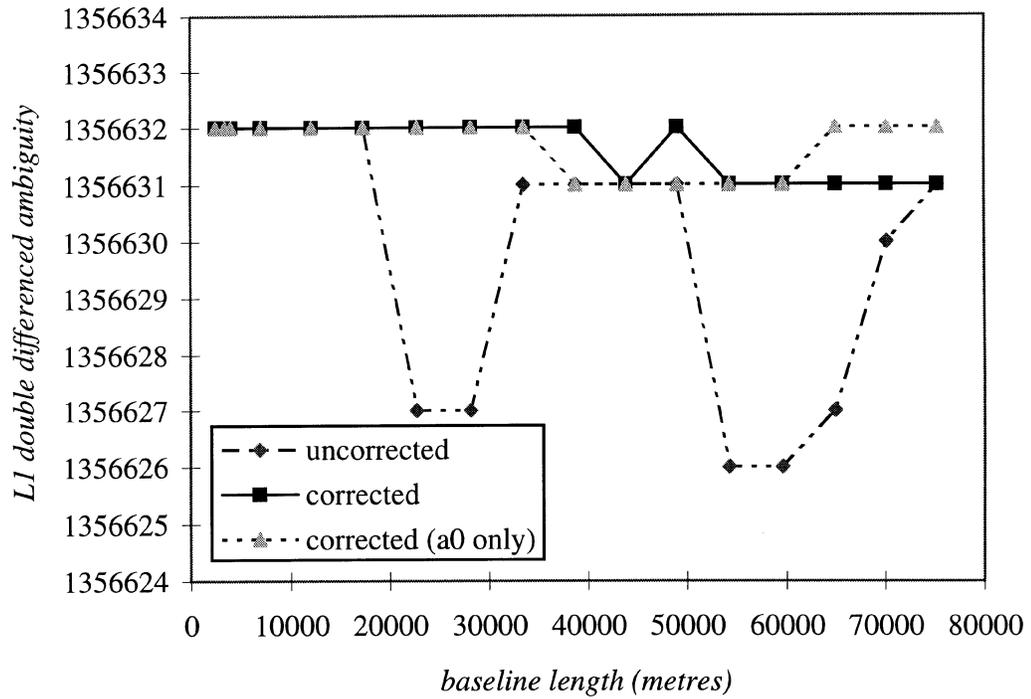


Figure 5.6 Resolved double differenced ambiguity values for SV#17

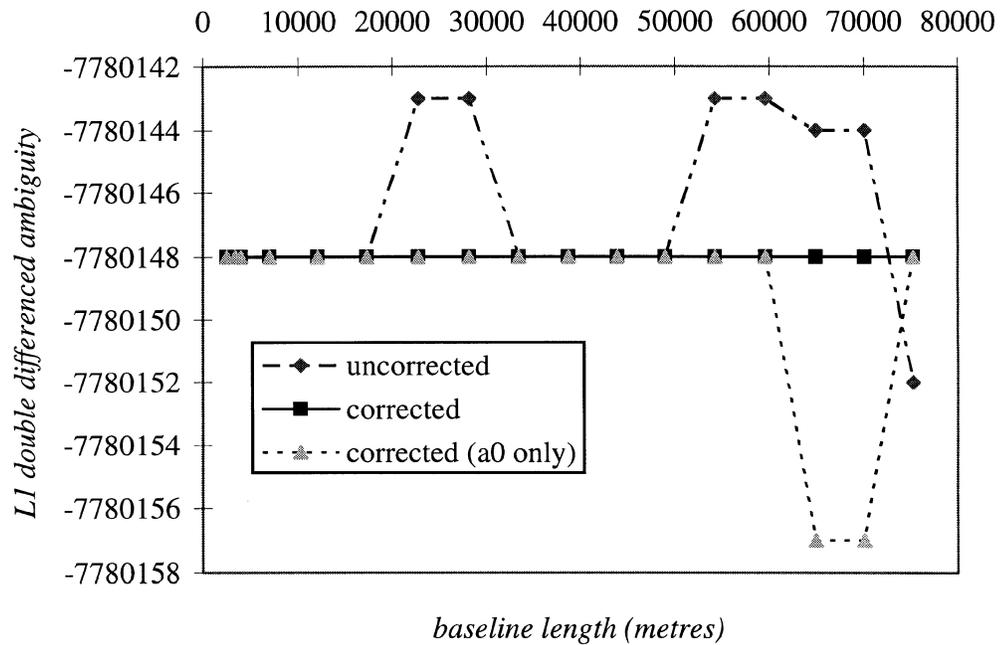


Figure 5.7 Resolved double differenced ambiguity values for SV#18.

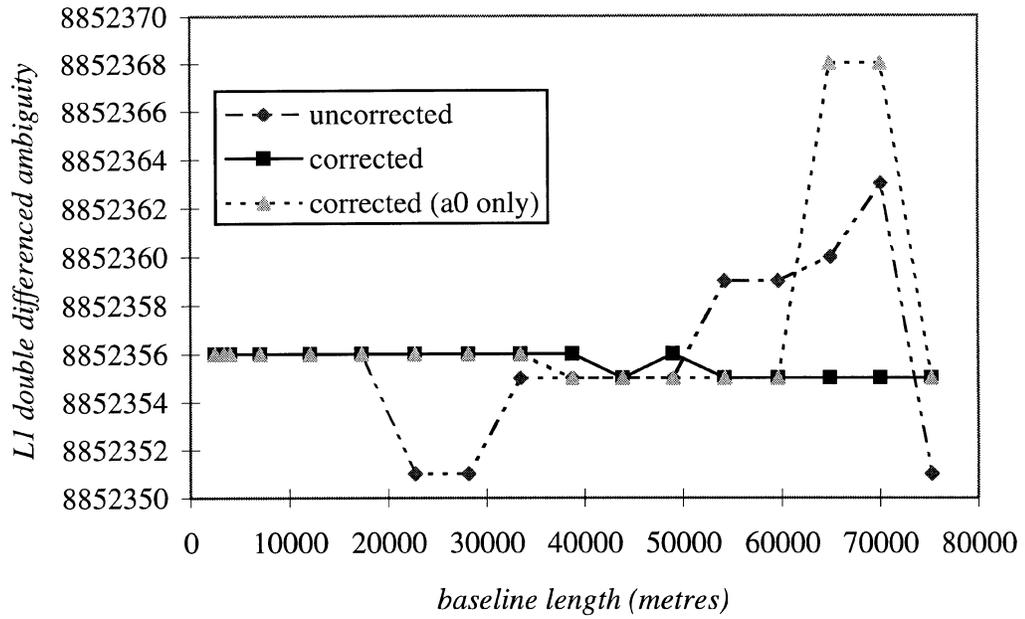


Figure 5.8 Resolved double differenced ambiguities for SV#25.

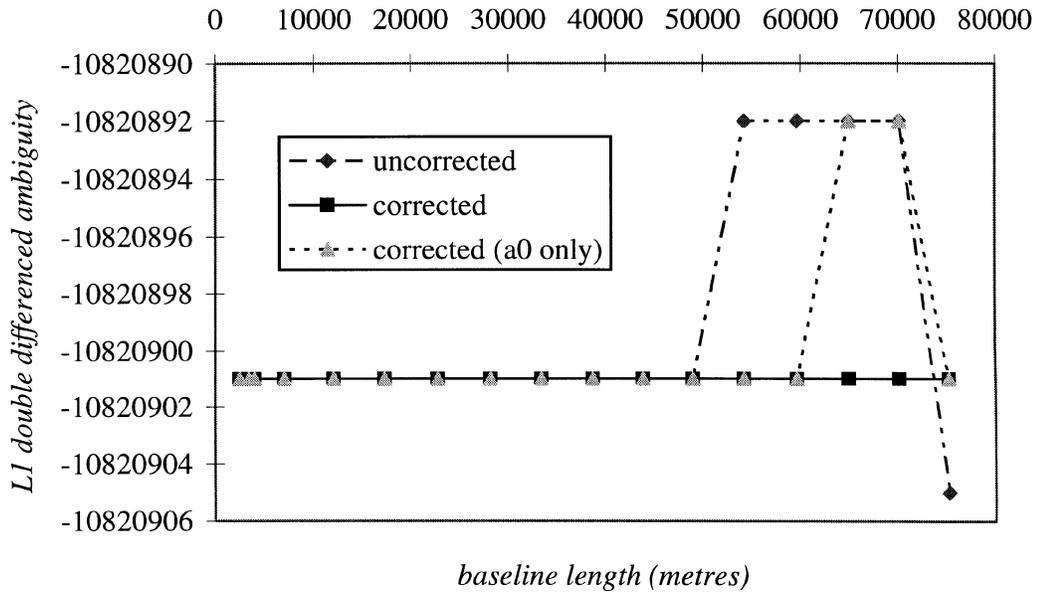


Figure 5.9 Resolved double differenced ambiguity values for SV#28.

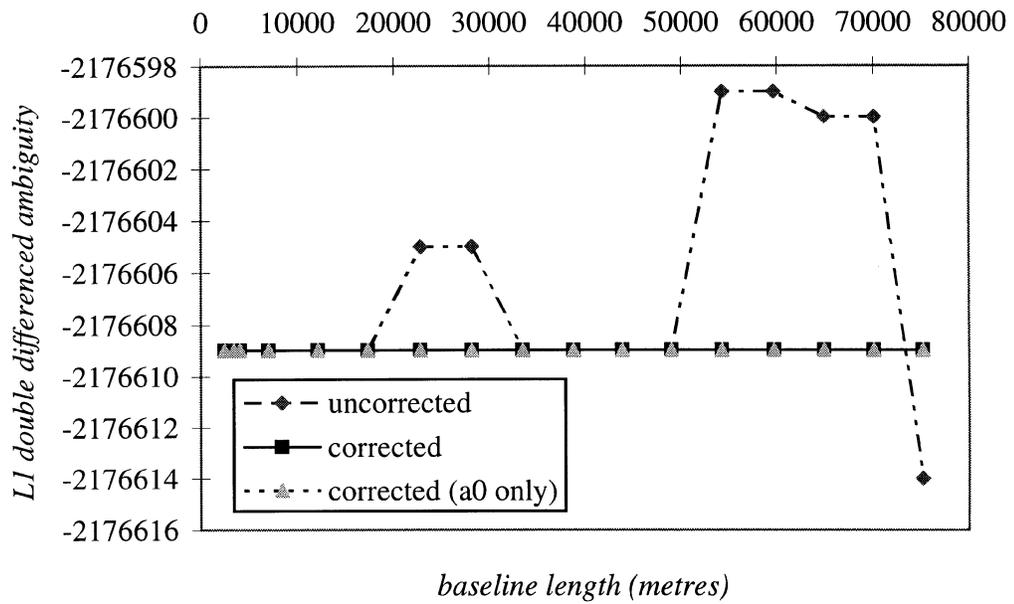


Figure 5.10 Resolved double differenced ambiguity values for SV#29.

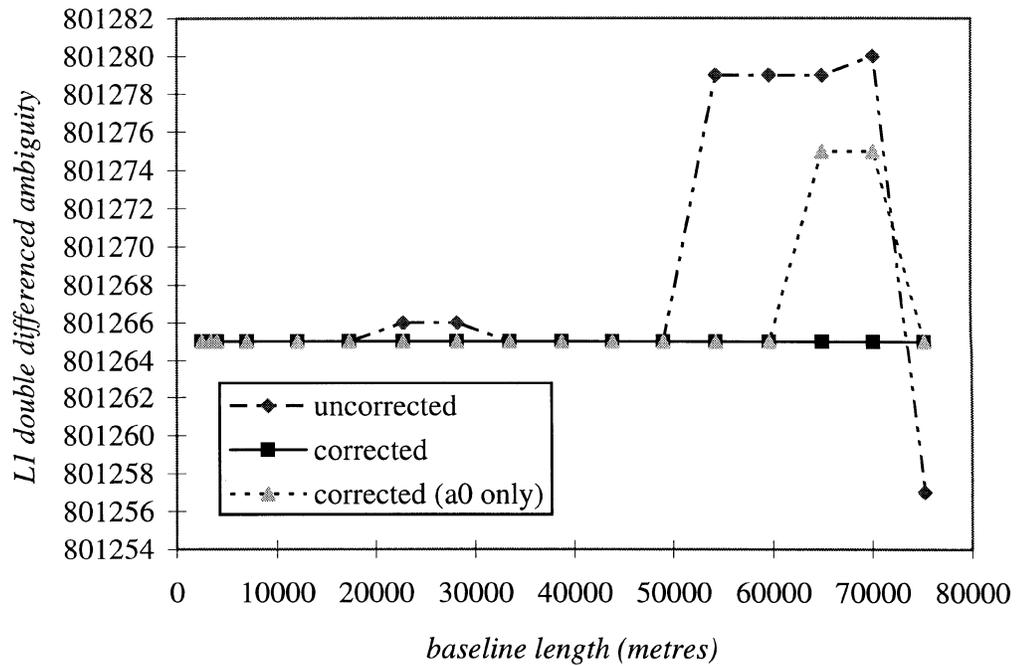


Figure 5.11 Resolved double differenced ambiguity values for SV#31.

As the station separation increases, so the difference in ionospheric delay increases, and the effect on ambiguity resolution is noticeable. Looking at Figure 5.6 as an example, the ambiguity values resolved remain constant over longer baselines for the corrected data. The uncorrected solution becomes unstable at a baseline length of approximately 15km, exhibiting changes of the order of 5 cycles. This pattern is repeated in the resolved L2 values, which are not shown here.

Now looking at the results from the corrected data, the values resolved may be considered reliable to approximately 40km.

It was also proposed to look into the effect of correcting the base and rover data with only the spatially constant part of the estimated delay. Thus only the a_0 coefficient was used, and the ambiguities resolved plotted in the above figures. The results show a pattern consistent with that of the data corrected with all coefficients for up to 35km baselines. Beyond this threshold, the ambiguity values obtained are less consistent, especially for satellites 18, 25, 28 and 31.

Thus it may be conjectured that the spatial modelling of ionospheric delay shows an improvement both over the original uncorrected data, and the data corrected with a values computed from a spatially constant slab thickness.

The quality of these resolved ambiguities may be investigated by looking at the double differenced carrier phase residuals for corrected and uncorrected data. Figure 5.12 shows the r.m.s. carrier phase residual for all satellites observed for the original data and that corrected for ionospheric delay. It would be expected that at longer baselines this residual would be smaller for the corrected data since most of the differential ionospheric delay effect has been modelled and removed.

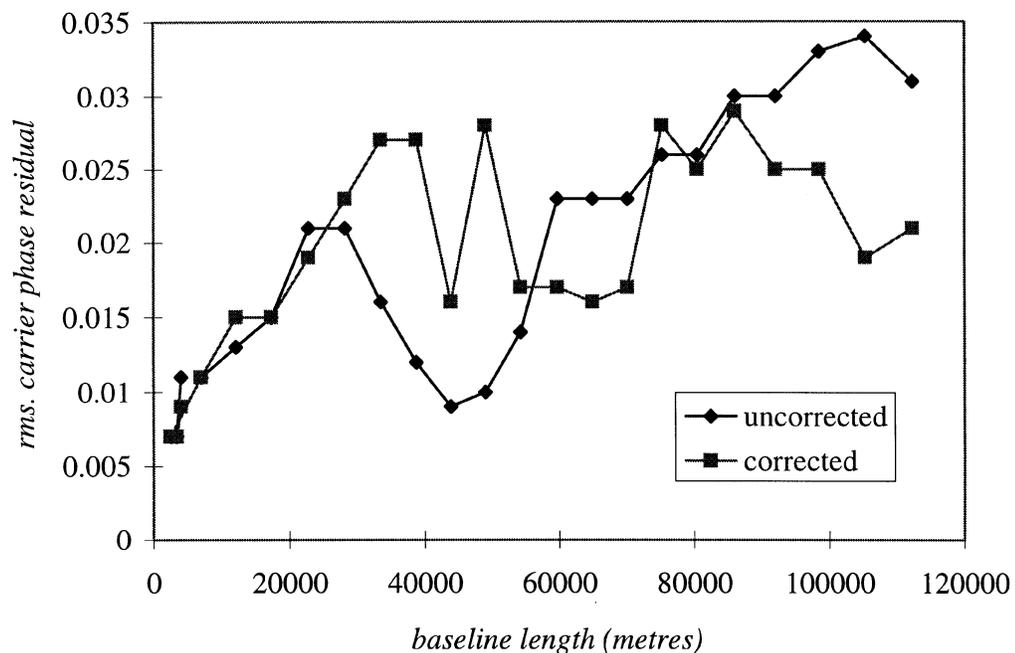


Figure 5.12 R.m.s. carrier phase residuals at time of ambiguity resolution plotted against baseline length.

The general trend apparent from Figure 5.12 is an increase in the carrier phase residual with baseline length. A comparison of the results for uncorrected data shows that for longer baselines, the residuals do indeed appear smaller for the corrected data. Residuals calculated for short baseline ambiguity resolution show similar results for both corrected

and uncorrected data since the difference in estimated ionospheric delay is extremely small at these distances. There is however an unexplained anomaly at medium (25 - 55km) distances where residuals for the corrected data are significantly larger than those for the uncorrected observations.

6 Chapter Six

Conclusions and Recommendations

This chapter provides a summary of the work carried out in the course of this thesis, along with a number of concluding remarks and suggestions for the furtherance of the ideas proposed.

6.1 Summary

Modelling of the local ionosphere around a fixed reference station has been attempted, producing a set of coefficients which describe a polynomial in latitude and longitude differences from the base. Latitude and longitude values here refer to those of the points at which the satellite signal intercepts the thin shell model of the ionosphere at the assumed height of maximum ionization. Computing the ionospheric delay values from this estimated slab model for each satellite observed at the base and rover receiver, and applying these values to the observations then significantly reduces the effect of the differential ionosphere. This is an important limiting factor for the resolution of carrier phase integer ambiguity values over long baselines ($> 15\text{km}$) since spatial decorrelation of such delays means that the double differencing operation will leave significant residual error.

Evaluation of the influence of this ionospheric delay modelling approach was undertaken by comparison of the results of attempts to initialise integer ambiguities over successively longer baselines. Beginning the ambiguity resolution process at 1 minute intervals as the aircraft containing the rover receiver travelled away from the reference station provided an ideal measure of the success or failure.

6.2 Conclusions

It appears that the ionospheric delay surface model has had a significant effect on the apparent range of ambiguity resolution. Baselines of greater than 35km were resolved within a few epochs once the ionospheric corrections were applied to remove much of the differential ionospheric effect. This compares with an apparent limit of less than 20km for the reliable initialisation of integer ambiguities using the same ambiguity resolution algorithm on uncorrected data.

Using a model of the ionospheric shell in which the slab thickness was held spatially constant made for a more reliable solution at longer baselines. The most consistent results were however obtained on modelling this spatial ionospheric delay variation. Using the delay values computed from the three coefficient estimated ionospheric delay surface allowed consistent resolution of ambiguities to baselines of the order of 40km.

Thus it may be suggested that the procedure described in this thesis has promise as an approach for the reduction of the differential ionospheric effect. It should be mentioned however that the work undertaken here describes only preliminary research, and a number of improvements should be made before a more general application is attempted.

6.3 Recommendations for Future Related Research

It was stated earlier that the work undertaken here was to serve as an initial investigation into the worth of this ionospheric surface estimation approach. Thus it is perhaps predictable that some additional work has been identified. A number of such suggestions are outlined below:

- A more rigorous approach to the calculation of the noise inputs to the Kalman filter should be attempted. This would allow a more generalised algorithm. The current approach produced a rather *ad hoc* set of process and measurement noise values by changing the inputs until a suitable output was obtained. External input sources such as an indication of the prevalent solar and geomagnetic activity indices could be used as a source of possible process noise values.
- The filter update rate is currently set as identical to the measurement rate. Thus a smoothing of the ionospheric surface estimates is carried out. Analysis of the temporal stability of the estimated coefficients would allow the filter update rate to

be increased. This would allow a reduction in the data required to be transmitted to the rover station in a real world situation.

- The position of the rover receiver to be used for the definition of the centre of the search space in the ambiguity resolution process may be better calculated from a current or predicted carrier phase based position solution.
- Further investigations into the spatial and temporal variations of the ionospheric coefficients would provide greater insight into the local ionospheric make-up. Considering the longitude dependent coefficient of the surface fit, it is suggested that the following pattern may be observable. At local dawn, the ionosphere in the direction of increasing time (and therefore decreasing longitude) would be expected to show greater ionization, with a tail off in the direction of increasing longitude. The opposite may be observable at local dusk, and a “tent shape” noticeable at local noon.
- Reduction of the effects of multipath could be reduced on spectral analysis of a number of days of data from the same site, allowing the frequency domain signature of this inherently localised multipath model to be built into the filtering of the coefficients.

References

- Abidin, H.A. (1993). *Computational and Geometrical Aspects of On-the-Fly Ambiguity Resolution*. Ph.D. dissertation, Department of Surveying Engineering Technical Report No. 164, University of New Brunswick, Fredericton, New Brunswick, Canada, 314 pp.
- Ashtech, Inc. (1994). *PNAV Software Users Guide*. Document Number 600248, Revision A, Ashtech, Inc. Sunnyvale, California.
- Bassiri, Sassan and George A. Hajj (1993). "Higher-order Ionospheric Effects on the Global Positioning System and Means of Modeling Them." *Manuscripta Geodaetica*, Vol. 13, pp. 280 - 289.
- Bishop, G.J. and J.A. Klobuchar (1990). "Ranging Errors due to Disturbances in the Polar Ionosphere." *Proceedings of ION-GPS 1990, The Third International Technical Meeting of the Satellite Division of the Institute of Navigation*, Colorado Springs, CO, Sept. 19 - 21, pp. 175 - 179.
- Blomenhofer, Helmut, Gunter W. Hein and David Walsh (1993). "On-The-Fly Carrier Phase Ambiguity Resolution for Precise Aircraft Landing." *Proceedings of ION-GPS 1993, The Sixth International Technical Meeting of The Satellite Division of The Institute of Navigation*, Salt Lake City, Utah, September 22-24, pp. 821-830.
- Braasch, Michael S. (1996). "Multipath Effects." In *Global Positioning System: Theory and Applications*, Eds. Bradford W. Parkinson and James J. Spilker Jr., Volume 163 of *Progress in Astronautics and Aeronautics*, Ed. Paul Zarchan, American Institute of Aeronautics and Astronautics, Inc., Washington DC, Vol. I, pp. 547 - 568.
- Brown, Robert Grover and Patrick Y.C. Hwang (1992). *Introduction to Random Signals and Applied Kalman Filtering*, John Wiley and Sons, New York.
- Brunner, Fritz K. and Walter M. Welsch (1993). "Effect of the Troposphere on GPS Measurements." *GPS World*, Vol. 4, No. 1, pp. 42 - 51
- Chao, Y.C., Y.J. Tsai, J. Evans, C. Kee, T. Walter, P.K. Enge, J.D. Powell, B.W. Parkinson (1996). "Generation of Ionospheric Correction and Confidence Estimates for WAAS." *Proceedings of the Institute of Navigation Annual Meeting*, Cambridge, MA, June 19 - 21, pp. 139 - 146.

- Chao, Yi-chung, Yeou-Ji Tsai, Todd Walter, Changdon Kee, Per Enge and Brad Parkinson (1995). "An Algorithm for Inter-Frequency Bias Calibration and Application to WAAS Ionosphere Modeling." *Proceedings of ION-GPS 1995, The Eighth International Technical Meeting of The Satellite Division of The Institute of Navigation*, Palm Springs, CA, Sept. 12 - 15, pp. 639 - 646.
- Chavin, Steven (1996). "Ionospheric Specification for the Wide Area Augmentation System (WAAS) Simulation Studies." *Proceedings of ION-GPS 1996, The Ninth International Technical Meeting of The Satellite Division of The Institute of Navigation*, Kansas City, MS, Sept. 17 - 20, pp. 585 - 594.
- Counselman, C.C. and S.A. Gourevitch (1981). "Miniature interferometric terminals for earth surveying: ambiguity and multipath with Global Positioning System." *IEEE Trans. Geosci. Remote Sens.*, Vol. GE-19, No. 4, pp. 244-252.
- Daniel, Robert E., Lincoln D. Brown and Robert W. Simon (1996a). "Modeling Low Latitude Ionospheric Dispersion." *Proceedings of the Institute of Navigation Annual Meeting*, Cambridge, MA, June 19 - 21, pp. 567 - 572.
- Daniel, Robert E., Lincoln D. Brown and Robert W. Simon (1996b). "A New, Improved Ionospheric Correction Algorithm for Single Frequency GPS Receivers." *Proceedings of ION-GPS 1996, The Ninth International Technical Meeting of The Satellite Division of The Institute of Navigation*, Kansas City, MS, Sept. 17 - 20, pp. 635 - 640.
- Dedes, George and Angela Mallett (1995). "Effects of the Ionosphere and Cycle Slips in Long Baseline Dynamic Positioning." *Proceedings of ION-GPS 1995, The Eighth International Technical Meeting of the Satellite Division of the Institute of Navigation*, Palm Springs, CA, Sept. 12 - 15, pp. 1081 - 1090.
- Doherty, Patricia, Eric Raffi, John Klobuchar and M. Bakry El-Arini (1994). "Statistics of Time Rate of Change of Ionospheric Range Delay." *Proceedings of ION-GPS 1994, The Seventh International Technical Meeting of The Satellite Division of The Institute of Navigation*, Salt Lake City, UT, Sept. 20 - 23, pp. 1589 - 1598.
- Fraile-Ordóñez, José (1995). "Real-Time TEC Determination for Ionospheric Modeling in WADGPS." *Proceedings of ION-GPS 1995, The Eighth International Technical Meeting of The Satellite Division of The Institute of Navigation*, Palm Springs, CA, Sept. 12 - 15, pp. 1193 - 1197.
- Friedman, Herbert (1985). *Sun and Earth*. W.H Freeman, New York.
- Georgiadou, Yola (1994). "Modelling the Ionosphere for an Active Control Network of GPS Stations." Report to the Faculty of Geodetic Engineering, No. 7, Technical University of Delft, Delft, The Netherlands.

- Georgiadou, Yola and Alfred Kleusberg (1988). "On the Effect of Ionospheric Delay on Geodetic Relative GPS Positioning." *Manuscripta Geodaetica*, Vol. 13, pp. 1 - 8.
- Gourevitch, Sergei, and John M. Nolan (1993). "Tracking Results: Implications to Precision Civilian Applications." *Proceedings of ION-GPS 1993, The Sixth International Technical Meeting of The Satellite Division of The Institute of Navigation*, Salt Lake City, Utah, September 22-24, pp. 1021-1025.
- Håkegård, Ole P. (1994). "A Real Time Ionospheric Model for Use in a WADGPS System." *Proceedings of the Institute of Navigation National Technical Meeting*, San Diego, CA, Jan. 24 - 26, pp. 921 - 926.
- Hargreaves, J.K. (1992). *The Solar Terrestrial Environment*. Volume 5 of *Cambridge Atmospheric and Space Science Series*, Eds. John T. Houghton, Michael J. Rycroft and Alexander J. Dessler, Cambridge University Press, Cambridge
- Hartmann, G.K. and Leitinger, R. (1984). "Range Errors Due to Ionospheric and Tropospheric Effects for Signal Frequencies Above 100 MHz." *Bulletin Géodésique*, Vol. 58, pp. 109 - 136.
- Janes, H.W., R.B. Langley and S.P. Newby (1991). "Analysis of Tropospheric Delay Prediction Models." *Bulletin Géodésique*, Vol. 65, No. 3, pp. 151 - 161.
- Kleusberg, Alfred and Peter J.G. Teunissen (1996). "GPS Observation Equations and Positioning Concepts." In *GPS for Geodesy*, Eds. Alfred Kleusberg and Peter J.G. Teunissen, Volume 60 of *Lecture Notes in Earth Sciences*, Eds. S. Bhattacharji, G.M. Friedman, H.J. Neugebauer, A. Seilacher, Springer Verlag, Berlin, Heidelberg, New York.
- Klobuchar, J.A. (1996). "Ionospheric effects on GPS." In *Global Positioning System: Theory and Applications*, Eds. Bradford W. Parkinson and James J. Spilker Jr., Volume 163 of *Progress in Astronautics and Aeronautics*, Ed. Paul Zarchan, American Institute of Aeronautics and Astronautics, Inc., Washington DC, Vol. I, pp. 485 - 516.
- Klobuchar, J.A., P.H. Doherty and M. Bakry El-Arini (1995). "Potential Ionospheric Limitations of the GPS Wide Area Augmentation System (WAAS)." *Navigation*, Vol. 42, No. 2, pp. 353 - 370.
- Klobuchar, John A. (1987). "Ionospheric Time Delay Algorithm for Single Frequency GPS Users." *IEEE Transactions on Aerospace and Electronic Systems*, Vol. AES-23, No. 3, pp. 325 - 331.
- Komjathy, A. and R.B. Langley (1996). "The Effect of Shell Height on High Precision Ionospheric Modelling Using GPS." Paper presented at the IGS Workshop, Silver Spring, MD, 19 - 21 March.

- Lancaster, Peter and Kestutis Šalkauskas (1986). *Curve and Surface Fitting: An Introduction*. Academic Press, London.
- Langley, Richard B. (1996). "Propagation of the GPS signals." In *GPS for Geodesy*, Eds. Alfred Kleusberg and Peter J.G. Teunissen, Volume 60 of *Lecture Notes in Earth Sciences*, Eds. S. Bhattacharji, G.M. Friedman, H.J. Neugebauer, A. Seilacher, Springer Verlag, Berlin, Heidelberg, New York.
- Leick, Alfred (1995). *GPS Satellite Surveying*. 2nd Edition, John Wiley and Sons, New York.
- Mader, Gerald L. (1992). "Rapid Static and Kinematic Global Positioning System Solutions Using the Ambiguity Function Technique." *Journal of Geophysical Research*, Vol. 97, No. B3, pp. 3271-3283.
- Mader, Gerald L. (1995). "Kinematic and Rapid Static (KARS) GPS Positioning: Techniques and Recent Experiences." *GPS Trends in Precise Terrestrial, Airborne and Spaceborne Applications, IAG Symposium No. 115*, Eds. Gerhard Beutler, Günter W. Hein, William G. Melbourne and Günter Seeber, Boulder, CO, July 3-4, pp. 170-174.
- Mannucci, A.J., B.D. Wilson and C.D. Edwards (1993). "A New Method for Monitoring the Earth's Ionospheric Total Electron Content using the GPS Global Network." *Proceedings of ION-GPS 1993, The Sixth International Technical Meeting of The Satellite Division of The Institute of Navigation*, Salt Lake City, UT, Sept. 22 - 24, pp. 1323 - 1332.
- Mannucci, Anthony, Brian Wilson and Dah-Ning Yuan (1995). "An Improved Ionospheric Correction Method for Wide-Area Augmentation Systems." *Proceedings of ION-GPS 1995, The Eighth International Technical Meeting of the Satellite Division of the Institute of Navigation*, Palm Springs, CA, Sept. 12 - 15, pp. 1199 - 1208.
- McNamara, Leo F. (1991). *The Ionosphere: Communications, Surveillance and Direction Finding*. Orbit, A Foundation Series, Ed. Edwin F. Strother, Krieger Publishing Co., Malabar, Florida.
- Meehan, T.K., J.M. Srinivasan, D.J. Spitzmesser, C.E. Dunn, J.Y. Ten, J.B. Thomas, T.N. Munson, C.B. Duncan (1992). "The TurboRogue GPS Receiver." *Proceedings of the Sixth International Geodetic Symposium on Satellite Positioning*, Columbus, OH, 17-20 March, pp. 209 - 218.
- Mendes, V.B., J.P. Collins and R.B. Langley (1995). "The Effect of Tropospheric Propagation Delay Errors in Airborne GPS Precision Positioning." *Proceedings of ION-GPS 1995, The Eighth International Technical Meeting of the Satellite Division of the Institute of Navigation*, Palm Springs, CA, Sept. 12-15, pp. 1681 - 1689.

- Niell, A.E., A.J. Coster, F.S. Solheim, V.B. Mendes, P.C. Toor, R.B. Langley and C.A. Ruggles (1996). "Measurement of Water Vapor by GPS, WVR and Radiosonde." <http://gauss.gge.unb.ca/papers.pdf/evgapro2.pdf>, 10 December.
- Parkinson, Bradford W. (1996). "GPS Error Analysis." In *Global Positioning System: Theory and Applications*, Eds. Bradford W. Parkinson and James J. Spilker Jr., Volume 163 of *Progress in Astronautics and Aeronautics*, Ed. Paul Zarchan, American Institute of Aeronautics and Astronautics, Inc., Washington DC, Vol. I, pp. 469 - 483.
- Schwarz, Klaus-Peter (1987). "Kalman Filtering and Optimal Smoothing." In *Papers for the CISM Adjustment and Analysis Seminars*, 2nd ed., Ed. Edward J. Krakiwsky, The Canadian Institute of Surveying and Mapping, Ottawa, pp. 230 - 264.
- Spilker, J.J. (1996). "Tropospheric effects on GPS." In *Global Positioning System: Theory and Applications*, Eds. Bradford W. Parkinson and James J. Spilker Jr., Volume 163 of *Progress in Astronautics and Aeronautics*, Ed. Paul Zarchan, American Institute of Aeronautics and Astronautics, Inc., Washington DC, Vol. I, pp. 517 - 545.
- St-Pierre, Claude and Rock Santerre (1996). "Modélisation du Délai Ionosphérique Relatif en vue de la Détermination des Ambiguïtés de Phase GPS." Report submitted to VIASAT Géo-Technologie inc., for *Proposition OTF-GPS en temps réel pour la Garde côtière canadienne (Région des Laurentides)*, Centre de Recherche en Géomatique, Université Laval, Quebec, Canada, September.
- Stull, Craig and A.J. Van Dierendonck (1995). "Test Results of Wilcox Electric's Ionospheric Monitoring Network." *Proceedings of ION-GPS 1995, The Eighth International Technical Meeting of the Satellite Division of the Institute of Navigation*, Palm Springs, CA, Sept. 12 - 15, pp. 1219 - 1228.
- Teunissen, Peter J.G. (1996). "GPS Carrier Phase Ambiguity Fixing Concepts." In *GPS for Geodesy*, Eds. Alfred Kleusberg and Peter J.G. Teunissen, Volume 60 of *Lecture Notes in Earth Sciences*, Eds. S. Bhattacharji, G.M. Friedman, H.J. Neugebauer, A. Seilacher, Springer Verlag, Berlin, Heidelberg, New York.
- Tscherning C.C. and C.C. Goad (1985). "Correlation Between Time Dependent Variations of Doppler Determined Height and Sunspot Numbers." *Journal of Geophysical Research*, Vol. 90, No. B6, pp. 4589 - 4596.
- U.S. Department of Transportation (nd.) *Federal Aviation Administration Specification - Wide Area Augmentation System (WAAS)*, FAA-E-2892A, U.S. Department of Transportation, Washington D.C., 49 pp.

- van Graas, Frank and Michael S. Braasch (1996). "Selective Availability." In *Global Positioning System: Theory and Applications*, Eds. Bradford W. Parkinson and James J. Spilker Jr., Volume 163 of *Progress in Astronautics and Aeronautics*, Ed. Paul Zarchan, American Institute of Aeronautics and Astronautics, Inc., Washington DC, Vol. I, pp. 601 - 621.
- Wanninger, Lambert (1995). "Improved Ambiguity Resolution by Regional Differential Modeling of the Ionosphere." *Proceedings of ION-GPS 1995, The Eighth International Technical Meeting of the Satellite Division of the Institute of Navigation*, Palm Springs, CA, Sept. 12 - 15, pp. 55-62.
- Webster, I.R. (1993). *A Regional Model for the Prediction of Ionospheric Delay for Single Frequency Users of the Global Positioning System*. M.Sc.E. thesis, Department of Surveying Engineering Technical Report No. 166, University of New Brunswick, Fredericton, New Brunswick, Canada, 124 pp.
- Weigel, Gary (1989). "Correlating Solar Activity with Uncompensated Ionospheric Effects on Doppler Determined Ellipsoid Heights." *Proceedings of the Fifth International Geodetic Symposium on Satellite Positioning*, New Mexico State University, Las Cruces, NM, pp. 400 - 409.
- Wilson, Brian D. and Anthony J. Mannucci (1993). "Instrumental Biases in Ionospheric Measurements Derived from GPS Data." *Proceedings of ION-GPS 1993, The Sixth International Technical Meeting of The Satellite Division of The Institute of Navigation*, Salt Lake City, UT, Sept. 22 - 24, pp. 1343 - 1351.
- Wu, B., P. Nicolaidis, T.N. Upadhyay, T.R. Jenkins (1996). "Ionospheric Error Compensation for GPS Receivers using Real-Time Ionospheric Model." *Proceedings of ION-GPS 1996, The Ninth International Technical Meeting of the Satellite Division of the Institute of Navigation*, Kansas City, MS, Sept. 17 - 20, pp. 575 - 583.

VITA

FULL NAME : Peter John Stewart

PLACE AND DATE OF BIRTH : Burnley
Lancashire, UK
20th May 1971

PERMANENT ADDRESS : 108 Drakies Avenue
Inverness
Scotland, UK
IV2 3SD

SCHOOLS ATTENDED : Elgin High School
High School Drive
Elgin, Moray
Scotland, UK

August 1983 - May 1989

UNIVERSITIES ATTENDED : Department of Geography and Topographic Science
University of Glasgow
Glasgow
Scotland, UK
G12 8QQ

1990 - 1994 B.Sc. (Hons.) Topographic Science

PUBLICATIONS

Kleusberg, Alfred, David Wells, Peter Stewart and John Hanley (1996). "The Design of GPS L1/L2 Carrier Phase Algorithms for Precision Position Determination." Contract Report for Applied Analytics Corporation, Markham, Ontario.
