Atmospheric Signal Delay Affecting GPS Measurements Made by Space Vehicles During Launch, Orbit and Reentry

by
Rachel Neville Thessin
B.S., California Institute of Technology (2003)
Submitted to the Department of Aeronautics and Astronautics
in partial fulfillment of the requirements for the degree of
Master of Science in Aeronautics and Astronautics
at the
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
June 2005
©2005 Rachel Neville Thessin. All rights reserved.
The author hereby grants to MIT permission to reproduce and to distribute publicly paper and electronic copies of this thesis document in whole or in part.

Author . . . . . .
Department of Aeronautics and Astronautics
May 20, 2005

Certified by . . . . . .
Anthony J. Bogner
Senior Member of the Technical Staff, Charles Stark Draper Laboratory
Thesis Supervisor

Certified by . . . . . .
Thomas A. Herring
Professor of Geophysics
Thesis Supervisor

Accepted by . . . . . .
Jaime Peraire
Professor of Aeronautics and Astronautics
Chair, Committee on Graduate Students

[This page intentionally left blank.]

Atmospheric Signal Delay Affecting GPS Measurements Made by Space Vehicles During Launch, Orbit and Reentry

by

Rachel Neville Thessin

Submitted to the Department of Aeronautics and Astronautics
on May 20, 2005, in partial fulfillment of the
requirements for the degree of
Master of Science in Aeronautics and Astronautics

Abstract

In this thesis, I present neutral atmosphere, ionosphere and total delays experienced by GPS signals traveling to space vehicles during launch, orbit and reentry. I calculate these delays for receivers at 0 km to 1700 km altitude by ray-tracing through the Global Reference Atmosphere Model (1999) and the International Reference Ionosphere (2001). These delays are potentially much larger than those experienced by signals traveling to GPS receivers near the surface of the Earth, but are primarily experienced at negative elevation angles, and are therefore most relevant for space vehicles with limited visibility of GPS satellites and during launch and reentry. I compare these signal delays to the delays predicted by three onboard delay models: the Altshuler and NATO neutral atmosphere delay models, and the Klobuchar ionosphere delay model. I find that these models are inadequate when the space vehicle is in orbit. The NATO model will suffice during the final period of reentry, where it predicts the neutral atmosphere delay to within 1 m of the ray-traced value, but it will not suffice when a satellite is rising or setting. I propose a method to extend the NATO model for receivers at higher altitudes. The Klobuchar model will suffice for most satellites during reentry, but will potentially predict ionosphere delays with errors up to 30 m, and will not suffice when a satellite is rising or setting. I find that a dual frequency GPS receiver will run into many problems if it is designed with near-surface use in mind. I examine "measure and hold" and "measure and propagate" dual frequency algorithms, which will have errors up to 30 m during orbit and up to 10 m during reentry. I propose a method by which to improve these algorithms for use in a GPS receiver aboard a space vehicle.

Thesis Supervisor: Anthony J. Bogner
Title: Senior Member of the Technical Staff, Charles Stark Draper Laboratory

Thesis Supervisor: Thomas A. Herring
Title: Professor of Geophysics

[This page intentionally left blank.]

Acknowledgments

Many people helped make this thesis possible. I would like to thank

- Tony Bogner, for giving me the freedom to pursue this topic as I saw fit while still guiding me toward relevant results
- Tom Herring, for lending his expertise and being willing to advise me from afar
- Tom Armstrong and Brian Kern, without whose help and support in the past I would have never reached the point where I could accomplish this thesis
- William Robertson, for his enthusiasm and extensive collection of journal articles
- Arnie Soltz, for input and advice
- Anil Rao, for always wondering how we were doing
- Alisa Hawkins and all the 5th floor A-core Draper Fellows
- Will Farr, for always being there for me

[This page intentionally left blank.]

Acknowledgment

May 20, 2005

This thesis was prepared at The Charles Stark Draper Laboratory, Inc. under contract N00030-04-C-0007, sponsored by the U.S. Navy.

Publication of this thesis does not constitute approval by Draper or the sponsoring agency of the findings or conclusions contained herein. It is published for the exchange and stimulation of ideas.

Rachel N. Thessin . . . . . .

[This page intentionally left blank.]

[This page intentionally left blank.]

[This page intentionally left blank.]

Contents

1 Introduction **29**

2 GPS Signal Propagation **33**

2.1 GPS Basics 33

2.2 Propagation in the Atmosphere 34

2.2.1 Dispersive vs. Non-Dispersive Media 34

2.2.2 Signal Delay 35

2.2.3 Signal Path 37

2.3 Refractivity in the Neutral Atmosphere 38

2.4 Refractivity in the Ionosphere 42

2.5 The Neutral Atmosphere 44

2.5.1 Ideal Gas Law Model 46

2.5.2 Variations in the Neutral Atmosphere 47

2.6 The Ionosphere 50

2.6.1 Layer Creation 52

2.6.2 Variations in the Ionosphere 52

2.7 Signal Propagation Features 55

2.7.1 Scintillation 55

2.7.2 Superrefraction and Ducting 56

3 Existing Methods to Correct GPS Signal Delay **59**

3.1 On-board Receiver Models 60

3.1.1 Models for the Ionosphere Signal Delay 61

3.1.2 Models for the Neutral Atmosphere Signal Delay 63

3.2 Dual Frequency Ionosphere Delay Correction 67

3.3 Single Frequency Ionosphere Delay Correction 68

3.4 Differential GPS Ionosphere Delay Correction 69

4 Ray-Tracing Model Options **71**
4.1 Neutral Atmosphere Model and Profile Options 72
4.1.1 Global Empirical Models 73
4.2 Refractivity Models of the Neutral Atmosphere 76
4.3 Ionosphere Model and Profile Options 77
4.3.1 Bent Ionospheric Model 78
4.3.2 International Reference Ionosphere (2001) 79

5 Implemented Ray-Tracing Approach **81**
5.1 Numerical Methods Used 81
5.1.1 Problem Set-up 82
5.1.2 Derivation of the Ordinary Differential Equations 83
5.1.3 Integrating the Initial Value Problem 84
5.1.4 Multidimensional Root Finder 85
5.1.5 Component Delay Calculations 88
5.1.6 Integration Implementation 90
5.2 Implementation of the Global Reference Atmosphere Model 1999 . . . 91
5.3 Implementation of the International Reference Ionosphere 2001 92

6 Ray-Traced GPS Signal Delays **93**
6.1 Delay as a Function of Elevation Angle and Receiver Altitude 94
6.1.1 Validation of Ray-Tracing Results 94
6.1.2 Typical Features 98
6.1.3 Conditions Examined 110
6.1.4 Delay as a Function of Season 110
6.1.5 Delay as a Function of Solar Activity 111
6.1.6 Delay as a Function of Time of Day 116
6.1.7 Delay as a Function of Latitude 116
6.1.8 Delay as a Function of Satellite Azimuth 119
6.2 Delay as Experienced by a GPS Receiver Aboard a Space Vehicle . . 126
6.2.1 Signal Delay 128
6.2.2 Delay Rates 133
6.2.3 Effect of Variables on Signal Delay Features 137
6.3 Error Analysis 144
6.4 Comparison of Ray-Traced Delay to Receiver Compensation Methods 147
6.4.1 Model Performance on Reentry 147

7 Correcting Signal Delay Aboard a Space Vehicle **157**
7.1 Correcting the Neutral Atmosphere Delay 157
7.1.1 Accept Existing Models 157
7.1.2 Develop New Model 158
7.2 Correcting the Ionosphere Delay 162
7.2.1 Modeling the Ionosphere Delay 163
7.2.2 Dual Frequency Correction of the Ionosphere Delay 164

8 Summary **175**
8.1 Aim 175
8.2 Results 175
8.3 Future Research 177

[This page intentionally left blank.]

List of Figures

1-1 A receiver on the surface of the Earth can only see satellites at positive elevation angles. (Receivers on the surface of the Earth can actually see satellites at very small negative elevation angles. Atmospheric refraction bends the signal slightly so that it curves with the surface of the Earth, allowing receivers to see satellites slightly below the horizon.) A receiver above the surface of the Earth, however, can see satellites at negative elevation angles. Figure is not to scale. 30

1-2 A signal from a GPS satellite at a negative elevation angle can potentially travel through the atmosphere twice: once on the "way in", and once on the "way out". Figure is not to scale. 31

1-3 As a spacecraft with a rear-mounted GPS antenna is reentering the atmosphere the spacecraft body and the Earth will block signals from most GPS satellites. Signals from visible satellites in region A will experience delay both on their way "into" the atmosphere and on their way "out". Signals from region B will only pass through the atmosphere on their way "in". Figure is not to scale. 32

2-1 Temperature, pressure and refractivity profiles from the GRAM-99 and U.S. Standard Atmosphere 1976 atmosphere models for noon on 3/21/80 at 28°N 80°W. Note that the shape of the pressure curve drives the shape of the refractivity curve. 45

2-2 The GRAM-99 temperature profile, which shows trends representative of the average atmosphere. TOP LEFT: Comparing the GRAM-99 profile with the U.S. Standard Atmosphere 1976 profile. TOP RIGHT: The GRAM-99 temperature profile does not change with time of day on a mid-latitude, spring day. BOTTOM LEFT: The GRAM-99 temperature profile varies greatly with latitude. Note the inversion layers at the higher latitudes. BOTTOM RIGHT: The GRAM-99 profile varies only slightly with season at the mid-latitudes. 49

2-3 Electron density and group refractivity curves during high solar activity and low solar activity periods at noon and at midnight, as specified by the International Reference Ionosphere Model. Profile located at 28°N 80°W. Layer markings are for the high solar activity, daytime curve. 51

2-4 The IRI-2001 electron density profile. TOP LEFT: Electron density on a high, medium and low solar activity, spring equinox day. TOP RIGHT: Electron density as a function of time of day. Not only does the density change, but so does the altitude of the F2-layer peak and the existence of the lower layers. BOTTOM LEFT: Electron density as a function of latitude. BOTTOM RIGHT: Electron density as a function of season. 54

2-5 Phase refractivity gradient in both the GRAM-99 and IRI-2000 models. Note that the GRAM has only standard refraction, and the IRI has both standard and subrefractive regions, due to the negative phase refractivity. 58

3-1 The zenith delay does not contain contributions from all of the layers of the atmosphere that the signal passes through. Figure is not to scale. 61

4-1 GRAM-99 water vapor pressure profile. 75

5-1 **Root-finder grid search.** Each dot represents one shot (i.e. one $\boldsymbol{v}$), and the star represents the root (the satellite). If the resulting $x(l = 1)$ from two successive shots are not within the tolerance ($5 \times 10^{-6} \times L$) but cross zero, the root-finder reduces the search region in the x-dimension and searches again. 87

5-2 An illustration of the problems in breaking the signal delay into components. The outer ring indicates the outer edge of the ionosphere and the inner ring indicates the outer edge of the neutral atmosphere. The receiver is at R and the GPS satellite at S. Figure is not to scale. . . 88

6-1 **Neutral Atmosphere Ray-Tracing Validation.** A comparison of neutral atmosphere signal delay ray-tracing results with the NATO and Altshuler delay models. Ray-tracing was done through the GRAM-99 atmosphere for a receiver on the ground at 28°N 80°W at noon on 3/21/80, and a satellite due east of the receiver. TOP: All elevation angles. BOTTOM: Low elevation angles. 96

6-2 **Ionosphere Ray-Tracing Validation.** TOP: A comparison of ionosphere signal delay ray-tracing results with both the Klobuchar delay model and the ray-traced zenith delay mapped by the geometrical mapping function presented in the lower plot. Ray-tracing was done through the IRI-2001 ionosphere for a receiver on the ground at 28N 80W at noon on 3/21/80, and a satellite due east of the receiver. The Klobuchar broadcast coefficients were for a high solar activity, spring equinox day, as presented in [1]. BOTTOM: A comparison of the Klobuchar mapping function to a geometrical mapping function that calculates the linear distance a straight line signal originating from the Earth's surface spends between 50 and 1050 km altitude for a given elevation angle, and divides this distance by 1000 km. (This normalizes the zenith mapping value to 1.) The Earth is assumed to be spherical, with a radius of 6,378 km. 97

6-3 Neutral atmosphere signal delay as a function of elevation angle for control conditions at six different receiver altitudes. LEFT: Close-in view. RIGHT: Full view. The delay curve for 1700 km is missing because the range of elevation angles at which there is any delay for a receiver at this altitude is less than the resolution of the ray-tracing (0.1° apparent elevation angle). 98

6-4 Geometrical effects on signal delay profile. LEFT: The geometrical mapping function that calculates the linear distance a straight line signal originating from the Earth's surface spends between 0 and 50 km altitude for a given elevation angle, and divides this distance by 50 km. (This normalizes the zenith mapping value to 1.) The Earth is assumed to be spherical, with a radius of 6,378 km. RIGHT: As a receiver gains in altitude, the range of elevation angles that the neutral atmosphere comprises shrinks and the horizon is located at lower and lower elevation angles. Figure is not to scale. 99

6-5 A sample horizon-grazing signal path. TOP: The altitude of the signal as it travels from the satellite to the receiver. Closest approach is ≈ 4 km. MIDDLE: The distance between the signal and the straight line connecting the satellite and the receiver in the direction toward the Earth. In order to reach the receiver, a signal must angle away from the straight line connecting the receiver and the satellite in the direction away from the center of the Earth. The bending that occurs when the signal is in the lowest layers of the atmosphere will then point the signal toward the receiver. Geometrically, this satellite is below the receiver's horizon since the signal deviates 60 km from the straight line and yet the altitude of closest approach is only 4 km. BOTTOM: Diagram of signal path. Figure is not to scale. 101

6-6 Receivers A, B and C should experience the same signal delay. Figure is not to scale. 102

6-7 A contour plot of neutral atmosphere signal delay as a function of receiver altitude and geometric elevation angle. The plot on the left examines the contour lines at the lower altitudes. Contour lines are at 1, 5, 10, 15, 20, 30, 50, 70, 100 m delay. Note how as altitude increases, the delay is reduced to a sliver of elevation angles. Also note that the delay at high receiver altitudes is missing because the range of elevation angles for which there is delay is less than the resolution of the ray-tracing (0.1° apparent elevation angle). The curves have some distortion at the low elevation angles due to the MATLAB interpolation routines used in the contour plotting and due to the resolution of the data input (every 25 km altitude for the higher altitudes). 103

6-8 Ionosphere signal delay as a function of elevation angle for control conditions at five different receiver altitudes. 104

6-9 Ionosphere signal delay increases as satellite elevation angle decreases, peaks when the signal is tangent to the densest region of the ionosphere, and then decreases until the satellite goes below the horizon. In this figure, darkness indicates a higher electron density. Note that while the signal from satellite C travels the longest distance in the ionosphere, it travels much of that distance in a region of low electron density. Of the three, the signal from satellite B travels the longest distance in the densest region of the ionosphere, and therefore experiences the largest ionosphere delay. Figure is not to scale. 105

6-10 For a receiver above the ionosphere, the ionosphere will fill a larger portion of the sky than the neutral atmosphere. There will therefore be ionosphere delay over a larger range of elevation angles (α) than there will be neutral atmosphere delay (β). Figure is not to scale. . . 106

6-11 A contour plot of ionosphere signal delay as a function of altitude and geometric elevation angle. Contour lines are at 1, 5, 10, 15, 20, 30, 50, 70, 100 m delay. Note that at higher altitudes, the data resolution decreases, causing the bumpiness of the 1 m contour line. As with the neutral atmosphere delay, the curves have some distortion at the low elevation angles due to the MATLAB interpolation routines used in the contour plotting. 107

6-12 Total signal delay as a function of elevation angle for four different receiver altitudes and control conditions. LEFT: Close-in view. RIGHT: Full view. 108

6-13 A contour plot of total signal delay as a function of altitude and geometric elevation angle. Contour lines are at 1, 5, 10, 15, 20, 30, 50, 70, 100 m delay. Note that this plot looks very similar to the ionosphere delay plot except at low receiver altitudes, where there is a kink in each of the otherwise smooth contour lines. Note that this plot also has the effects of resolution and interpolation visible. 109

6-14 Neutral atmosphere, ionosphere and total signal delay as a function of **season** for 0 km and 350 km receiver altitude. The delay during the summer solstice (6/21/80) is indicated with a dashed line, and the delay during the winter solstice (12/21/80) is indicated with a dash-dotted line. 112

6-15 Neutral atmosphere, ionosphere and total signal delay as a function of **season** for 700 km and 1700 km receiver altitude. The delay during the summer solstice (6/21/80) is indicated with a dashed line, and the delay during the winter solstice (12/21/80) is indicated with a dash-dotted line. 113

6-16 Neutral atmosphere, ionosphere and total signal delay as a function of **solar activity** at 0 km and 350 km receiver altitude. 114

6-17 Neutral atmosphere, ionosphere and total signal delay as a function of **solar activity** at 700 km and 1700 km receiver altitude. 115

6-18 Neutral atmosphere, ionosphere and total signal delay as a function of **time of day** at the receiver for 0 km and 350 km receiver altitude. . 117

6-19 Neutral atmosphere, ionosphere and total signal delay as a function of **time of day** at the receiver for 700 km and 1700 km receiver altitude. 118

6-20 Neutral atmosphere, ionosphere and total signal delay as a function of receiver **latitude** at 0 km and 350 km receiver altitude. Delay at 15°N is indicated by a thick dashed line; delay at 45°N by a thick dash-dotted line; and delay at 75°N by a thin dashed line. 120

6-21 Neutral atmosphere, ionosphere and total signal delay as a function of receiver **latitude** at 700 km and 1700 km receiver altitude. Delay at 15°N is indicated by a thick dashed line; delay at 45°N by a thick dash-dotted line; and delay at 75°N by a thin dashed line. 121

6-22 The "relevant" atmosphere for a spacecraft above the atmosphere is approximately the same number of degrees away from the receiver as the magnitude of the elevation angle to the satellite. 123

6-23 Neutral atmosphere, ionosphere and total signal delay as a function of satellite **azimuth** for 0 km and 350 km receiver altitude. Delay at 90° azimuth is indicated by a dashed line, and delay at 270° azimuth is indicated by a dash-dotted line. 124

6-24 Neutral atmosphere, ionosphere and total signal delay as a function of satellite **azimuth** for 700 km and 1700 km receiver altitude. Delay at 90° azimuth is indicated by a dashed line, and delay at 270° azimuth is indicated by a dash-dotted line. 125

6-25 LEFT: Receiver altitude as a function of time into flight. RIGHT: Receiver altitude as a function of longitude. Lines indicate the top of the neutral atmosphere and the top of the ionosphere in both plots. 126

6-26 The trajectory of the receiver and the distribution of GPS satellites that will be visible during the flight. Labeling of satellite PRN number is at the starting latitude and longitude of the satellite path during the 36 minutes of receiver flight. The horizon lines are not at the latitude and longitude of the horizon, but at the latitude and longitude of a satellite at GPS satellite orbital altitude that will be visible at the geometrical horizon. 127

6-27 The neutral atmosphere, ionosphere and total signal delay experienced by signals from 6 satellites during the sample receiver flight. The top plot gives the elevation to each satellite for reference. 129

6-28 The neutral atmosphere, ionosphere and total signal delay experienced by signals from 8 satellites during the sample receiver flight. The top plot gives the elevation to each satellite for reference. 130

6-29 The neutral atmosphere, ionosphere and total signal delay experienced by signals from 8 satellites during the sample receiver flight. The top plot gives the elevation to each satellite for reference. 131

6-30 The altitude of the receiver during the launch period of the receiver flight. 133

6-31 The neutral atmosphere, ionosphere and total signal delay experienced by signals from 6 satellites during the launch period of the sample receiver flight. Note that the receiver does not go above local sea level until 10 s into flight, and thus the ray-tracing program did not report any delay. 134

6-32 The neutral atmosphere, ionosphere and total signal delay experienced by signals from 8 satellites during the launch period of the sample receiver flight. Note that the receiver does not go above local sea level until 10 s into flight, and thus the ray-tracing program did not report any delay. 135

6-33 The neutral atmosphere, ionosphere and total signal delay experienced by signals from 8 satellites during the launch period of the sample receiver flight. Note that the receiver does not go above local sea level until 10 s into flight, and thus the ray-tracing program did not report any delay. 136

6-34 The altitude of the receiver during the reentry period of the receiver flight. 137

6-35 The neutral atmosphere, ionosphere and total signal delay experienced by signals from 6 satellites during reentry of the sample receiver flight. 138

6-36 The neutral atmosphere, ionosphere and total signal delay experienced by signals from 8 satellites during reentry of the sample receiver flight. 139

6-37 The neutral atmosphere, ionosphere and total signal delay experienced by signals from 8 satellites during reentry of the sample receiver flight. 140

6-38 Delay rates introduced by uncorrected total, neutral atmosphere and ionosphere signal delay experienced by signals from 6 satellites during the sample receiver flight. 141

6-39 Delay rates introduced by uncorrected total, neutral atmosphere and ionosphere signal delay experienced by signals from 8 satellites during the sample receiver flight. 142

6-40 Delay rates introduced by uncorrected total, neutral atmosphere and ionosphere signal delay experienced by signals from 8 satellites during the sample receiver flight. 143

6-41 Signal delays for the same trajectory flight with the same satellite positions on two different dates: 3/21/80 (high solar activity; solid) and 3/21/78 (low solar activity; dashed). Note that the time at which the peak delay occurs has shifted. 144

6-42 Total delay minus the sum of neutral atmosphere delay and ionosphere delay: one measure of error in delay calculations due to ray-tracing implementation. 146

6-43 Performance of the Altshuler and NATO delay models at 20, 40 and 60 km receiver altitude compared to neutral atmosphere ray-tracing results. Note that the Altshuler and NATO models do not cover negative elevation angles, and so predict no delay at 60 km receiver altitude. . 148

6-44 Performance of the Klobuchar delay model at 100, 350 and 700 km altitude compared to ionosphere ray-tracing results. Note that the Klobuchar algorithm did not determine at what elevation angle the signal would no longer be visible. 149

6-45 A comparison of ray-traced neutral atmosphere signal delays with delays predicted by the NATO model during reentry for 6 satellites. Note that the NATO model continues to predict delays even once a satellite has set. I have removed the NATO predictions for satellites that are not visible at any point during the reentry period. 151

6-46 A comparison of ray-traced neutral atmosphere signal delays with delays predicted by the NATO model during reentry for 8 satellites. Note that the NATO model continues to predict delays even once a satellite has set. I have removed the NATO predictions for satellites that are not visible at any point during the reentry period. 152

6-47 A comparison of ray-traced neutral atmosphere signal delays with delays predicted by the NATO model during reentry for 8 satellites. Note that the NATO model continues to predict delays even once a satellite has set. I have removed the NATO predictions for satellites that are not visible at any point during the reentry period. 153

6-48 A comparison of ray-traced ionosphere signal delays with the Klobuchar delay prediction during reentry for 6 satellites. Note that the Klobuchar model continues to predict delay even once a satellite has set. I have removed the Klobuchar predictions for satellites that are not visible at any point during the reentry period. 154

6-49 A comparison of ray-traced ionosphere signal delays with the Klobuchar delay prediction during reentry for 8 satellites. Note that the Klobuchar model continues to predict delay even once a satellite has set. I have removed the Klobuchar predictions for satellites that are not visible at any point during the reentry period. 155

6-50 A comparison of ray-traced ionosphere signal delays with the Klobuchar delay prediction during reentry for 8 satellites. Note that the Klobuchar model continues to predict delay even once a satellite has set. I have removed the Klobuchar predictions for satellites that are not visible at any point during the reentry period. 156

7-1 Ray-traced neutral atmosphere signal delay. Contour lines at 0.01, 1, 5, 10, 15, 20, 30, 50, 70, 100 m delay. Equation 7.1 is plotted over the points of 0.01 m delay. 159

7-2 Signals to a spacecraft that is within the atmosphere will pass through some layers of the atmosphere once and others twice. The angle at which a signal passes through each layer is different. 160

7-3 The L1 and L2 signals do not follow precisely the same path from the satellite to the receiver due to the dispersive nature of the ionosphere. The ray-tracing root-finder only requires the ray-traced signals to strike within 100 m of the true satellite position, causing the two signals to strike slightly different satellite positions. 163

7-4 A comparison between the dual frequency ionosphere correction and the ray-traced ionosphere delay. 164

7-5 Signals traveling to spacecraft potentially pass through two subionospheric points. 165

7-6 LEFT: The measure and hold algorithm makes a dual frequency calculation of the ionosphere signal delay for five seconds, and then uses the average of those measurements as the ionospheric correction over the next 20 seconds. RIGHT: The measure and propagate algorithm maintains the slope of the dual frequency ionosphere correction over the next 20 seconds. 166

7-7 TOP: Measure and hold ionosphere delay calculation for PRN 3. SECOND: Error in measure and hold delay calculation. THIRD: Measure and propagate ionosphere delay calculation. BOTTOM: Error in measure and propagate calculation. 167

7-8 TOP: Measure and hold ionosphere delay calculation for PRN 10. SECOND: Error in measure and hold delay calculation. THIRD: Measure and propagate ionosphere delay calculation. BOTTOM: Error in measure and propagate calculation. 168

7-9 TOP: Measure and hold ionosphere delay calculation for PRN 25. SECOND: Error in measure and hold delay calculation. THIRD: Measure and propagate ionosphere delay calculation. BOTTOM: Error in measure and propagate calculation. 169

7-10 TOP: Measure and hold ionosphere delay calculation for PRN 11. SECOND: Error in measure and hold delay calculation. THIRD: Measure and propagate ionosphere delay calculation. BOTTOM: Error in measure and propagate calculation. 170

7-11 TOP: Measure and hold ionosphere delay calculation for PRN 18. SECOND: Error in measure and hold delay calculation. THIRD: Measure and propagate ionosphere delay calculation. BOTTOM: Error in measure and propagate calculation. 171

7-12 TOP: Measure and hold ionosphere delay calculation for PRN 9. SECOND: Error in measure and hold delay calculation. THIRD: Measure and propagate ionosphere delay calculation. BOTTOM: Error in measure and propagate calculation. 172

[This page intentionally left blank.]

List of Tables

2.1 The error (in meters) between the true neutral atmosphere signal delay and the line-of-sight approximation for various elevation angles for a ground user at a northern-hemisphere, mid-latitude site as calculated by Hopfield [2] using ray-tracing. 37

2.2 Commonly cited coefficients for the neutral atmosphere refractivity equation (eqn. 2.10). "S and W" indicates Smith and Weintraub. . . 40

2.3 Height parameters derived with Hopfield's two-quartic refractivity model. [3] 50

2.4 Phase refractivity gradients for each signal propagation condition, as specified by Hitney et al. [4] 56

5.1 The effects of varying the precision of the root-finding algorithm for $L \approx 20,000$ km. 87

7.1 Variations in neutral atmosphere signal delay from that observed at 28°N as a function of latitude for 0 km receiver altitude. EA stands for geometrical elevation angle. 161

7.2 Variations in neutral atmosphere signal delay from that observed at 28°N as a function of latitude for 50 km receiver altitude. EA stands for geometrical elevation angle. 161

7.3 Variations in neutral atmosphere signal delay from that observed at 28°N as a function of latitude for 350 km receiver altitude. EA stands for geometrical elevation angle. 161

[This page intentionally left blank.]

Chapter 1

Introduction

A Global Positioning System (GPS) signal is both refracted (bent) and retarded (slowed) as it travels through the Earth's atmosphere, causing the signal to arrive at a GPS receiver later than it would have had it travelled through a vacuum. If not corrected for, this signal delay causes the receiver to calculate the originating satellite to be farther away than it really is and introduces an error into the position calculation of the receiver.

Signal delay is typically broken into two components: delay caused by the neutral atmosphere and delay caused by the ionosphere. The neutral atmosphere delays the signal as it travels near the surface of the Earth, up to approximately 50 km altitude. The ionosphere delays the signal in the region above the neutral atmosphere, from approximately 50 km to 1000 km altitude. For a signal traveling to a receiver that is on the surface of the Earth, neutral atmosphere delays range from 2–25 m and are a function primarily of the elevation angle to the originating satellite. Ionosphere delays for the same signal have roughly the same range of values as the neutral atmosphere delays, except that the ionosphere delay is strongly a function of the solar activity level and other conditions (such as the receiver latitude and the time of day) in addition to elevation angle.

Most GPS receivers employ two methods to correct for this signal delay. Single frequency receivers approximate both delays with models. The ionosphere model takes into account the time of day, receiver latitude, recent ionosphere conditions (via coefficients broadcast by the GPS satellites) and elevation angle to the GPS satellite. There are many available neutral atmosphere models; the ones that I examined take into account the time of day, day of year, latitude, height of receiver (up to $\approx$ 20 km) and elevation angle to the satellite. For a receiver that is on the surface of the Earth, these models correct for roughly 50% of the ionosphere delay and over 50% of the

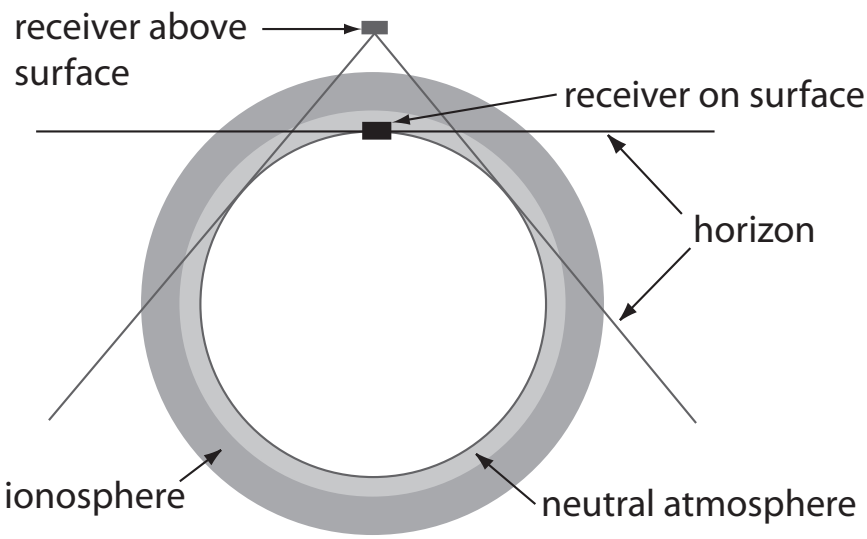

Figure 1-1: A receiver on the surface of the Earth can only see satellites at positive elevation angles. (Receivers on the surface of the Earth can actually see satellites at very small negative elevation angles. Atmospheric refraction bends the signal slightly so that it curves with the surface of the Earth, allowing receivers to see satellites slightly below the horizon.) A receiver above the surface of the Earth, however, can see satellites at negative elevation angles. Figure is not to scale.

neutral atmosphere delay. Dual frequency receivers take advantage of the dispersive nature of the ionosphere and are typically able correct for the vast majority of the ionosphere delay, for receivers both on and above the surface of the Earth. However, dual frequency receivers must still rely upon delay models to correct for the neutral atmosphere delay.

These delay correction methods run into two main problems when used in receivers aboard space vehicles.

First, the delay models are designed only for receivers on or near the surface of the Earth, and are designed only for satellites at positive elevation angles. Space vehicles both travel far from the surface of the Earth and are able to view GPS satellites at negative elevation angles (figure 1-1). Signals from satellites at negative elevation angles potentially travel through a given layer of the atmosphere twice — once on the way in and once on the way out (figure 1-2). Signals traveling such paths typically have much larger delays than do signals from satellites at positive elevation angles, which only pass through a given layer of the atmosphere once. Existing delay models are therefore inadequate for a receiver onboard a space vehicle.

Second, space vehicles travel much faster than do vehicles on or near the surface

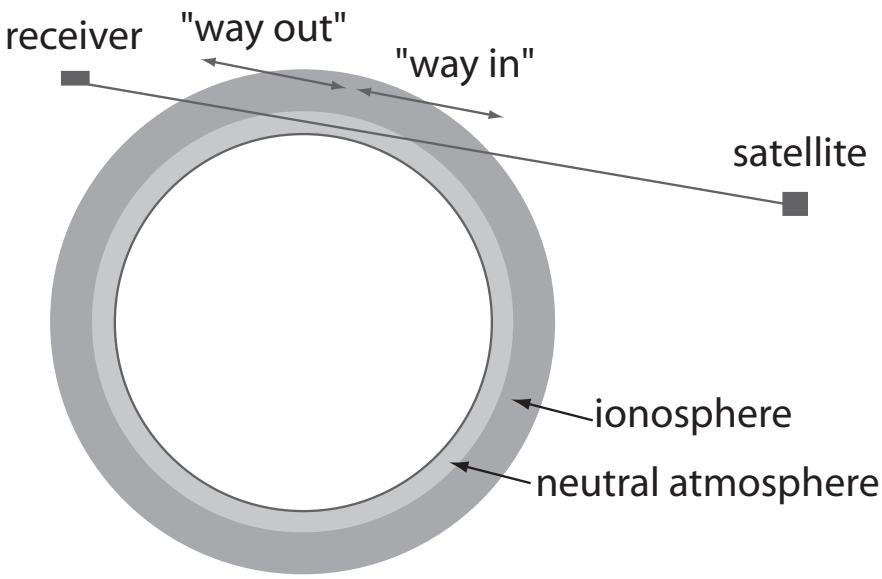

Figure 1-2: A signal from a GPS satellite at a negative elevation angle can potentially travel through the atmosphere twice: once on the "way in", and once on the "way out". Figure is not to scale.

of the Earth. The elevation angle and azimuth to a given satellite will therefore change quicker, as will the atmospheric conditions near the space vehicle. Both of these factors cause the delay on signals from a given satellite to change much more rapidly than would the delay of a signal from that same satellite to a receiver near the surface of the Earth. If a dual-frequency receiver does not have sufficient channels to constantly monitor each satellite on two frequencies, the ionosphere delay calculation may become quite inaccurate.

These problems are generally irrelevant aboard space vehicles — a sufficient number of GPS satellites are visible above the atmosphere, and the signals from these satellites experience no delay. However, there are times when a space vehicle receiver must rely upon satellites whose signals pass through the atmosphere. If the space vehicle has a rear-mounted antenna, the vehicle body may block the receiver antenna from receiving signals from over half the sky, and the Earth may block another quarter of the sky. This problem is particularly acute right after launch or during reentry, when the space vehicle is in the atmosphere but not close enough to the surface of the Earth to use the delay models. In figure 1-3, the GPS receiver must rely upon signals from region A both to achieve a good dilution of precision[1] and potentially to

[1]Dilution of precision is a measure of how well the GPS satellites are distributed about the receiver. A GPS receiver needs satellites both above it and to its sides in order to accurately

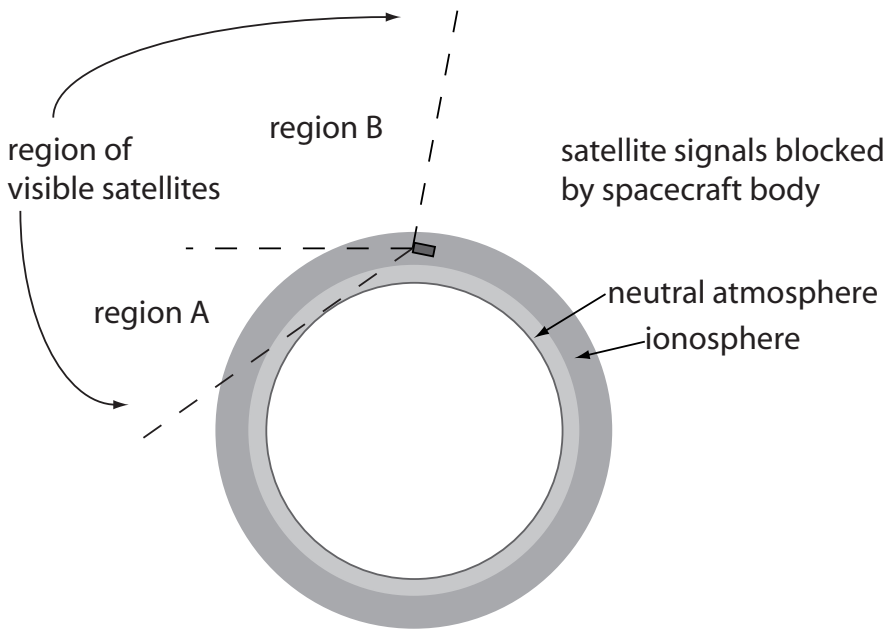

Figure 1-3: As a spacecraft with a rear-mounted GPS antenna is reentering the atmosphere the spacecraft body and the Earth will block signals from most GPS satellites. Signals from visible satellites in region A will experience delay both on their way "into" the atmosphere and on their way "out". Signals from region B will only pass through the atmosphere on their way "in". Figure is not to scale.

attain enough satellites for a navigation solution.

In this thesis I examine the delays experienced by signals to a GPS receiver aboard a space vehicle during launch, low Earth orbit and reentry by ray-tracing through the Global Reference Atmosphere Model (1999) and the International Reference Ionosphere (2001). The results are average monthly delays that are a function of a number of variables: season, solar activity, time of day, receiver latitude, receiver altitude and satellite azimuth. Since these are average delays, they are smooth approximations of the delays signals would realistically experience. They contain none of the features that actual delay curves would have due to small scale disturbances in the atmosphere such as thunderstorms, ionosphere scintillation and traveling ionospheric disturbances. I discuss methods for a GPS receiver to correct for these delays that would avoid the problems with current delay correction methods mentioned above.

pinpoint its position in all dimensions.

Chapter 2

GPS Signal Propagation

2.1 GPS Basics

The Global Positioning System (GPS) is a constellation of satellites designed for the navigation of objects near the surface of the Earth. Each satellite in the constellation broadcasts a uniquely coded signal indicating the time at which the signal was transmitted and information that can be used to determine the position of the transmitting satellite. A GPS receiver then uses the time at which the signal was received to calculate the distance (range) between the satellite and the receiver:

$$\rho_r^s = (t_r - t^s)\ c \tag{2.1}$$

where ρ_r^s is the range between the satellite and the receiver, t^s is the time at which the signal leaves the satellite, t_r is the time at which the signal arrives at the receiver and c is the speed of light. This range, combined with the position of the satellite, defines the surface of a sphere on which the receiver must lie. If all of the GPS satellite and receiver clocks are synchronized, we can pinpoint the precise position that the receiver is located on the sphere by making range measurements to two other satellites and calculating the intersection of the three resulting spheres.

The positioning described above is how GPS would work in an ideal world. In the real world, the satellite clocks and receiver clock are not synchronized, and each has a clock error. Therefore, the calculated range to the satellite is not quite correct and is called a pseudorange:

$$P_r^s = [(t_r + \Delta t_r) - (t^s + \Delta t^s)]\ c \tag{2.2}$$

where t_r and t^s are truth, and Δt_r and Δt^s are clock errors. Equation 2.2 is not complete: the signal both travels a curved path and experiences signal delays in the neutral atmosphere and the ionosphere; the receiver has measurement noise error; and relativistic effects must be accounted for.

$$P_r^s = \rho_r^s + (\Delta t_r - \Delta t^s)\, c + (\Delta NA_r^s + \Delta I_r^s + \nu_r + \Delta R_r^s)\, c \tag{2.3}$$

The GPS constellation contains up to 31 satellites, one for each available code. A working constellation contains as few as 24 satellites, but there are often active replacement satellites in orbit, increasing the number available. These satellites are distributed between six orbital planes, each orbiting with an approximate radius of 26,560 km and inclined 55°. In this configuration, four satellites will always be visible above 15° elevation angle everywhere on the surface of the Earth. As many as eight satellites may be visible above 15° elevation angle, and twelve satellites may be visible above 5° elevation angle.

The current GPS satellites broadcast on two frequencies: L1 (1.57542 GHz) and L2 (1.22760 GHz). Future satellites will broadcast on additional frequencies.

2.2 Propagation in the Atmosphere

A GPS user is interested in the undisturbed GPS signal as transmitted by the satellite. From this undisturbed signal, the user can ideally measure the exact distance between the receiver and each satellite, and pinpoint the receiver's position. Any GPS signal that travels through the Earth's atmosphere, however, will be disturbed. The atmosphere changes both the speed and direction of the signal propagation according to Snell's Law [5]

$$n_i \sin\theta_i = n_t \sin\theta_t$$

GPS signals will also be attenuated by the atmosphere, but this topic will not be covered in this thesis.

2.2.1 Dispersive vs. Non-Dispersive Media

In a vacuum, a GPS signal travels at the speed of light. In other media, the signal speed is characterized by the index of refraction:

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

For each signal, there are two indices of refraction: the group index of refraction, and the phase index of refraction. The group index of refraction characterizes the speed at which the wave group (i.e. information contained in the signal) travels, and the phase index of refraction characterizes the speed at which the phase of a given wave travels. These two values are the same in neutral, uncharged media. However, if the signal frequency is close to the atomic frequency of an electrically charged medium, the resulting resonance affects the propagation speed of the signal in a frequency-dependent manner [6], making the medium *dispersive*. For a narrow bandwidth signal, such as a GPS signal, the relation between the phase and group velocities is [7]:

$$v_g = v_p - \lambda \frac{dv_p}{d\lambda}$$

where λ is the wavelength of the signal. This equation, the Rayleigh equation, can be modified to describe the relation between the indices of refraction as well:

$$n_g = n_p - \lambda \frac{dn_p}{d\lambda}$$

Since information cannot travel faster than the speed of light, $n_g \geq 1$, and the information in a signal is always delayed by the atmosphere. However, in electrically charged media n_p can be less than 1, advancing the carrier phase.

2.2.2 Signal Delay

The length of time it takes the information contained in a GPS signal to reach a receiver is determined by both the group and phase indices of refraction along the signal's path. The phase index of refraction determines the path (see section 2.2.3), and the group index of refraction determines the speed the signal travels along that path. We are interested in the additional time it takes the signal to reach the receiver traveling through the atmosphere over the time it would have taken if the signal had travelled through a vacuum, so that we can remove that delay (ΔNA and ΔI) from equation 2.3 and make an accurate geometric range measurement to the GPS satellite. In the following equation expressing the combined neutral atmosphere and ionosphere delay, ds is the actual signal path and dl is the geometrical line connecting the satellite and the receiver:

$$\Delta = \int_{rcvr}^{sat} n_g \; ds - \int_{rcvr}^{sat} 1 \; dl \tag{2.4}$$

The delay can also be expressed in terms of delay due to slowing and delay due to bending:

$$\Delta = \int_{rcvr}^{sat} (n_g - 1)\ ds + \left[\int_{rcvr}^{sat} 1\ ds - \int_{rcvr}^{sat} 1\ dl \right] = 10^{-6} \int_{rcvr}^{sat} N_g\ ds + \Delta_{bend}$$

where

$$N = 10^6 (n - 1) \tag{2.5}$$

is the refractivity of a medium, and is a function of position in the atmosphere. Since both atmospheric and ionospheric indices of refraction are often quite close to unity, the scale of the refractivity of a medium is often easier to use.

For computational simplicity, many scientists and engineers assume that the curvature of the signal path is negligible and calculate the signal delay along the straight line between the satellite and the receiver:

$$\Delta \approx \int_{rcvr}^{sat} n_g\ dl - \int_{rcvr}^{sat} 1\ dl = 10^{-6} \int_{rcvr}^{sat} N_g dl \tag{2.6}$$

In the neutral atmosphere, this calculation provides an upper bound to the actual signal delay. As a signal travels into a more refractive region, it bends toward the normal to the boundary of the regions, minimizing the time spent in the "slower" region. So, while the curvature of the signal path adds distance travelled, it actually minimizes the total neutral atmosphere signal delay as compared to the straight line path. In the ionosphere, the phase index of refraction is less than unity. As a signal travels into a more (higher magnitude) refractive region, it bends away from the normal to the boundary of the regions. Therefore, the refracted signal actually spends more time in the ionosphere than would a signal along the straight line path, so this straight line calculation provides a lower bound to the actual ionospheric signal delay.

Assuming a spherically layered atmosphere, path bending increases as the elevation angle of the signal decreases because the signal is further from the normal to the boundary of the mediums. The error between the true signal delay and the line-of-sight approximation is

$$\epsilon = \left[\int_{rcvr}^{sat} n_g\ dl - \int_{rcvr}^{sat} 1\ dl \right] - \left[\int_{rcvr}^{sat} n_g\ ds - \int_{rcvr}^{sat} 1\ dl \right] = \int_{rcvr}^{sat} n_g\ dl - \int_{rcvr}^{sat} n_g\ ds$$

According to ray-tracing done by Hopfield [2] at a northern-hemisphere, mid-latitude site, this error was centimeters large for the neutral atmosphere when elevation angles

Elevation Angle	January, 1967	July, 1967
60°	0.00008	0.00007
45°	0.00030	0.00026
30°	0.00125	0.00111
20°	0.0045	0.0040
10°	0.034	0.0310
5°	0.22	0.194
4°	0.38	0.33
3°	0.71	0.60
2°	1.5	1.25
1°	4.2	3.1

Table 2.1: The error (in meters) between the true neutral atmosphere signal delay and the line-of-sight approximation for various elevation angles for a ground user at a northern-hemisphere, mid-latitude site as calculated by Hopfield [2] using ray-tracing.

decreased to around 10°, and could get as large as 4.2 m for elevation angles of 1° (see Table 2.1). For this reason, any signal delay calculation by integration of indices of refraction must be conducted along the true signal path when the signal is at a low elevation angle.

2.2.3 Signal Path

The path that a signal travels through the atmosphere nominally obeys Fermat's Principle of Least Time, which states that "the actual path between two points taken by a beam of light is the one that is traversed in the least time."[5] As Hecht goes on to state, "As we shall see, even this form of the statement is incomplete and a bit erroneous at that... Fermat's Principle in its modern form reads: *a light ray in going from point S to point P must traverse an optical path length that is stationary with respect to variations of this path.*" This is a direct result of Huygen's principle, and can be used to derive Snell's Law. Essentially, it is saying that all wavelets traveling along paths close to the true path of the signal will arrive at the point P with nearly the same phase, and therefore reinforce each other. (This is because the path length is **stationary** with respect to variations – i.e.,

$$\int_{rcvr}^{sat} n_p \, ds \tag{2.7}$$

where n_p is the phase index of refraction, is stationary. In practice, this will either be a minimum or a saddle point.) Wavelets traveling other paths will arrive with a

random phase and cancel each other out. I will use the fact that eqn. 2.7 is stationary to derive differential equations that describe the signal path in section 5.1.2.

2.3 Refractivity in the Neutral Atmosphere

We have now seen that the atmospheric signal delay can be fully calculated knowing only the index of refraction (or equivalently, the refractivity) along the signal path. In the neutral atmosphere the dielectric constant of air ϵ_a is greater than one and the magnetic permeability of air μ_a is approximately one, causing the index of refraction of air n_a to be greater than one. Consequently, the GPS signal travels slower in the neutral atmosphere (at speed v_a) than it would in a vacuum (c) [8]:

$$n_a = \sqrt{\epsilon_a \mu_a}$$
$$\frac{c}{n_a} = v_a < c$$

Since there are, by definition, few charged particles in the neutral atmosphere, the neutral atmosphere is a non-dispersive medium at microwave frequencies and the group index of refraction there is equal to the phase index of refraction.

The radio refractivity of the neutral atmosphere is often expressed in the form

$$N(T, P_d, e) = k_1 \frac{P_d}{T} + k_2 \frac{e}{T} + k_3 \frac{e}{T^2} \tag{2.8}$$

where T is the temperature in Kelvin, P_d is the partial pressure due to dry gasses in millibars and e is the partial pressure of water vapor, in millibars. Classically, the first term accounts for the induced dipole moment of the dry constituents of the atmosphere. The second and third terms account for the induced and permanent dipole moments of the water vapor in the atmosphere, respectively [9]. An alternate version of eqn. 2.8 absorbs k_2 into k_1 by replacing P_d with the total pressure of air $P = P_d + e$ and slightly modifying k_3 to compensate:

$$N(T, P, e) = k_1' \frac{P}{T} + k_3' \frac{e}{T^2} \tag{2.9}$$

This equation is accurate to roughly 0.5% (approximately 2 N, where N is the "units" of refractivity) given Smith and Weintraub's [10] values for the k coefficients, and was the standard equation used in the 1950's and 1960's.

In 1967, Owens [11] derived compressibility factors (Z_d^{-1}, Z_w^{-1}) to account for the

nonideal behavior of the atmospheric constituents. He fit terms involving P and T to tabulated data for the dry compressibility factor, and fit terms involving P and T to values calculated from an equation of state for the wet compressibility factor. He found that they were of the form:

$$\begin{aligned} Z_d^{-1} &= 1 + P\left(a + \frac{b}{T} + \frac{c}{T^2}\right) \\ Z_w^{-1} &= 1 + P(1 + dP)\left(e + \frac{f}{T} + \frac{g}{T^2} + \frac{h}{T^3}\right) \end{aligned}$$

Incorporating these factors into equation 2.8 increases the accuracy of the refractivity calculation.

$$N(T, P_d, e) = k_1 \frac{P_d}{T} Z_d^{-1} + \left(k_2 \frac{e}{T} + k_3 \frac{e}{T^2}\right) Z_w^{-1} \tag{2.10}$$

where the compressibility terms have been somewhat modified from the original. The following compressibility terms are presented in multiple sources [12, 13, 9]. Thayer [12] and Davis [9] claim these equations originate from Owens [11], but as best as I could tell these equations were not derived in [11].

$$\begin{aligned} Z_d^{-1}(T, T_c, P_d) &= 1 + P_d\left[57.97 \times 10^{-8} \times \left(1 + \frac{0.52}{T}\right) - 9.4611 \times 10^{-4} \times \frac{T_c}{T^2}\right] \\ Z_w^{-1}(T, T_c, e) &= 1 + 1650\ \frac{e}{T^3}\left[1 - 0.01317\ T_c + 1.75 \times 10^{-4}\ T_c^2 + 1.44 \times 10^{-6}\ T_c^3\right] \end{aligned}$$

(T_c is the temperature in Celsius.) Compressibility factors are a few thousandths larger than unity for the atmosphere [9]; neglecting these terms leads to errors of 0.05% to 0.1% (0.14 N to 0.5 N) for dry to humid conditions at sea level [12]. Owens determined these compressibility factors using a least-squares fit to thermodynamic data. They are accurate to within a few parts per million [9]. Smith and Weintraub's formulation (eqn. 2.8) is now rarely used in the literature; it has been replaced by equation 2.10.

The constants k_1, k_2 and k_3 can be solved for experimentally in the lab, fit to observed atmospheric data, or solved theoretically. Table 2.2 lists the commonly cited values for these coefficients.

Smith and Weintraub [10] took existing laboratory measurements, adjusted them for 0.03% CO_2 and so that they represented "real" rather than "ideal" properties, averaged them and solved for k_1, given that T is not the absolute temperature, but rather $T_c + 273$ (and so is different from the absolute temperature by ≈ 0.16). They used the Debye constants provided by Birnbaum and Chatterjee [17] to solve for

Source	k_1(K/mb)			k_2(K/mb)			k_3(K^2/mb)		
S and W (1953) [10]	77.607	$\pm$	0.013	71.6	$\pm$	8.5	374700	$\pm$	3100
Boudouris (1963) [8]	77.6	$\pm$	0.08	71.9	$\pm$	10.5	375400	$\pm$	3000
Thayer (1974) [12]	77.604	$\pm$	0.014	64.79	$\pm$	0.08	377600	$\pm$	400
Hill et al. (1982) [14]	$--$			98	$\pm$	1	358300	$\pm$	300
Hill (1988) [15]	$--$			102	$\pm$	1	357800	$\pm$	300
Bevis et al. (1994) [16]	77.60	$\pm$	0.05	70.4	$\pm$	2.2	373900	$\pm$	1200

Table 2.2: Commonly cited coefficients for the neutral atmosphere refractivity equation (eqn. 2.10). "S and W" indicates Smith and Weintraub.

k_2 and k_3. Their results are valid for temperatures between 223 and 313 K, total pressures between 200 and 1,100 mb, partial water vapor pressures between 0 and 30 mb, and frequencies between 0 and 30,000 mHz. (Note: Smith and Weintraub fit the coefficients to eqn. 2.8, and thus may have had slightly different results than if fitting to eqn. 2.10.)

Boudouris [8] measured indices of refraction for both dry air and water vapor using a cavity spectrometer for pressures between 0 and 1013 mb, and temperatures between 273 and 323 K. He adjusted his dry air measurements to include CO_2. Since the errors in k_2 and k_3 are correlated, his seemingly large errors (14.5% for k_2 and 0.8% for k_3) reduce to approximately 0.2% when taken together [12]. (Note: Boudouris fit the coefficients to eqn. 2.8, and thus may have had slightly different results than if fitting to eqn. 2.10.)

Thayer [12] derived his k_1 and k_2 coefficients from laboratory measurements of the optical refractivity of air. He corrected for the different magnetic permeability at optical and radio frequencies, and removed the temperature and pressure dependence of published refractivity values to create his radio frequency coefficients. Thayer derived k_3 by incorporating his value of k_2 in a weighted average with Boudouris' [8] radio refractivity, temperature and pressure measurements. Thayer states that his values have errors of 0.018%, 0.12% and 0.11%, respectively, but that in temperate regions, the total error on N will be approximately 0.05% (0.2 N).

Hill et al. [14] solve for the k coefficients more indirectly, by summing the refraction due to the infrared resonances of a water vapor monomer. Each absorption resonance adds to the radio refraction of water vapor. Infrared resonances do not quite contribute like $\frac{e}{T}$, but are minor contributors to radio refractivity. Far-infrared resonances vary like $\frac{e}{T^2}$, but energy changes of the order K_BT (i.e. rotational transitions at the resonance frequency $K_BT/\hbar c$) will <u>not</u> vary like either $\frac{e}{T}$ or $\frac{e}{T^2}$. Despite

the failure of their approach to conform with the dipole formulation of refractivity, Hill et al. fit their summed refractivity to the form of eqn. 2.10 and added the effects at optical frequencies to get their coefficients. Note that they are significantly different from all other quoted values, but since their k_2 is larger and their k_3 is smaller than the other sources' values, the diverging contributions of these two terms somewhat cancel and the resulting wet refractivity is only 2.4% less than Boudouris' at 20°C. The difference between Hill et al.'s and Boudouris' refractivities decreases as temperature increases. Hill et al. offer possible explanations for why they derived different coefficients; one is that the rotational resonances contribute approximately like $1/T^{1.97}$, not $1/T^2$. The contribution of the rotational resonances is thus partially assigned to k_2, increasing Hill et al.'s k_2 and decreasing Hill et al.'s k_3 with respect to the other sources' values. The posted coefficients are valid for 273 to 333 K; deviations caused by temperatures between 263 and 523 K are also within the errors given. Hill later recalculated coefficients using updated resonance line parameters [15]. His calculations differed even more from empirical results than before.

As you can see, Thayer's calculation contains the smallest errors overall, and so appears to be the best choice. There is, however, some disagreement over his values of k_2 and k_3. Hill et al. [14] argue that Thayer applies the optical k_2 to radio frequencies and ignores the contribution of infrared water vapor resonances to k_2. Thayer [12] *does* modify the optical term to account for the different magnetic permeability at radio frequencies, but Hill et al.'s arguments are not affected by this fact.

Bevis et al. [16] sought to provide robust k coefficients with conservative errors to provide an empirical alternative to Hill's deviant and Thayer's disputed values. They averaged 20 previously published values of k_1 and seven values each of k_2 and k_3 derived experimentally between 1951 and 1970. (Hasegawa and Stokesbury used these same values to compute an average of the empirical k coefficients, but Hill et al. [14] and Bevis et al. both dispute the weighting used in that study.) Bevis et al. calculated the mean, the standard error to that mean (based on the dispersion about the mean), the weighted mean, and the median value of the 20 empirical values of k_1. After removing three outliers, the mean, weighted mean and median value were all within 0.01 K/mb of each other, so the mean value was adopted. For k_2 and k_3, Bevis et al. felt that the quoted errors were statistically incompatible. Lacking a method to judge which errors were incorrect, they calculated an unweighted average of the seven values for each coefficient and calculated a standard error in the same manner as for k_1.

The coefficients k_1 and k_2 must be chosen simultaneously. If k_2 is reduced slightly,

the refractivity can be brought approximately back to its previous value at a given T and e by increasing k_3 [14]. The reverse also applies, and therefore we must use coefficients from the same source.

Because Thayer and Hill et al.'s values are not accepted reliably in the literature, I have chosen to use Bevis et al.'s refractivity coefficients.

In equations 2.10 and 2.8 the refractivity is split into a dry component and a wet component:

$$\begin{aligned} N_d(T,P) &= k_1 \frac{P}{T} Z_d^{-1} \\ N_w(T,e) &= \left(k_2 \frac{e}{T} + k_3 \frac{e}{T^2} \right) Z_w^{-1} \end{aligned}$$

(Since k_2 was absorbed into k_1 in Smith and Weintraub's simplification (eqn. 2.9), it cannot be split into dry and wet components.) This separation will be important in some of the neutral atmosphere signal delay calculation methods described later (see section 3.1). Generally, for a receiver on the surface of the Earth, the dry component will contribute to approximately 90% of the zenith signal delay [13, 18, 7]. At low elevation angles, however, the wet component's contribution will increase considerably [2].

When deriving k_1 the composition of dry air is assumed to be standard. The atmospheric CO_2 level is typically 0.04%[1]. If the CO_2 level is likely to exceed 0.06%, the dry refractivity must be calculated as [12, 18]:

$$N_d(T, P_{CO2-free}, P_{CO2}) = k_1' \frac{P_{CO2-free}}{T} + 133 \frac{P_{CO2}}{T}$$

2.4 Refractivity in the Ionosphere

The ionosphere has, by definition, a large number of charged particles, and is therefore a dispersive medium. This means that while the group index of refraction is greater than unity, the phase index of refraction is less than unity, causing the signal to bend counter-intuitively. Further, the bending and delay experienced by the signal in this region will be a function of frequency.

The full expression of the index of refraction in the ionosphere, the Appleton

[1] Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory. Measurements from Mauna Loa, Hawaii. URL: http://cdiac.esd.ornl.gov/trends/co2/sio-mlo.htm

formula, is derived by Davies [19] and described more succinctly by Klobuchar [20]:

$$n^2 = 1 - \frac{X}{1 - iZ - \frac{Y_T^2}{2(1-X-iZ)} \pm \left[\frac{Y_T^4}{4(1-X-iZ)^2} + Y_L^2\right]^{1/2}} \tag{2.11}$$

where

$$\begin{aligned}
Z &= \frac{f_\nu}{f} \\
X &= \frac{N_e^2}{\epsilon_0 m\omega^2} = \frac{f_N^2}{f^2} \\
Y_L &= \frac{eB_L}{m\omega} = \frac{f_H}{f}\cos\theta \\
Y_T &= \frac{eB_T}{m\omega} = \frac{f_H}{f}\sin\theta \\
Y &= \frac{f_H}{f} \\
B_L &= B\cos\theta \\
B_T &= B\sin\theta
\end{aligned}$$

where f is the frequency of the signal ($\approx 10^9$ Hz for GPS), f_ν is the electron-neutral collision frequency ($\approx 10^4$ Hz), f_N is the plasma frequency of the ionosphere, f_H is the electron gyro frequency (≈ 1.5 MHz) and θ is the angle of the signal with respect to the Earth's magnetic field. N_e is the electron density of the ionosphere, ϵ_0 is the permittivity of free space (8.854×10^{-12} farad/m), m is the rest mass of an electron (9.107×10^{-31} kg), $\omega = 2\pi f$, and B is the strength of the Earth's magnetic field.

The complex part of this equation (Z) describes the effect of energy absorbed from collisions on the index of refraction. Evaluating this ratio for GPS, we see that Z ($\approx 10^{-5}$) is quite small, and can be neglected. Likewise, we see that $Y \approx 10^{-3}$.

Klobuchar states that eqn. 2.11 can thus be reduced to:

$$n_p = 1 - \frac{X}{2} \approx 1 - \frac{40.3}{f^2} N_e \tag{2.12}$$

to better than 1%. Brunner and Gu (cited in [20]) kept the next two higher order terms, leaving:

$$n_p = 1 - \frac{X}{2} \pm \frac{XY}{2}\cos\theta - \frac{X^2}{8} \tag{2.13}$$

Klobuchar goes on to state that the first order term is at most $\approx 2 \times 10^{-4}$, where as the higher order terms are at most $\approx 2 \times 10^{-7}$ and $\approx 2 \times 10^{-8}$, respectively, at GPS

frequencies, indicating that the first-order approximation (eqn. 2.12) is good to 0.1%. (This is equivalent to $\approx$ 15 cm of signal delay in the worst-case scenario of 150 m of ionosphere signal delay.)

If we use eqn. 2.12 to represent the phase index of refraction of the ionosphere and the following corresponding equation for the group index of refraction of the ionosphere:

$$n_g \approx 1 + \frac{40.3}{f^2} N_e \tag{2.14}$$

the modeled signal path and delay will be strictly a function of the electron density of the ionosphere.

2.5 The Neutral Atmosphere

The atmosphere extending from the surface of the Earth to approximately 50 km is responsible for the neutral atmosphere signal delay experienced by a GPS signal. In this region, where the atmosphere is dense enough to have a significant neutral atmospheric refractivity, molecules collide often and charged particles are quickly neutralized. Therefore, this region has a negligible amount of charged particles and is not a dispersive medium.

A vast majority of the mass of the atmosphere lies in the troposphere, or the region extending from the surface of the Earth to 12 km altitude [21]. Therefore, approximately 75–80% of the neutral atmosphere signal delay takes place in the troposphere [22, 13], and the neutral atmosphere signal delay is often referred to as "tropospheric" signal delay. (Together, the troposphere and stratosphere, extending from the surface of the Earth up to 45 km altitude, account for 99.9% of the mass in the atmosphere [21], and essentially all of the neutral atmosphere delay.)

We are most interested in the temperature, pressure and water vapor distribution in this region so that we may determine the refractivity (eqn. 2.10) and then the signal delay (eqn. 2.4) that occurs in the neutral atmosphere.

The neutral atmosphere is often described by its temperature regimes. The troposphere extends from the surface of the Earth to approximately 12–16 km altitude. In this region, the temperature decreases roughly linearly with altitude. The temperature then remains constant throughout the tropopause, which is the boundary between the troposphere and the stratosphere and extends from approximately 12–16 to 20 km altitude [13]. In the stratosphere the temperature is initially constant, but then increases with altitude until the stratopause, at approximately 45 km altitude.

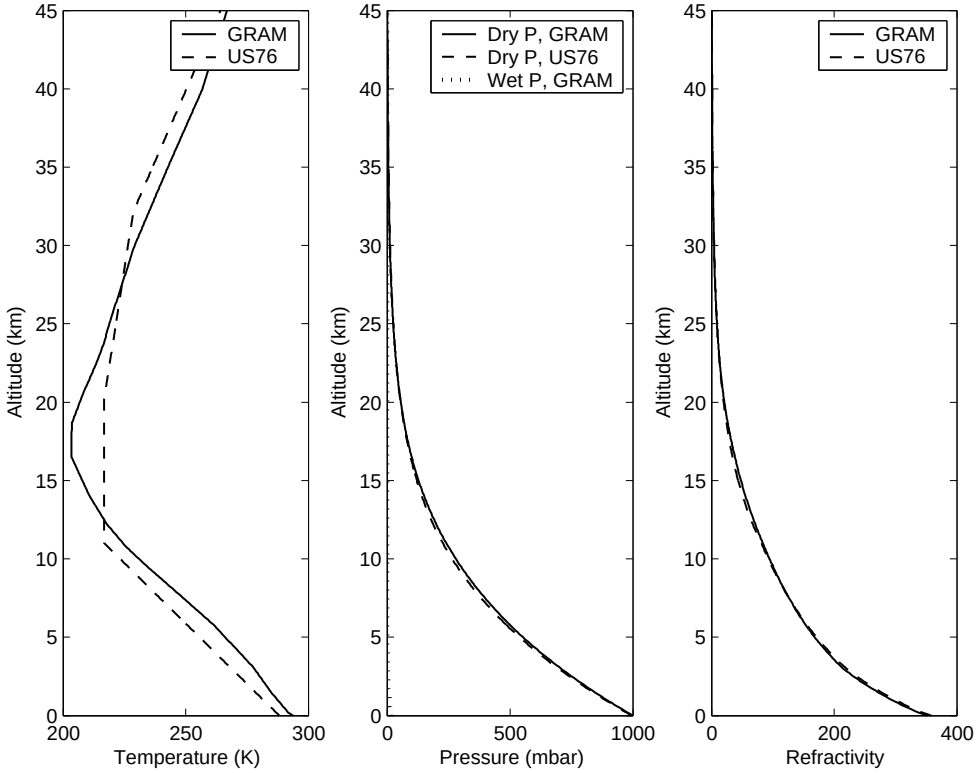

Figure 2-1: Temperature, pressure and refractivity profiles from the GRAM-99 and U.S. Standard Atmosphere 1976 atmosphere models for noon on 3/21/80 at 28°N 80°W. Note that the shape of the pressure curve drives the shape of the refractivity curve.

The three final temperature regions of the neutral atmosphere are not dense enough to affect the refractivity of the atmosphere: the mesosphere, which extends from 45 km to approximately 95 km, in which temperature decreases with altitude and reaches its lowest point; the thermosphere, which extends from 95 km to approximately 400 km; and the exosphere, which extends from 400 km to approximately 1000 km. [23]

The lower 100 km of the atmosphere is largely homogeneous, composed of molecular nitrogen and oxygen. There is a large amount of mixing between the lowest region of the stratosphere (below 21 km) and the troposphere, but little mixing with the regions above 21 km [23].

Water vapor is not distributed homogeneously in the atmosphere. It exists primarily in the troposphere, with approximately half located below 1.5 km altitude and approximately 95% located below 5 km altitude [13, 24].

The atmospheric pressure profile is well approximated by the ideal gas law, discussed in the following section.

2.5.1 Ideal Gas Law Model

The atmosphere is well approximated by the ideal gas law, and the following derivation by Spilker [13] gives a good feel for the temperature and pressure behavior of the atmosphere. These equations provide spatial values of T and P provided that values at the Earth's surface are known.

Assuming that only reversible, adiabatic processes take place in the troposphere, the temperature in the troposphere decreases with altitude at an approximately constant lapse rate (β):

$$T = T_0 + \beta h \tag{2.15}$$

$$\beta = -\left(\frac{\gamma - 1}{\gamma}\right)\frac{Mg}{R} \approx -6.5 \text{ km/K} \tag{2.16}$$

where M is the molar mass of the atmospheric gases, g is the acceleration due to gravity, R is the gas constant and $\gamma = \frac{C_p}{C_v}$ is the ratio of the heat capacities of the atmospheric gases. (The value of g changes slightly with altitude, so that β is not truly constant; further, changes in water vapor content affect the value of γ.) The temperature in the tropopause is constant, and then the temperature in the stratosphere increases roughly linearly with altitude. Spilker chose to assume that the temperature of the stratosphere is constant, arguing that it varies more slowly than in the troposphere. (As one can see in figure 2-1, the GRAM-99 creators did

not feel this assumption reflected available data.)

Using the same reversible, adiabatic assumptions, we can describe atmospheric pressure in the neutral atmosphere:

$$\begin{aligned} P = &\quad P_0 \left(\frac{T_0+\beta h}{T_0}\right)^{\frac{-Mg}{R\beta}} && h < h_{tpp} \\ P = &\quad P_0 \exp\left(\frac{-Mg(h-h_{tpp})}{R\ T_{tpp}}\right) && h \geq h_{tpp} \end{aligned}$$

where the subscript *tpp* refers to the tropopause.

Since the water vapor in the atmosphere is so variable and unpredictable, there is no good way of predicting the water vapor pressure (e) at altitude given a value at the Earth's surface. Saastamoinen [25] argues that despite this variability the water vapor pressure still decreases with altitude, although more quickly than the dry pressure. He provides the following "very approximate" equation for e in the troposphere:

$$e = e_0 \left(\frac{T}{T_0}\right)^{\frac{-4g}{R\beta}}$$

Many measurements have shown "extreme dryness" in stratosphere [26]. More recent measurements using microwave radiometry [27] agree with earlier radiosonde measurements that e is negligible in the tropopause and stratosphere (on clear days), but too many measurements conflict to conclusively say $e = 0$ there.

2.5.2 Variations in the Neutral Atmosphere

The characteristics of the atmosphere vary with season, latitude, surface altitude and whether or not the atmosphere is above land or water. These variations are important to understanding how broadly applicable the ray-tracing results presented in chapter 6 are, and the motivation behind which independent variables were chosen to examine variations in signal delays.

The Troposphere

Atmospheric water vapor content varies greatly across the globe in both magnitude and vertical distribution, and can vary drastically in a given location on time scales of less than a day, making it very unpredictable and difficult to model. In comparison, the dry atmosphere is relatively consistent in composition and vertical distribution. [13]

Atmospheric density varies seasonally, peaking slightly after the equinoxes and hitting a minimum slightly after the solstices. This variation results from the Earth-Sun separation distance and the changing solar zenith angle. [23]

The temperature profile can vary with latitude, season and time of day. In the upper latitudes, both in winter and at night, a temperature inversion layer often forms near the surface in which the temperature increases with altitude for 2 km, and then decreases with altitude as expected in the troposphere. At all latitudes, the region below 2 km is more variable that the region above [28]. The maximum altitude of the troposphere decreases as you move toward the poles, from approximately 12 km at the equator to 8 km at the poles in winter, and from 12 km at the equator to 10 km at the poles in summer [13]. Equations for its height are often expressed in terms of surface temperature, however, which roughly correlates with latitude and helps take into account the seasonal variation [18]:

$$h_w = 13.268 - 0.09796\ T_0$$

where h_w is in km and T_0 is the surface temperature, in Kelvin.

The tropospheric lapse rate, β, varies seasonally (decreasing in magnitude in the summer and increasing in the winter) and with latitude from its mean of -6.5 K/km, and is also affected by the local water content of the atmosphere. Because both T_0 and β in equation 2.15 vary seasonally, the temperature at the tropopause varies seasonally, but with a lower amplitude than at the surface of the Earth [28]. Further, the rate at which T_0 varies with latitude varies seasonally: in the winter hemisphere, the temperature decreases twice as rapidly from the equator to the pole as in the summer hemisphere. The seasonal temperature variation also depends on whether the location is over land or ocean. All of these dependencies combined make the actual situation quite complex.

The Stratosphere

The maximum altitude of the stratosphere (the altitude of the stratopause) varies roughly with latitude and/or surface temperature. Initially, Hopfield [3] postulated that this height was greatest at the equator and decreased toward the poles, proposing an equation of the form:

$$h_d = h_{d,equator} + A_d \sin^2 \phi \tag{2.17}$$

where h_d is the altitude of the stratopause at the chosen latitude (ϕ), $h_{d,equator}$ is the altitude of the top of the stratosphere at the equator, and A_d is the (negative)

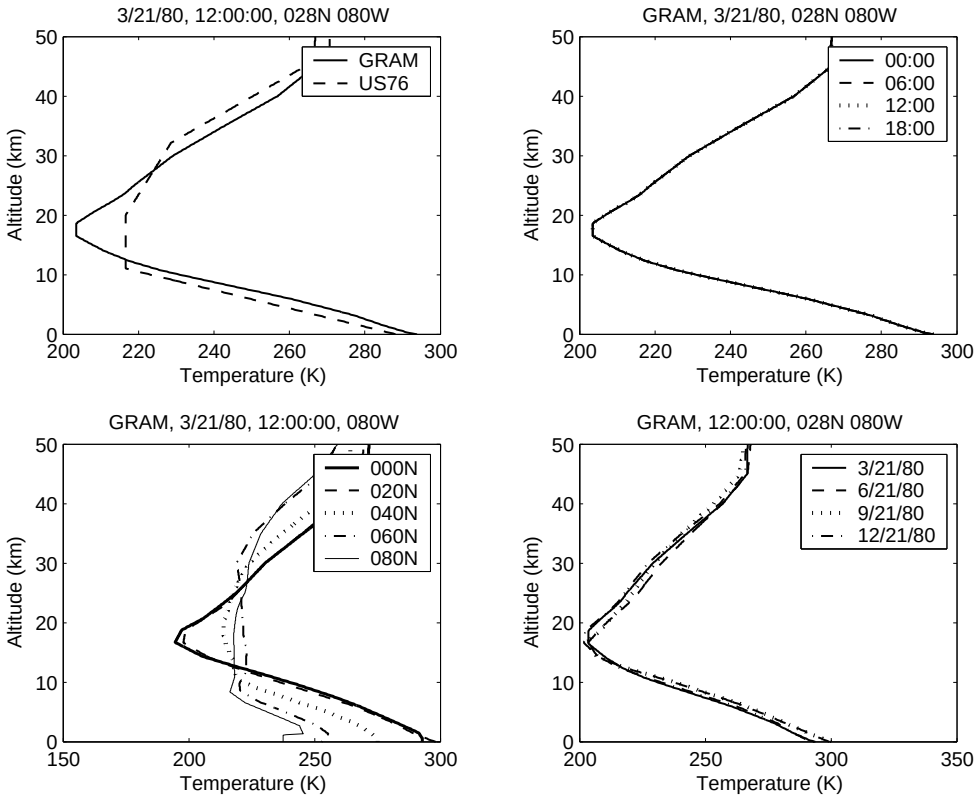

Figure 2-2: The GRAM-99 temperature profile, which shows trends representative of the average atmosphere. TOP LEFT: Comparing the GRAM-99 profile with the U.S. Standard Atmosphere 1976 profile. TOP RIGHT: The GRAM-99 temperature profile does not change with time of day on a mid-latitude, spring day. BOTTOM LEFT: The GRAM-99 temperature profile varies greatly with latitude. Note the inversion layers at the higher latitudes. BOTTOM RIGHT: The GRAM-99 profile varies only slightly with season at the mid-latitudes.

h_w, km	$h_{d,equator}$, km	A_d, km
10	43.858	−5.986
12	43.130	−5.206
14	42.402	−4.426

Table 2.3: Height parameters derived with Hopfield's two-quartic refractivity model. [3]

amplitude at which the height of the troposphere decreases with latitude. She found $A_d = -5.206$ and $h_{d,equator} = 43.130$ to be suitable in most instances; additional values based on the maximum altitude of the troposphere are given in table 2.3. Later, Hopfield [22] stated that h_d was proportional to surface temperature:

$$h_d = 40.136 + 0.14872 \times (T_s - 273.16) \tag{2.18}$$

where T_s is the surface temperature in Kelvin and h_d is given in km above the geoid of the Earth. Hopfield's approach, which involved fitting data to integrals of her atmospheric profile (see section 4.2), ignored the actual shape of the atmospheric profile when the data was taken, making the values produced by these equations only approximately correct unless used in conjunction with her profile.[2]

In the stratosphere the temperature distribution changes with season. In the winter hemisphere, the temperature is highest over the middle latitudes and lower above the pole and equatorial regions; in the summer hemisphere, the equator is colder than the pole [21].

2.6 The Ionosphere

The ionosphere is a region containing ions and free electrons extending from approximately 60 km to 2000 km. The ions and electrons are present in roughly equal quantities, making the ionosphere charge neutral. Despite the name *iono*sphere, however, ions and electrons only constitute approximately 0.4% of the particles at a given altitude [23]. The low atmospheric density in this region allows them to persist without recombining for extended periods of time, creating sufficient quantities to delay

[2]Hopfield assumed that the height of the tropopause was constant, and then found the stratopause height which made the zenith tropospheric signal delay using her refractivity profile (eqn. 4.1 with $N_{d,0}$ calculated from Smith and Weintraub's simplified refractivity equation (eqn. 2.9)) equal that from balloon data. She then did a least-squares fit of these heights to an equation of the form 2.17. She found A_d and $h_{d,equator}$ to be functions of the chosen h_w, and found $h_w = 12$ km to be suitable in most instances. (See table 2.3.)

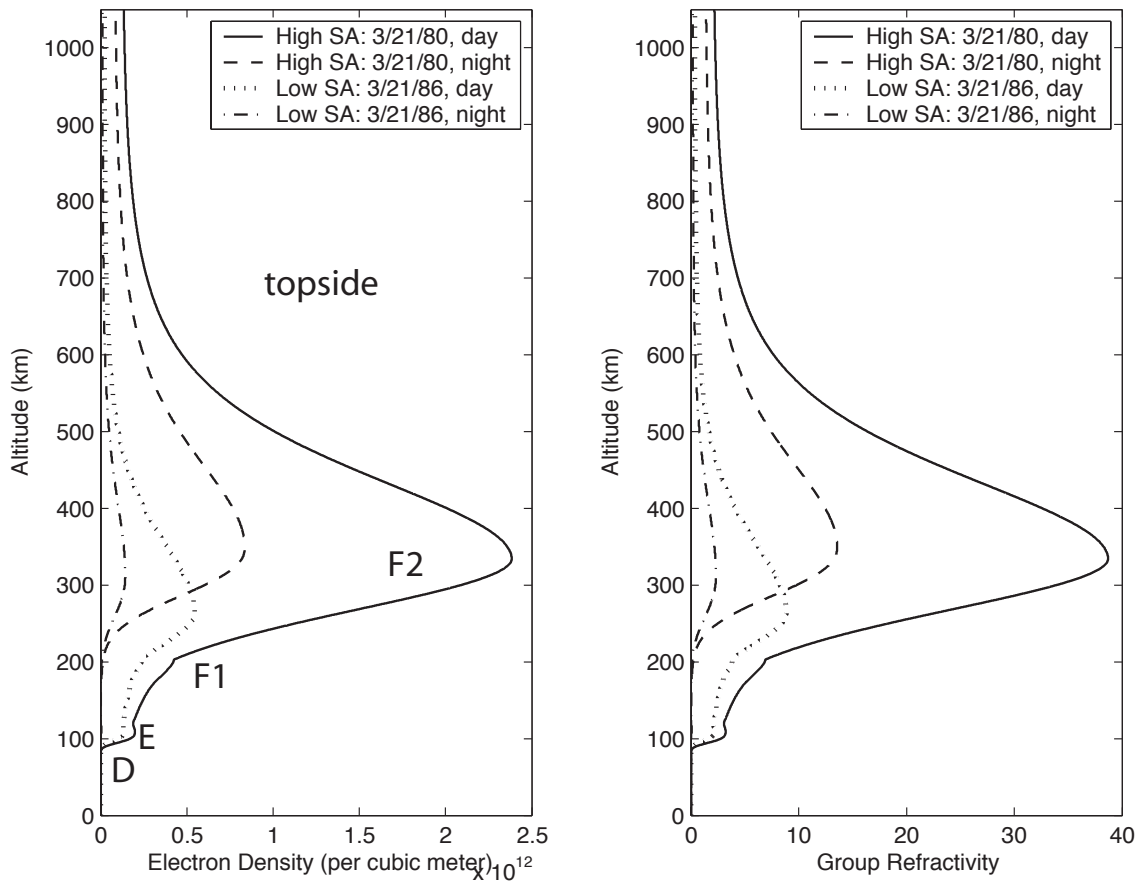

Figure 2-3: Electron density and group refractivity curves during high solar activity and low solar activity periods at noon and at midnight, as specified by the International Reference Ionosphere Model. Profile located at 28°N 80°W. Layer markings are for the high solar activity, daytime curve.

GPS signals. Since there are a large number of charged particles, the ionosphere is a dispersive medium.

We are most interested in the electron density in this region, so that we may determine the refractivity (eqn. 2.12) and then the signal delay (eqn. 2.4).

The electron density profile can generally be described in terms of five regions: the D, E, F1 and F2 layers, and the topside (nominally at 90 km, 110 km, 200 km, 350 km, and > 450 km, respectively). The lower four regions each contain a local peak in the electron density. These peaks result from changes in neutral atmosphere's density and composition with altitude. The altitude ranges, density and even existence of each of these layers varies with time of day, season, solar activity level, geomagnetic storm level and latitude. Figure 2-3 shows a sample mid-latitude electron density

profile at noon and midnight during high and low solar activity periods.

2.6.1 Layer Creation

Many processes create the ionosphere: solar X-ray and UV radiation ionizes air molecules; high-energy solar and extra-solar cosmic rays that pass through the atmosphere leave a trail of ionized air molecules; radioactive substances in the Earth's crust decay, providing ions; and electric charges separate in clouds. Most ionizing solar radiation is absorbed above 90 km, and the photoionization of N_2, O_2, and O creates the E- and F-layers of the ionosphere. The D-layer is created by radiation around 0.12 μm, which is able to penetrate to around 60 km altitude where it ionizes nitric oxide (NO). Tascione [23] provides a complete description of the chemical processes that take place in the ionosphere.

The charged particles persist in the ionosphere, not recombining immediately with oppositely charged particles, for two reasons. First, most unstable species of ions are too heavy to exist at the altitudes of the ionosphere. Second, at ionospheric altitudes the atmosphere has thinned out sufficiently that electrons and ions are not often colliding and recombining. The mean free path of free electrons is higher, and thus their lifetimes are much longer than those at lower altitudes [21].

2.6.2 Variations in the Ionosphere

As with the neutral atmosphere, it is important to understand how the ionosphere varies and which variables the ionosphere electron density profile is a function of. Only then can we asses the applicability of the ray-tracing results presented in chapter 6, and the motivation behind which independent variables were chosen to examine variations in signal delays.

Note two features in figure 2-3: the variation of the ionosphere with time of day and solar activity level.

With no solar radiation to replenish what ions and electrons do recombine, the D and F1 layers disappear at night and the E layer becomes negligible. The F2 density decreases by a factor of two as the night progresses, but does not reach zero. The layers build themselves back up as the day progresses, with peak F2 density occurring around 1 to 3 p.m. local time. This trend can be disrupted by winds. Equator-ward winds raise the altitude of the F2 peak. At these higher altitudes the air is less dense and the ions and electrons recombine at a slower rate. It is more common, however, for winds to blow pole-ward. In this situation, plasma travels down magnetic field

lines to lower altitudes, decreasing the F2 density and increasing the D- and E-region densities.

We also see in figure 2-3 that the ionosphere has more than four times the electron density during periods of high solar activity in the sun's eleven year cycle than it does during periods of low solar activity. Solar flares cause above-average levels of X-rays to strike the Earth, increasing the ionization of particles. Solar disturbances increasing the speed and density of the solar wind can also cause "anamolous currents" in the ionosphere and introduce high-energy particles into the polar ionosphere.

Events called geomagnetic storms can increase the electron density of the ionosphere by approximately 10%. These storms occur approximately one day after a solar flare and last for a few days. They are centered around the auroral latitudes (67° geomagnetic) but have a global effect increasing electron densities. The storm ring current weakens the geomagnetic field, which in turn slows the vertical transport of gases at the equator, allowing the equatorial latitudes to retain the electrons and ions created there. The effect on electron densities is most significant at night. [23]

A more localized form of ionospheric storm is induced by a solar flare. When protons from an energetic flare reach the geomagnetic field, they are directed by field lines to the magnetic poles where they penetrate the lower layers of the ionosphere and increase the density of the D-region. These storms are localized to the polar and auroral regions, are worst during the day, and last on the order of days. [23]

The ionosphere varies significantly with geomagnetic latitude. Near the poles, the ionosphere is coupled to the tail of the magnetosphere. The resulting interaction determines the ionosphere there. Near the equator, plasma instabilities and changes in the magnetospheric ring current affect densities, as was seen with geomagnetic storms. The magnetic field lines are horizontal at the equator, creating the equatorial electrojet: a sheet of current that flows around the equator at 100 km altitude, driving plasma convection. Plasma from the electrojet is raised to higher altitudes by the strong daytime vertical transport present at the equator. Once at these altitudes, the plasma is carried pole-ward by magnetic field lines. This process causes the F2-region electron density to peak with respect to latitude between 20° and 30° geomagnetic latitude — not closer to the equator as would be expected from the solar zenith angle. This variation is known as the "equatorial anomaly" [23].

The electron density profile also depends on the distribution of neutral particles. In the middle latitudes, the F2-region peak occurs where the photoionization of atomic oxygen balances the recombination of electrons and ions. Any change in the ratio of atomic oxygen to molecular oxygen and nitrogen can change the altitude of this peak.

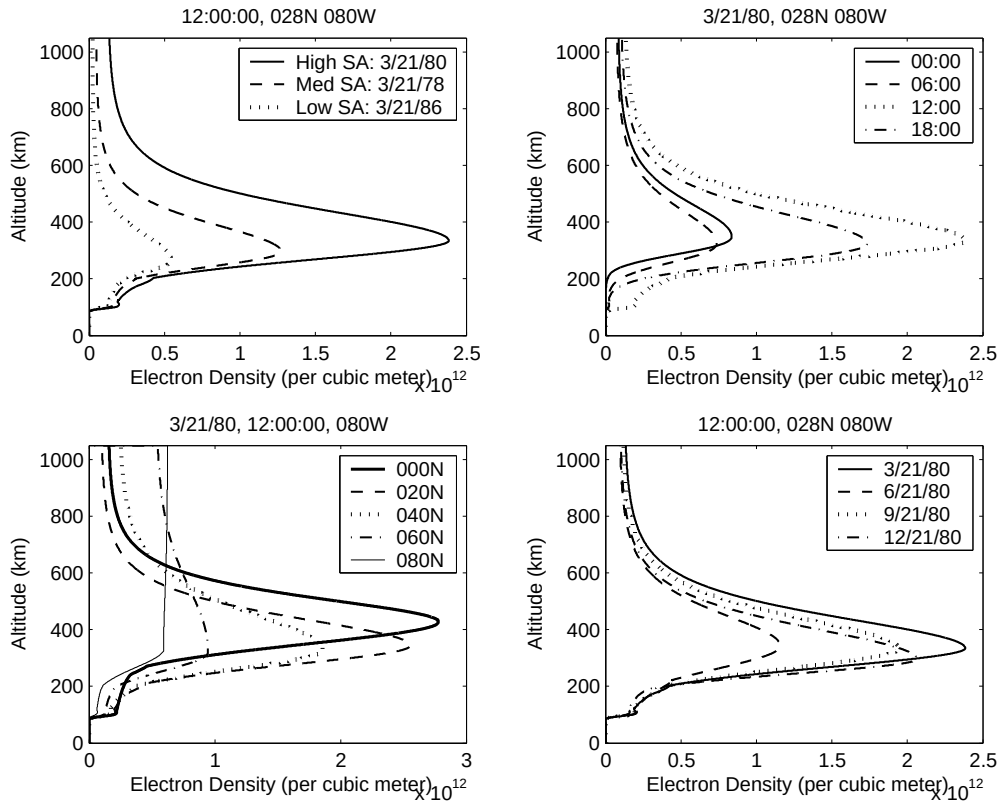

Figure 2-4: The IRI-2001 electron density profile. TOP LEFT: Electron density on a high, medium and low solar activity, spring equinox day. TOP RIGHT: Electron density as a function of time of day. Not only does the density change, but so does the altitude of the F2-layer peak and the existence of the lower layers. BOTTOM LEFT: Electron density as a function of latitude. BOTTOM RIGHT: Electron density as a function of season.

The ionosphere varies seasonally. The density of the E- and F1-regions peak in the summer, and the density of the F2-region peaks in winter. Varying the season can change densities by as much as a factor of four. The variation is largest in the mid-latitudes [23].

There are many other variations in the ionosphere that are either less predictable and have a major effect, or are predictable but have a small effect. Electron density can change with atmospheric tides caused by the Sun and the Moon; solar flares can cause a temporary (on the time-scale of an hour or two) increase in the density of the D-region known as a sudden ionospheric disturbance, causing the ionosphere to absorb radio waves that normally would pass through; thunderstorms, meteor showers and periods of high geomagnetic activity can cause a "sporadic" E-layer of variable density, height and size; and auroral heating can send traveling ionospheric disturbances equator-ward. The ionosphere is not predictable: densities can vary 20–25% from the monthly mean at the 1 sigma level [20, 29].

2.7 Signal Propagation Features

2.7.1 Scintillation

If the electron density varies rapidly along the signal path, the signal path will follow a seemingly erratic path and defocus. The signal scintillates, its amplitude and phase randomly varying rapidly. The atmospheric phenomena that causes scintillation, "spread F", occurs primarily in the night-time equatorial anomaly latitudes (20°–30°) and the polar latitudes and is most common during the equinoxes. When the ionosphere experiences "spread F", the F region spreads to encompass a larger range of altitudes and its density can vary by a factor of 100–1000 on scales of a few kilometers, although scales as short as tens of meters have been observed [23].

Scintillation is particularly harmful to GPS users aboard spacecraft. Since both the GPS satellite and the spacecraft are moving on the order of kilometers per second, the path the signal takes at one second is shifted by kilometers from the path the signal takes the next second. If the F-region is experiencing "spread F", the signal will pass through irregularities extremely rapidly, and the changes in incoming phase and amplitude of the GPS signal will cause loss of signal lock. This phenomena was not included in the ionosphere model used in this thesis and therefore was not examined.

Condition	Phase Refractivity Gradient (N/km)
Trapping	$\frac{dN}{dh} \leq -157$
Superrefractive	$-157 < \frac{dN}{dh} \leq -79$
Standard	$-79 < \frac{dN}{dh} \leq 0$
Subrefractive	$0 < \frac{dN}{dh}$

Table 2.4: Phase refractivity gradients for each signal propagation condition, as specified by Hitney et al. [4]

2.7.2 Superrefraction and Ducting

Superrefraction and ducting can cause a signal to deviate significantly from the line-of-sight signal path, greatly increasing the signal delay.

In a standard, spherically symmetric atmosphere, phase refractivity typically decreases monotonically with altitude until the F2 peak of the ionosphere, bending a GPS signal toward the Earth. The extent of this bending depends on the gradient of the phase refractivity (see table 2.4).

If the refractivity in this spherically symmetric atmosphere decreases especially rapidly with altitude ($-157 < \frac{dN}{dh} \leq -79$, where h is in km), the signal will be superrefracted: it will bend toward the Earth more quickly than normal, although not at a rate greater than the Earth's curvature. Superrefraction will lengthen the path of a signal, and for horizon-glancing signals, help carry the signal below the horizon [4].

If the refractivity decreases even more rapidly with altitude ($\frac{dN}{dh} \leq -157$), a signal will be "trapped": it will bend toward the Earth at a rate greater than the Earth's curvature. Refractivity gradients of this magnitude form ducts, or vertical regions where a signal can be trapped, bouncing off of the top and bottom of the duct, only able to travel horizontally. Since trapped signals are no longer able to radiate as in free space, signals traveling in ducts experience very little power loss, and since they do not leave the atmosphere, are able to travel very long distances, potentially circumnavigating the globe [4].

Ducts are best described using the modified refractivity, M:

$$M = N + \frac{h}{R_{Earth}} \times 10^6 \approx N + 157\ h \tag{2.19}$$

where h is altitude in kilometers and R_{Earth} is the Earth's radius in kilometers. Stephanakis [30] provides a derivation of M, based on the idea that two signals at different altitudes bending with the same curvature as the Earth will traverse equal

arcs when those arcs are expressed in wavelengths. In non-trapping conditions, M increases monotonically with altitude. However, a trapping layer will decrease the value of M. The region extending from the local minima of M down to the altitude where that value of M last existed is a duct [4], and the region where $\frac{dM}{dh} = 0$ produces around-the-world propagation of radio waves. A stable around-the-world propagation of the signal occurs when the region where $\frac{dM}{dh} = 0$ coincides with a local maximum in M (and roughly the center of the duct), since the refractive gradient on either side of this region will tend to push the signal back toward the region of maximum M [30]. This means that even though a trapping layer occurs at altitude, the resulting duct may extend all the way to the surface of the Earth if the trapping layer decreases M sufficiently [4].

Since only the magnitude of the phase refractivity in the ionosphere is frequency dependent, the location of ducts in the ionosphere will not be frequency dependent, but their effectiveness will be. The ionosphere is likely to have significant horizontal refractivity gradients, especially at the day-night terminator, which will affect the propagation of signals.

Many situations, in both the neutral atmosphere and the ionosphere, can cause ducts to form. In a modeled atmosphere such as the Global Reference Atmosphere Model 1999 (GRAM-99) atmosphere, we don't typically see these duct-forming features. Note that in figure 2-5, the GRAM-99 profile has only standard refraction, and the International Reference Ionosphere 2000 profile only has standard and subrefractive regions.

In some geographical regions, however, the typical profile will rarely look like the the standard profile, and ducts often form. Evaporation over the oceans creates a layer water vapor that quickly drops off to drier air, creating a surface duct anywhere from 2 to 40 meters high. Hitney [4] notes that the frequencies affected by these surface ducts are height dependent, but typically only those 3 GHz and higher are affected, sparing GPS signals. Elevated neutral atmosphere ducts, occurring anywhere from slightly above the surface to 6 km altitude (although mostly below 3 km altitude), affect signals 100 MHz and higher, including GPS signals [4]. The high variability of the ionosphere suggests that ducts can form for many reasons.

Just because a duct exists does not mean that a signal will necessarily become trapped in it. Signals must enter the duct at grazing angles in order to become trapped.

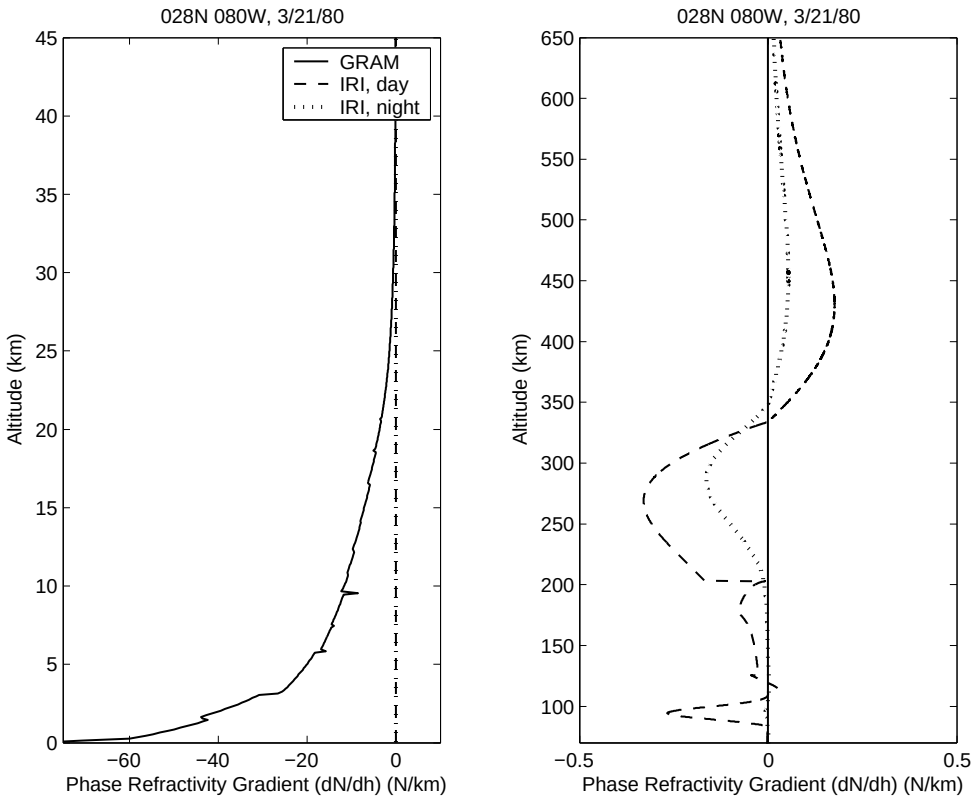

Figure 2-5: Phase refractivity gradient in both the GRAM-99 and IRI-2000 models. Note that the GRAM has only standard refraction, and the IRI has both standard and subrefractive regions, due to the negative phase refractivity.

Chapter 3

Existing Methods to Correct GPS Signal Delay

Methods currently exist to correct for both neutral atmosphere and ionosphere signal delay. However, most of these methods have been designed with ground-based receivers in mind and are often not suitable for receivers aboard space vehicles.

The only way to correct for neutral atmosphere signal delay is with a delay model (section 3.1.2). Models designed for receivers on the surface of the Earth can correct for neutral atmosphere signal delay with centimeter level accuracy if provided with detailed conditions, such as surface temperature, pressure, local tropopause height, troposphere lapse rate and the water vapor distribution [31]. This information can be obtained in real-time by downloading current numerical weather prediction model results off the internet. If the water vapor distribution is not known, these delay models are only accurate to within 20–30% of the true value. Models for receivers aboard commercial aircraft will potentially be more accurate since the aircraft is above the majority of the water vapor and is likely only observing satellites at positive elevation angles. No models exist for receivers aboard space vehicles.

The ionosphere signal delay can also be corrected with an on-board delay model (section 3.1.1), and this is the primary method of correction for single-frequency ground-based receivers. Methods have been developed for single-frequency ground-based users to directly estimate the ionosphere signal delay (sections 3.3 and 3.4), but the most accurate ionosphere correction is reserved for those users with access to the code of the L2 signal, who can take advantage of the frequency dependence of the ionosphere signal delay and perform dual-frequency correction (section 3.2)[1].

[1]It is anticipated that future GPS satellites will transmit C/A code at both L2 and L5 (1.17645 GHz), in addition to at L1, allowing all users to implement dual frequency correction.

Dual-frequency correction is the only method that is completely compatible with space-based receivers. If a space-based dual-frequency receiver does not have sufficient channels to constantly monitor each satellite on two frequencies, a modification to the single-frequency delay estimation method will allow it to still use the dual-frequency method.

3.1 On-board Receiver Models

Signal delay models typically model the zenith signal delay above the receiver, and then scale that delay according to the elevation angle of the satellite. For the neutral atmosphere, the signal delay model takes the form:

$$\Delta^{na} = \Delta^{na}_{d,zen} \times m_d(E) + \Delta^{na}_{w,zen} \times m_w(E) \tag{3.1}$$

since the distribution of the water vapor contributing to the "wet" terms in the delay differs from the distribution of the "dry" atmospheric gases which cause the remaining delay. For the ionosphere, the model takes the form:

$$\Delta^{i} = \Delta^{i}_{zen} \times m(E) \tag{3.2}$$

We can immediately see that these mapping functions will run into three problems in a receiver aboard a space vehicle. First, even though the zenith delay can be modified accordingly as the receiver gains altitude, the mapping function must also change with receiver altitude to account for the changing geometries (i.e. $m(E) \rightarrow m(E,h)$). Second, as the receiver gains in altitude it will be able to see satellites at negative elevation angles. For the neutral atmosphere, the delay calculation can no longer maintain the form of eqn. 3.1, because the zenith delay no longer contains contributions from all of the layers of the atmosphere that the signal may be passing through. For an extreme example, consider a receiver above the atmosphere, looking down at a satellite near the horizon. Even though there is no zenith neutral atmosphere signal delay, there will be a large neutral atmosphere signal delay. (See figure 3-1.) For the ionosphere, equation 3.2 will break down in a similar manner once the receiver reaches ionospheric altitudes ($\approx$ 60 km). Third, as a receiver gains altitude, the signal will potentially be traveling through the atmosphere at points farther and farther away from the receiver. For a ground user, a signal to a satellite low on the horizon can pass through the thickest part of the ionosphere 14° (Earth centered angle) away from the receiver [1]; for a receiver at 1700 km altitude, a signal whose closest approach to

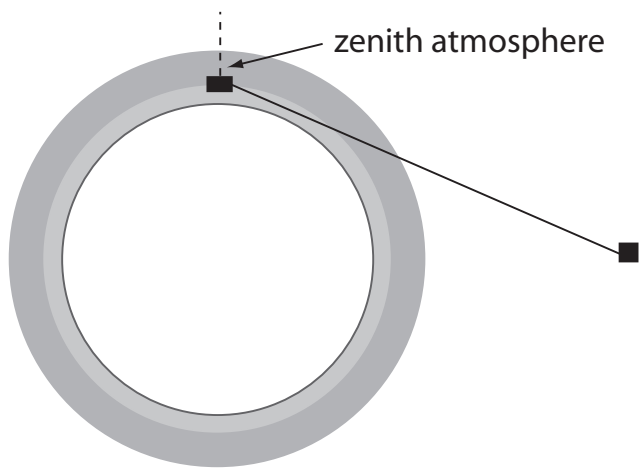

Figure 3-1: The zenith delay does not contain contributions from all of the layers of the atmosphere that the signal passes through. Figure is not to scale.

the Earth occurs at 350 km altitude will pass through this part of the ionosphere 34° (Earth centered angle) away from the receiver. The farther the relevant ionospheric or neutral atmosphere "pierce point" is from the receiver, the more the conditions at these points will differ from those at the receiver, and so the zenith delay may no longer be adequate. As the pierce point becomes farther and farther away from the receiver, the manner in which the conditions differ will depend more and more on the azimuth to the satellite, complicating any potential mapping function. For these reasons, most delay calculations assume that the receiver is on the ground and cannot see satellites at negative elevation angles.

3.1.1 Models for the Ionosphere Signal Delay

Likely due both to the existence of dual frequency correction and to the high variability of the ionosphere, the ionosphere signal delay has only one commonly used correction model. Klobuchar developed this model in the 1970s to correct for at least 50% (in practice $\approx$ 60%) of the ionosphere signal delay *for a ground user*, on the argument that trying to correct for any more of the delay would require a significantly more complex model and more frequent updates, while only providing at most a 75% correction[2] due to the variability of the ionosphere. He found that a half cosine of

[2] "Percent correction," as used by Klobuchar, is one minus the ratio of the rms of the residual delay after correction to the rms of the delay without correction.

the form:

$$\Delta^i_{zen} = A + B\cos\left(\frac{2\pi\ (LT - \phi)}{P}\right)$$

(where only the positive half of the cosine is used) fit the monthly median zenith ionosphere delay well. A is a constant offset term, and is the nighttime zenith delay value; B is the amplitude of the delay half-cosine; LT is the local time at the receiver and ϕ is the local time of peak ionosphere delay, both in hours; and P is the period of the half-cosine, and is not necessarily 24 hours. (Since only the positive half of the cosine is fit to data, the cosine does not have to have a 24 hour period.)

Klobuchar fit his delay model to delays calculated by ray-tracing through the Bent ionosphere model for multiple latitudes, every 10 days for one year, for multiple solar flux conditons. To conserve coefficients, he determined that ϕ and A varied much less than B and P, and set them to 14:00:00 local time and 5 ns (1.5 m), respectively. B and P were then each represented by four coefficients of a cubic polynomial in geomagnetic latitude[3]. In this manner, Klobuchar was able to incorporate many of the variables that the electron density (and therefore ionosphere signal delay) depends upon. He incorporated time of day and geomagnetic latitude explicitly into his algorithm, and seasonal, solar activity and storm related variations are accounted for in the eight changing coefficients.

Klobuchar used a relatively simple mapping function:

$$m(E) = 1.0 + 16.0 \times (0.53 - E)^3 \tag{3.3}$$

where E is in semi-circles. Instead of using the zenith delay above the receiver, his algorithm calculates the latitude and local time of the mean ionospheric location: the point at which the signal passes through 350 km altitude, which is approximately the thickest part of the ionosphere and where most of the delay occurs.

The majority of single frequency GPS receivers use the Klobuchar model to correct for ionosphere signal delay. The eight coefficients describing B and P are broadcast by all GPS satellites and are updated by the GPS Control Segment anywhere from every day to every 10 days, depending on the current magnitude of variations in the ionosphere. The full Klobuchar algorithm can be found in many places [1, 20].

[3]Klobuchar did not use a more complex form than the cubic polynomial, as might be required in the equatorial anomaly region, because the Bent model did not cover the low-latitudes adequately. Therefore, we must conclude that the Klobuchar model is less accurate in the low-latitude regions that in the middle latitudes. The model is also less accurate in the highly variable polar regions.

3.1.2 Models for the Neutral Atmosphere Signal Delay

Many very accurate mapping functions for multiplying the neutral atmosphere zenith signal delay were created to aid in precise geodetic positioning. These mapping functions are for ground users (with some allowance for the Earth's surface being located above the geoid), and typically require user input of meteorological data such as surface dry pressure, wet pressure, temperature, height above the geoid and the temperature lapse rate in the troposphere. Mendes [31] compares the five most accurate mapping functions at low elevation angles, and found that they were accurate on the level of a few centimeters at $6° - 10°$ elevation angle (assuming a "truth" zenith delay input). These mapping functions often rely upon zenith delay calculations by Saastamoinen [25] or Davis [9] in order to calculate the signal delay. Davis' calculation of the dry zenith delay relies upon knowledge of the geodetic latitude, surface height above the geoid, and surface pressure. Davis also claims that Saastamoinen's calculation of the wet zenith delay, based upon average profiles, could introduce a 10–20% error in wet delay, but he does not provide an alternate equation and suggests using a water vapor radiometer to measure the water vapor distribution in real-time.

These very accurate models are not suitable for a receiver aboard a spacecraft for two reasons. First, they do not account for the altitude of the spacecraft and the resulting possibility of negative elevation angles. Second, they require the meteorological conditions, such as temperature, pressure or refractivity, nominally at the receiver. Since these models assume the receiver is on the surface of the Earth, the conditions at a receiver aboard a space vehicle are not suitable replacements. Conditions on the surface of the Earth beneath the signal's point of closest approach would be most appropriate, but would be difficult for a space-based receiver to obtain in real-time.

The following four models have limited utility in GPS receivers used during the launch and reentry of a space vehicle – none requires surface conditions, and all can accommodate receivers at some altitude. The four are of varying complexity and accuracy. Altshuler and Kalaghan have the most computationally complex equation, valid to 5° elevation angle and an unknown receiver altitude. Altshuler later derived an alternate expression, accurate to within 6.1% of the true neutral atmosphere signal delay and valid down to 5° elevation angle and up to 20 km receiver altitude. The third model presented is the NATO Standard Troposphere Model, which uses a modified zenith delay calculation with a ground-user mapping function. Finally, I present a computationally simple (and thus inaccurate) equation from Mendes [18] which Mendes claims is commonly used in GPS receivers.

Altshuler and Kalaghan Mapping Function

Altshuler and Kalaghan [32] took seasonal, 5-year averaged surface refractivity values from 268 stations, extended the values into three dimensions using the CPRL Reference Atmosphere 1958 Refractivity Model, and ray-traced through the resulting profiles to create a neutral atmosphere signal delay model:

$$\Delta^{na} = m(E)H(h)F(h, N_s) \tag{3.4}$$

where N_s is the refractivity at the surface of the Earth beneath the receiver, E is the elevation angle in degrees, h is the height of the receiver above sea level in thousands of feet, and the resulting signal delay is in feet.

$$\begin{aligned}
m(E) &= \left(0.1556 + \frac{138.8926}{E} - \frac{105.0574}{E^2} + \frac{31.5070}{E^3}\right)\left[1.0 + 10^{-4}(E - 30.0)^2\right] \\
H(h) &= \left[0.00970 - \frac{2.08809}{h + 8.6286} + \frac{122.73592}{(h + 8.6286)^2} - \frac{703.82166}{(h + 8.6286)^3}\right] \\
F(h, N_s) &= 3.28084 \\
&\quad \times \left[\frac{6.81758}{h + 3.28084} + 0.30480(h + 3.28084) + 0.00423N_s - 1.33333\right] \\
&\quad \times \left[1 - 1.41723 \times 10^{-6}(N_s - 315.0)^2\right]
\end{aligned}$$

While the equations are only valid for $E \geq 5°$, there is no stated limit on the height of the receiver, thus providing a formulaic option for space-based signal delay calculations with $E \geq 5°$. If the refractivity beneath the receiver is not known, Altshuler and Kalaghan provide the following equation:

$$\begin{aligned}
N_s(h, \phi, M) &= 369.03 + (-)0.01053h + (-)0.92442\phi + (-)0.00160h \sin^2\left(\frac{\pi M}{12}\right) \\
&\quad +(-)0.19361\phi \sin^2\left(\frac{\pi M}{12}\right) + 0.00063h \cos\left(\frac{\pi M}{12}\right) \\
&\quad -0.05958\phi \cos\left(\frac{\pi M}{12}\right)
\end{aligned}$$

where ϕ is the receiver's latitude, in degrees, M is the calendar month, and h is now in feet. They claim this value is correct to within 5% of the mean refractivity value. (-)'s are indicated where Spilker [13] disagrees with the signs presented in [32]; while Altshuler and Kalaghan do change notational convention in reporting this last equation, the sign changes are not noted in the Altshuler and Kalaghan errata sheet, and Spilker made numerous other typographical errors in reporting the above

equations.

The stated standard deviation of the delay given N_s is unknown (but h, M, and ϕ are known) range from 0.14 m at zenith to 1.56 m at $E = 5°$. Spilker claims that this delay calculation has errors of 3.7% if surface refractivity is known, and 6% if it is not and the equation for N_s is used. Recall that Altshuler and Kalaghan extended their surface values with the CPRL refractivity profiles, and therefore do not reflect the full range of possible water vapor, temperature or pressure profiles. These errors only reflect the statistical variation in surface refractivities.

A somewhat simpler form of this equation was later derived for use with the FAA's Wide Area Augmentation System [18].

Altshuler Delay Model

Altshuler [33] later derived an alternate delay model for aircraft flying up to 20 km altitude based on the model of Bean and Dutton[4]. He modified their model so that it fit Millman's[5] signal delays (calculated by ray-tracing through CPRL Reference Atmosphere 1958 refractivity profiles up to 30 km):

$$\Delta_{gu}^{na}(E, N_s) = \frac{1.4588 + 0.002961 N_s}{\sin E} - \left[0.001786(N_s - 360)^2 + 89.4\right] E^{-2.3}$$

where E is elevation angle in degrees[6]. Millman also calculated the signal delay occurring between sea level and aircraft altitudes; by subtracting an equation for this delay from the total signal delay from the ground to the satellite, Altshuler created an equation for users at altitude:

$$\Delta^{na}(E, N_s, h_{rcvr}) = \Delta_{gu}^{na} \times \exp\left(-\left[(0.0002N_s + 0.070)\, h_{rcvr} + \left(\frac{0.83}{N_s} - 0.0017\right) h_{rcvr}^2\right]\right) \tag{3.5}$$

where h_{rcvr} is the altitude of the receiver in kilometers and N_s is the surface refractivity beneath the receiver. If the surface refractivity is not known, Altshuler provides the following equation:

$$N_s = 369.0 - |\phi| \left[0.9244 - 0.1936 \sin^2\left(\frac{\pi doy}{365}\right) + 0.0596 \cos^2\left(\frac{\pi doy}{365}\right)\right]$$

[4]cited as Bean, B.R. and Dutton, E.J., *Radio Meteorology* New York: Dover, 1966.

[5]cited as Millman, G.H., "Tropospheric effects on space communications," in *AGARD Conf. Proc. 70 Troposph. Radio Wave Propogat.,* Düsseldorf, Germany, Sept. 1970, vol. AGARD CP-7071, p. 4.

[6]NOTE: The second term of this equation is stated incorrectly in the text of [33]; the term is correct as derived in Appendix A.

where ϕ is latitude in degrees and *doy* is the day of year. If the receiver is in the southern hemisphere, the day of year should be replaced with $\mod (doy + 182, 365)$.

Using the surface refractivity equation (as would be required in most receivers), Altshuler quotes 1σ errors of 6.1%. Again, these errors only reflect the statistical variation in surface refractivities. This model is designed for elevation angles down to 5°.

NATO Standard Troposphere Model

The NATO Standard Troposphere Model uses the Chao[7] ground-based dry mapping function with a modified total zenith delay, thus introducing errors on two fronts. Not only is the dry mapping function used to map a delay containing wet components, but the mapping function is not adjusted for the changing geometries at altitude:

$$\Delta^{na}(E, h_{rcvr}, N_s) = \Delta^{na}_{zen} \times \frac{1}{\sin E + \frac{0.00143}{\tan E + 0.0455}} \tag{3.6}$$

where

$$\begin{aligned} c &= \ln\left(\frac{N_1}{105}\right) \\ N_1 &= N_s + \Delta N \\ \Delta N &= -7.32 \exp\left(0.005577 N_s\right) \end{aligned}$$

$$\begin{aligned} \Delta^{na}_{zen} &= 1.430 + 0.732 + (0.5\Delta N - (N_s h_{rcvr} + 0.5\Delta N h^2_{rcvr})) \times 10^{-3} && 0 \le h_{rcvr} \le 1 \\ \Delta^{na}_{zen} &= 0.732 + \left(\tfrac{-8N_1}{c}\exp(-c) + \tfrac{8N_1}{c}\exp\left(0.125(1 - h_{rcvr})c\right)\right) \times 10^{-3} && 1 \le h_{rcvr} \le 9 \\ \Delta^{na}_{zen} &= \left(\tfrac{105}{0.1424}\exp\left(-0.1424(h_{rcvr} - 9)\right)\right) \times 10^{-3} && h_{rcvr} \ge 9 \end{aligned}$$

where h_{rcvr} is in kilometers, N_s is the mean sea-level surface refractivity and E is the elevation angle.

A Common Model

The following model is used in many GPS receivers (see [18] for references):

$$\Delta^{na}(E, h_{rcvr}) = \frac{2.4224}{0.026 + \sin E} \exp(-0.13345 h_{rcvr}) \tag{3.7}$$

[7] cited as Chao, C.C. "A model for tropospheric calibration from daily surface and radiosonde balloon measurement." Technical Memorandum 391-350, Jet Propulsion Laboratory, Pasadena, CA.

where h_{rcvr} is in km. Two main principles are used in this equation: the mapping function uses the $\frac{1}{\sin E}$ flat earth approximation and the zenith delay decreases exponentially with altitude. Altshuler [33] cites that refractivity (and hence zenith delay) decreases linearly for the first kilometer above the surface and then exponentially. At 9 km above the surface, the rate of exponential decrease changes. Both of these approximations limit the accuracy of this equation; Braasch [34] claims that it corrects for 90% of the signal delay, but based on the accuracies of more complex models, it may not even do this well.

3.2 Dual Frequency Ionosphere Delay Correction

Since the ionosphere is dispersive, the ionosphere signal delay is frequency dependent. This dependency allows us to measure the signal on two separate frequencies and calculate the ionosphere delay, and is the main reason why GPS transmits on both the L1 and L2 frequencies.

If we rewrite equation 2.3 in terms of frequency dependent ($\Delta I_r^s(f)\ c$) and frequency independent ($\delta_r^s\ c$) errors, we have:

$$P_r^s(f) = \rho_r^s + \Delta I_r^s(f)\ c + \delta_r^s\ c \tag{3.8}$$

Combining equations 2.14 and 2.4, we have:

$$\begin{aligned} \Delta I_r^s(f)\ c &= \int_{rcvr}^{sat} \left(1 + \frac{40.3}{f^2} N_e\right)\ ds - \int_{rcvr}^{sat} 1\ dl \\ &= \left[\int_{rcvr}^{sat} 1\ ds - \int_{rcvr}^{sat} 1\ dl\right] + \frac{40.3}{f^2} \int_{rcvr}^{sat} N_e\ ds \end{aligned}$$

If we approximate that the paths the L1 and L2 signals take are the same and write the terms that are not frequency dependent as constants,

$$\Delta I_r^s(f)\ c = \alpha_r^s + \frac{\beta_r^s}{f^2}$$

Subtracting the pseudoranges at the two frequencies leaves us:

$$P_r^s(L1) - P_r^s(L2) = \beta_r^s \left(\frac{1}{f_{L1}^2} - \frac{1}{f_{L2}^2}\right) \tag{3.9}$$

allowing us to solve for one of the ionospheric terms, β, and remove it from the

pseudorange calculation. Equivalently,

$$P_r^s(L1)f_{L1}^2 - P_r^s(L2)f_{L2}^2 = (\rho_r^s + \delta_r^s\ c + \alpha_r^s)(f_{L1}^2 - f_{L2}^2) \tag{3.10}$$

allowing us to solve for the pseudorange involving mostly non-frequency dependent quantities ($\rho_r^s + \delta_r^s\ c + \alpha_r^s$). Note that the ionospheric term is not completely eliminated in either case: α, the difference in length between the line-of-sight and actual signal path due to the ionosphere, remains unknown and not corrected for. Also note that the paths the L1 and L2 signals take are not in fact the same: the path the L2 signal takes will bend more and therefore be longer, since it is at a lower frequency and thus has a higher refractivity.

The phase pseudorange is corrected in a similar manner.

3.3 Single Frequency Ionosphere Delay Correction

A few approaches have been proposed to remove the ionosphere signal delay using only the L1 frequency. These approaches rely upon the fact that the code is delayed by approximately the same amount as the carrier phase is advanced, and therefore the ΔI_r^s term in both the code and phase pseudorange equations are approximately identical. (See section 2.4 for discussion of this approximation.) If we subtract the phase pseudorange

$$(P_\phi)_r^s(f) = \rho_r^s - \Delta I_r^s(f)\ c + (\delta_\phi)_r^s\ c + N\lambda \tag{3.11}$$

from the code pseudorange (eqn. 3.8) we are left with

$$(P_{code})_r^s(f) - (P_\phi)_r^s(f) = 2\ \Delta I_r^s(f)\ c + [(\delta_{code})_r^s - (\delta_\phi)_r^s]\ c - N\lambda \tag{3.12}$$

Written more succinctly, with the difference in frequency-independent errors reducing to ϵ and all errors written in terms of length delay, for a given satellite-receiver pair:

$$P_{code} - P_\phi = 2\ \Delta I + \epsilon - N\lambda \tag{3.13}$$

If N, the phase cycle integer ambiguity, can be calculated and continuously tracked (without cycle slips), we can continuously estimate the ionospheric delay.

Cohen, Pervan and Parkinson [35] propose a method in which the ionosphere term in equation 3.13 is split into a zenith delay component and a obliquity mapping func-

tion, and incorporating the elevation angle to each satellite and a local ionosphere model (essentially taking into account the variation in delay based on azimuth in addition to elevation angle) are able to separate N from ΔI. Trethewey, Catchpole and Hansla [36], and later Trethewey and Nisner [37], built upon this method and developed a Kalman filter to estimate ionospheric delay. These approaches are explicitly developed for ground receivers and are not suitable for spacecraft, although it is possible that they could be modified to form an estimate of the delay for space users.

There are approaches that would be suitable for spacecraft, however. If a dual frequency receiver does not have sufficient channels to continuously make a dual frequency correction for every satellite, we can calculate ΔI during a period of dual frequency ionosphere correction and solve for N for that satellite. Then, provided N is continuously tracked without cycle slips, ΔI can be approximated when L2 tracking is subsequently dropped.

3.4 Differential GPS Ionosphere Delay Correction

The ionospheric delay can be removed from a signal using differential GPS. If the receiver remains sufficiently close to a receiver at a precisely known location so that both receivers are experiencing the same ionospheric conditions, the fixed receiver can transmit corrections to the mobile receiver [20]. This solution is obviously not feasible for space users.

[This page intentionally left blank.]

Chapter 4

Ray-Tracing Model Options

If we wish to understand the magnitude of signal delays to expect aboard a space vehicle, we must ray-trace signals through a model atmosphere. Ray-tracing allows us to see how the signal delay changes as a function of neutral atmosphere conditions, ionosphere conditions and geometry. We could examine ionosphere signal delays using a dual frequency GPS receiver aboard a real space vehicle; however, since the neutral atmosphere is not dispersive, we cannot accomplish similar measurements of neutral atmosphere delay. Ray-tracing through a model atmosphere instead of data collection aboard a space vehicle has many other advantages: it is cheaper; it is easier to explore how different parameters affect the delay (for example, we do not have to wait 11 years to sample the ionosphere delays through all levels of solar activity); and it gives us representative, statistical signal delays, as opposed to delays based on a unique set of conditions. Knowing the delay for a very specific situation can be useful. For example, there are features, like ionosphere scintillation, that are common in some areas but not present in the average. Scintillation can have a significant effect on a signal's path, and the delay features that result will affect the design of receiver filters and algorithms. However, I have chosen to only provide statistical results (an average and a standard deviation) since these features are unpredictable.

There are many atmosphere and ionosphere models through which we could ray-trace GPS signals. Since I use equations 2.10 and 2.12 with equation 2.7 to ray-trace delays, I require a neutral atmosphere model that provides temperature, pressure and water vapor pressure as a function of altitude, and an ionosphere model that provides electron density as a function of altitude. As was seen in section 2.5.2, receiver latitude, season and time of day can all affect the neutral atmosphere temperature and pressure profiles, and therefore can affect the signal delay. Therefore, any acceptable neutral atmosphere model must be able to vary as a function of these conditions in

order to fully describe the range of possible neutral atmosphere delays. Likewise, in section 2.6.2 we saw that the electron density profile of the ionosphere varies with receiver geomagnetic latitude, season, time of day, solar activity level and geomagnetic storm conditions. Any acceptable ionosphere model must be able to vary as a function of these conditions in order to fully describe the range of possible ionosphere delays. The following two sections discuss existing neutral atmosphere and ionosphere models. Of those available, I found that the Global Reference Atmosphere Model (1999) produced at the NASA/Marshall Space Flight Center and the International Reference Ionosphere Model (2001) maintained by the International Union of Radio Science (URSI) and the Committee on Space Research (COSPAR) satisfied all of these requirements and were freely available. I have used these two models in my ray-tracing analysis.

4.1 Neutral Atmosphere Model and Profile Options

As was stated above, I am seeking a neutral atmosphere model that provides the temperature, pressure and water vapor pressure as a function of altitude, and can vary as a function of latitude, season and time of day. There are three categories of models that might satisfy these requirements: numerical weather prediction models, global empirical models and standard atmosphere profiles / empirical profiles. The models in these categories vary in accuracy and generality. Numerical weather prediction models will give precise values and accurately model realistic conditions in the atmosphere, but are only relevant for a very specific set of conditions. Standard atmosphere models are described by U.S. Committee on the Extension to the Standard Atmosphere (COESA) in *The U.S. Standard Atmosphere, 1976* [38] as a "hypothetical vertical distribution of atmospheric temperature, pressure and density which, by international agreement, is roughly representative of year-round, mid-latitude conditions." Standard atmospheres will thus be general and not necessarily capture features of the atmosphere that often occur, but are not present in the average.

For the ray-tracing performed in this thesis, my goal was to pick a model that spanned the features of the above models. The model should both represent typical conditions and ideally contain features that may often affect the ray-traced delay significantly but are not present in the average. For this reason, I chose to use a global empirical model for my model atmosphere. Global empirical models incorporate mea-

sured data with detailed physical models to create a model with an accuracy that lies between that of a numerical weather prediction model and that of a standard atmosphere. Numerical weather prediction models were too specific and computationally intensive (requiring a super computer) for the space vehicle trajectory studied, and standard atmospheres were too averaged and general.

4.1.1 Global Empirical Models

Global empirical models use historical data from radiosondes, satellites, rockets, radar and other measurements to model the atmosphere. Monthly mean atmospheric variables such as temperature, pressure, density and species concentrations are often given as a function of altitude, longitude, latitude and month or season. Many of the models only characterize one altitude region of the atmosphere, but a few combine these models to create super-models of the atmosphere. It is these super-models that I am ultimately interested in. Below I have briefly described three available super-models. Further information on these and their component atmosphere models can be found in the AIAA *Guide to Reference and Standard Atmosphere Models* [39].

NRLMSISE-00

The Naval Research Lab Mass-Spectrometer-Incoherent-Scatter (2000) (NRLMSISE-00) model uses Middle Atmosphere Program (MAP) data, National Meteorological Center (NMC) (now the National Centers for Environmental Prediction (NCEP)) averages and a number of other empirical measurements to describe temperatures and densities from the surface of the Earth up to the thermosphere as a function of year, day of year, time of day, altitude, latitude, longitude, F10.7 solar flux[1] and geomagnetic Ap index[2]. However, this model does not model water vapor, and so is inadequate for my analysis. This model can be downloaded from the National Space Science Data Center (NSSDC) at NASA/Goddard web page[3]. This web page also contains further infromation on the data sources for this model.

[1]The F10.7 solar flux is a measurement of the solar radiation density that reaches the radius of the Earth's orbit at a wavelength of 10.7 cm (2800 MHz). Solar radiation near this wavelength is responsible for much of the ionization in the ionosphere.

[2]The geomagnetic Ap index measures average global geomagnetic activity on a scale of 0 to 400. Despite this range, Ap values above around 30 indicate storm conditions.

[3]URL: http://nssdc.gsfc.nasa.gov/space/model/atmos/atmos_index.html

CIRA-86

The COSPAR International Reference Atmosphere (1986) (CIRA-86) provides atmospheric temperatures and densities from 0 km to 2000 km. It is based on a number of data sources which can be found on the NSSDC web page. The upper atmosphere region (above 120 km) of this model is very similar to the NRLMSISE-00 model. While it does model water vapor concentrations, the AIAA *Guide to Reference and Standard Atmosphere Models* [39] states that "the quality of the database describing some observables is variable." Further, this model does not provide standard deviations on the mean atmospheric parameters, making error analyses difficult. The model can be obtained from the NSSDC web page.

GRAM-99

The NASA/Marshall Space Flight Center (MSFC) Global Reference Atmosphere Model (1999) (GRAM-99) best meets the criteria laid out earlier in this chapter. It is a global atmosphere model that covers both seasonal and geographical variations, and extends from the surface of the Earth to orbital altitudes.

GRAM-99 uses the Global Gridded Upper Air Statistics (GGUAS) (based upon twice-daily measurements provided by European Center for Medium-Range Weather Forecasts (ECMWF)) to model the lower region of the atmosphere, MAP data to model the middle region of the atmosphere, and the Jacchia model as coded in the Marshall Engineering Thermosphere (1999) (MET-99) to model the upper atmosphere. These three models are smoothly combined using fairing techniques so that the output values and their first derivatives are continuous. Water vapor data is compiled from GGUAS data, MAP data, the Air Force Geophysics Laboratory atmospheric constituent profiles and the NASA/Langley Research Center water vapor climatology data. Combined, these models allow GRAM-99 to output monthly averaged atmospheric parameters (pressure, density, temperature, winds, water vapor concentrations and concentrations of 11 other atmospheric constituents) globally as a function of latitude, longitude, altitude, month, time of day, geomagnetic Ap index and F10.7 solar flux, covering altitudes from the surface of the Earth to orbital altitudes [40].

GRAM-99 performs both vertical and horizontal interpolation within data sets and fairing between data sets so that parameters vary smoothly in space, eliminating the grid nature of the data sets used to construct the model. Many of the data sets are incomplete near the poles. GRAM-99 interpolates to fill in these missing values as-

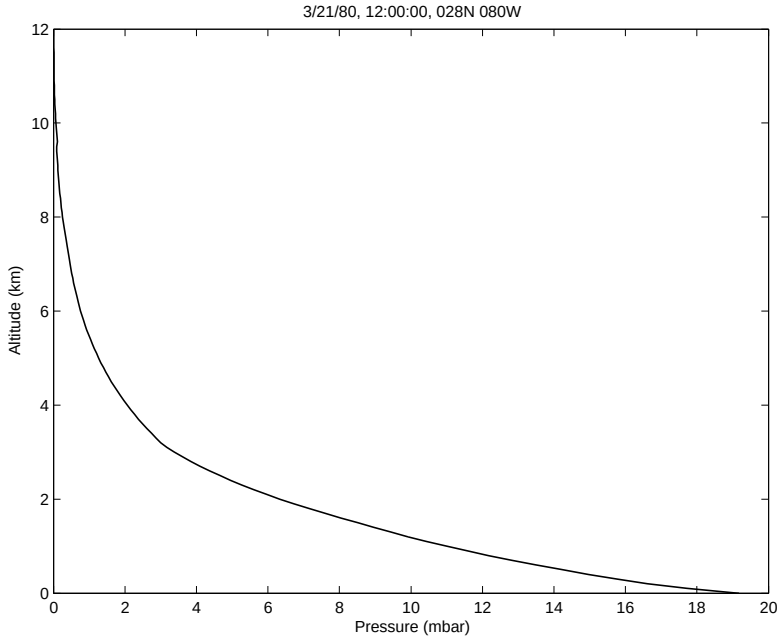

Figure 4-1: GRAM-99 water vapor pressure profile.

suming a parabolic variation across the poles. Many of the species concentration and water vapor databases have seasonal data. To provide monthly resolution, GRAM-99 converts the seasonal data to monthly data using either an annual harmonic variation fit to the seasonal data or a weighted average of adjacent seasonal data points [41].

Through the GGUAS database, GRAM-99 takes into account variations in atmospheric parameters due to local geographical variations. However, GRAM-99 also interpolates all values down to sea level using hydrostatic assumptions. This can cause some anomalous data points where the surface of the Earth is significantly above sea level (such as in the Andes and the Himalayas) [41].

GRAM-99 provides average water vapor pressures (see figure 4-1) and will therefore poorly represent typical water vapor distributions. This characteristic of GRAM-99 will affect the ray-traced delays of signals that pass close to the surface of the Earth (i.e. signals from satellites near the horizon). Signals that pass close to the surface of the Earth experience significant bending due to the large refractivity gradient there (see figure 2-5). A change in the water vapor distribution would affect the signal bending, and therefore the observed delay. GRAM-99 does not account for this potential variability. Ray-traced delays to satellites near the horizon will therefore have larger errors than will delays to satellites not near the horizon.

Users can optionally input Range Reference Atmosphere (RRA) data into GRAM-99, making the model more accurate at those locations for which RRA data is available. GRAM-99 transitions from the conditions specified in the RRA to the standard GRAM climatology using a weighted average.

GRAM-99 allows the user to add spatial and temporal perturbations, both large and small, to the atmosphere. These perturbations represent small-scale features such as turbulence, mesoscale storm processes and gravity waves, and large-scale features such as tides and Rossby wave processes. Small perturbations are modeled by a one-step Markov process. All large perturbations are formulated as cosine-waves, although their amplitude and wavelength may vary over the wave. The perturbations are scaled according to the standard deviation of the local value of the variable to be perturbed. The user can further scale all standard deviations by a constant factor. The phase of the cosine perturbations are chosen randomly by GRAM-99. The GRAM-99 perturbation feature is particularly useful for Monte Carlo analyses of the range of values derived from atmospheric conditions.

GRAM-99 is available upon request from the developers at MSFC. Contact information is available in the GRAM-99 technical memorandum [40].

4.2 Refractivity Models of the Neutral Atmosphere

Alternatively, we can calculate neutral atmosphere signal delays by ray-tracing through refractivity profiles instead of through refractivities calculated from atmospheric profiles.

The Central Radio Propagation Laboratory (CRPL) Reference Atmosphere (1958) provides refractivity profiles as part of its neutral atmosphere model, and was used in the ray-tracing work of Altshuler. This model has since been found to be inaccurate above 9 km altitude [33].

The Hopfield refractivity profile is often cited as a straightforward method to calculate atmospheric delays based solely on surface conditions, either via ray-tracing, integration of a mapped Hopfield profile or by mapping an approximate Hopfield zenith delay value onto the appropriate elevation angle. Many modified Hopfield approaches exist to calculate neutral atmosphere delays [7]. These modified approaches attempt to take the signal path out of the signal delay integral (eqn. 2.4), making the delay calculation simply an integral over altitude. The Hopfield profile has many inaccuracies which are detailed below.

Hopfield [3] simplified the ideal gas law extensions for T and P in the troposphere

in her two-quartic refractivity profile. For the dry component,

$$N_d \propto \frac{P}{T} = \frac{P_0}{T_0 + \beta h} \left(\frac{T_0 + \beta h}{T_0} \right)^{\frac{-Mg}{R\beta} - 1}$$

Note that $\frac{T_0}{\beta}$ has units of distance and $\beta < 0$, so that if we divide out $-\beta$ from the top and bottom of the enclosed term and simplify

$$N_d(h) = N_{d,0} \left(\frac{h_\beta - h}{h_\beta} \right)^{\frac{-Mg}{R\beta} - 1} \tag{4.1}$$

Hopfield found that replacing $\frac{-Mg}{R\beta} - 1$ with 4, representing $\beta \approx -6.8$ K/km, both approximated the shape of observed refractivity profiles and was computationally simple. This leaves refractivity as a function of altitude and surface refractivity. h_β (eqn. 2.18) was selected to make the integrated zenith profile match observed values via a least squares fit; it is approximately the altitude of the stratopause, but should not be considered an accurate physical altitude.

Note that Hopfield modified the ideal gas law extension for the troposphere to create her equation, and yet applies it to the dry air in both the troposphere and stratosphere despite their different characteristics.

Hopfield uses an equation of the same form as eqn. 4.1 (with h_β now representing the altitude of the tropopause) to approximate the "wet" refractivity profile "for convenience", even though the atmospheric water vapor is not distributed predictably. The wet profile is thus less accurate than the dry profile, but contributes less to the overall signal delay.

It is not always valid to assume that the atmosphere has a constant temperature lapse rate or that lapse rate is ≈ -6.8 K/km, as Hopfield does. Such invalid assumptions have an opposite effect on the wet and dry refractivity profiles (making one have a higher refractivity than is correct, and the other a lower refractivity than is correct), but those effects do not fully cancel out [22].

4.3 Ionosphere Model and Profile Options

As was stated earlier, we are seeking an ionosphere model that provides electron density as a function of altitude, geomagnetic latitude, season, time of day, solar activity level and geomagnetic storm conditions. As with the neutral atmosphere, ionosphere models come in many forms. There are theoretical models, statistical

models and empirical models. Unlike the neutral atmosphere, there are many models that would suffice, although some are overly computationally intensive (requiring a super computer) or not freely available. The AIAA *Guide to Reference Standard Ionosphere Models* [42] provides details on many of these models, including how to obtain them. Rush [43] provides further details on many models. Below I briefly describe the Bent ionospheric model, used by Klobuchar in creating his receiver delay model (see section 3.1.1), and the International Reference Ionosphere, which I chose to use in my analysis.

4.3.1 Bent Ionospheric Model

The Bent Ionospheric Model [44] is an empirical, global model. The developers hoped that an empirical model would improve upon results obtained from the Chapman and other theoretical profiles. The model was created using soundings and satellite measurements taken from 1962 to 1969, a period which covered both a solar minimum and a solar maximum. The fourteen stations that contributed to bottomside soundings covered 12°S–76°N geographic latitude (0°N–85°N geomagnetic latitude) and were only located over land. Topside soundings were made from 1962 to 1966, and while they did not cover as large a range of solar activity levels, they did cover an entire range of geomagnetic latitudes: 75°S–88°N. Probe data filled in the topside solar activity gap for geographic latitudes 70°S–70°N. No data set had complete longitudinal coverage.

The Bent model takes into account geographical and geomagnetic effects, solar activity, local time and seasonal variations. The model requires the user to input a satellite position, receiver position, time, date and solar information. Adding foF2[4] and electron content data can make the model more accurate.

This model is relatively simple, and ignores the D, E and F1 layers of the ionosphere because of their irregularities. The electron content from these regions was added to the general F2 layer, so while the profile is somewhat inaccurate, the vertical total electron content above a point on the Earth's surface remains accurate. (Llewellyn and Bent [44] felt that this simplification would not affect refraction calculations significantly.) The lower ionosphere (the F-layer) is described by a bi-parabola, and the upper ionosphere by an exponential. The topside model has a resolution of 5° latitude and 10° longitude; the bottomside is presumably continuous, but this fact is not quoted in [44].

[4]foF2 is a measurement of the highest frequency ordinary mode, zenith directed signal that will bounce off of the F2 layer of the ionosphere back toward the Earth.

Llewellyn and Bent predict that the model is 75–80% accurate without updates from on-site measurements, which agrees with Klobuchar's assumption (see section 3.1.1) that the high variability in the ionosphere will prevent any model from being more than 75% accurate without knowledge of specific conditions.

The Bent model code can be obtained from the National Space Science Data Center at NASA/Goddard[5].

4.3.2 International Reference Ionosphere (2001)

The International Reference Ionosphere (2001) (IRI) is an empirical model developed by a joint working group of the International Union of Radio Science (URSI) and the Committee on Space Research (COSPAR) to be an international standard for the specification of ionosphere densities and temperatures. This joint working group continues to meet yearly to improve the model as new data and theories become available [29]; the results from these workshops are available in special issues of *Advances in Space Research* [42]. The URSI's prime interest in the model is for radio propagation studies, and so their endorsement of the IRI is significant reason to use it in this research.

The IRI was developed from numerous sources of empirical data and uses theory to bridge data gaps and to verify results. The data is weighted toward North American and European latitudes, and therefore the model is most accurate in Northern-hemisphere mid-latitudes above land [42]. Very little data was available on the D-region of the ionosphere, and therefore this highly variable region is poorly modeled in the IRI [29].

The IRI takes into account solar activity, geomagnetic storm conditions, geographic and geomagnetic latitude, solar zenith angle, date and local time. To gauge the solar influence on the ionosphere, the IRI depends on the twelve month running mean of the number of dark sunspots on the solar disk and the twelve month running mean of the global ionospheric index, which is based upon measured F-peak plasma frequencies and the International Radio Consultative Committee (CCIR) ionosphere model. The sunspot mean correlates well with variations in the ionosphere above the F2-peak, and is used by the IRI to determine the F2-peak altitude and the topside profile. The global ionospheric index is used by the IRI to determine the F2-peak density. The Ap geomagnetic index is used to determine geomagnetic storm conditions.

[5]URL: http://nssdc.gsfc.nasa.gov/space/model/models_home.html

The IRI has several nice features designed to improve its accuracy. Users can enter measured values of the F-peak density, the F-peak critical frequency, the F-peak altitude, the M(3000)F_2 propagation factor, electron densities or exospheric neutral temperatures to improve the accuracy of the IRI profile. The IRI does not, however, accept data from multiple sites or times; the user must weight her own data before entering it for the desired profile.

The IRI gives the user a large amount of flexibility in deciding precisely what sub-models to use and what parameters to calculate, allowing the user to modify both the precision and run speed of the IRI. For example, the user can turn off electron density, ion density, electron temperature, ion temperature and ion drift calculations to shorten the run time of the program. The user can also choose between two options for the magnetic field, foF2, ion density, electron temperature, electron density, F1-region, electron temperature in the topside region, and D-region sub-models.

Chapter 5

Implemented Ray-Tracing Approach

5.1 Numerical Methods Used

There are two main steps to calculating the signal delay between a receiver and a satellite. First, one must calculate the path the signal takes through the atmosphere. Next, one must integrate the group index of refraction along that path.

The path the signal takes is determined by a set of four differential equations derived from Huygen's Principle (equivalently from Snell's Law). These four differential equations are the basis of a two-point boundary value problem: the signal must travel from the satellite to the receiver. There are two classes of numerical methods to solve a two point boundary value problem: shooting and relaxation. I have chosen to use shooting for reasons described below.

Numerically, two-point boundary problems are difficult. Therefore, instead of directly solving the boundary value problem, shooting propagates the four differential equations as if they were part of an initial value problem, and uses a multi-dimensional root-finder to run trial propagations until the signal travels from the satellite to the receiver. You can think of the process as throwing darts at a dart board: on each subsequent throw you change your aim based on where the last dart hit the board until you strike the bull's eye.

Once the multi-dimensional root-finder and integrator have found the path that goes from the satellite to the receiver (expressed simply as the direction the signal leaves the satellite), I integrate the group index of refraction along that path to find the total time it takes the signal to reach the receiver. I then subtract the time it

would take a signal to travel in a vacuum between the satellite and the receiver to find the signal delay. Each of these steps is described further below, with details regarding the specific algorithms used.

Shooting does run into some problems computationally. Sometimes, the root-finder is unable to converge to within the desired level of accuracy. In other words, it cannot determine the direction with which the signal must leave the satellite well enough to strike the receiver to within a chosen distance. In this case, relaxation becomes a better option.

With relaxation, instead of integrating the differential equations along the path between the satellite and the receiver, I solve the equivalent finite difference equations on a mesh of points between the satellite and the receiver. I start by choosing values for the dependent variables at each mesh point, and then successively modify the values at each point until the values all agree with the finite difference equations. This method is best if the boundary conditions are difficult to satisfy and the solution is smooth, but not oscillatory. If the solution has a complex or oscillating structure, the problem will require many grid points and take a long time to solve.

I have chosen to use the shooting method. It is not essential in my analysis to satisfy the boundary conditions with extremely high accuracy since the atmospheric conditions the signal will have passed through are nearly identical (see section 5.1.4 for discussion), and I am only trying to obtain a statistical (and not a specific) solution. Further, I can use a variable stepsize integrator with shooting which will automatically determine which regions along the signal path need higher resolution. With relaxation, this allocation would need to be done manually. Finally, the resolution required in the lower atmosphere would require thousands of mesh points and make an unwieldy relaxation problem.

5.1.1 Problem Set-up

I have chosen to solve for the signal path using a convention where the signal travels backwards, from the receiver to the satellite. This approach is physically equivalent to a signal traveling in the correct direction.

My coordinate system is driven by the straight line between the receiver and the satellite. The receiver is the origin; the l-axis points from the receiver to the satellite; the x-axis points perpendicular to the l-axis in the direction of the center of the Earth (i.e., toward the line parallel to the l-axis which passes through the center of the Earth); and the y-axis completes the right-handed coordinate system. I used the

satellite position at signal transmit time and the receiver position at the time of signal arrival. The l, x and y coordinates of the path are in units of L, the distance between the satellite and the receiver. The receiver is therefore at $(0, 0, 0)$ and the satellite is at $(1, 0, 0)$, and we expect the values of x and y to be quite small.

The integrator stops integrating when $l = 1$, in the plane containing the satellite that is perpendicular to the straight path between the satellite and the receiver.

5.1.2 Derivation of the Ordinary Differential Equations

Equation 2.7 described the change in signal phase at the receiver as a function of the path the signal took from the satellite. We saw in section 2.2.3 that equation 2.7 must be stationary with respect to small deviations from the true signal path — in other words, its functional derivative must be equal to zero for the true signal path. In this manner we can solve for differential equations that describe the path the signal takes through the atmosphere.

Equation 2.7 can be expressed in terms of the convention laid out in the previous subsection (5.1.1), using the shorthand $Dx(l) = \frac{dx(a)}{da}\Big|_{a=l}$ and $Dy(l) = \frac{dy(a)}{da}\Big|_{a=l}$ to represent values of the functions that are the derivatives of x and y, respectively:

$$\int_{l=0}^{l=1} n_p(l, x(l), y(l)) \sqrt{1 + (Dx(l))^2 + (Dy(l))^2} \; dl \tag{5.1}$$

Functionally differentiating this integral with respect to x and setting it equal to zero yields[1]

$$D^2x(l) = \frac{\left(1 + (Dx(l))^2 + (Dy(l))^2\right) \times \left(\partial_1 n_p(l, x(l), y(l)) - Dx(l) \times \partial_0 n_p(l, x(l), y(l))\right)}{n_p(l, x(l), y(l))} \tag{5.2}$$

Doing the same with respect to y yields

$$D^2y(l) = \frac{\left(1 + (Dx(l))^2 + (Dy(l))^2\right) \times \left(\partial_2 n_p(l, x(l), y(l)) - Dy(l) \times \partial_0 n_p(l, x(l), y(l))\right)}{n_p(l, x(l), y(l))} \tag{5.3}$$

leaving us with two second-order ordinary differential equations, which can be rewritten as four first-order ordinary differential equations.

A fifth ordinary differential equation is integrated to find the time (multiplied by

[1] I am indexing arguments from 0, so that $\partial_1 n$ is the partial derivative with respect to the second argument.

c, and so in units of distance) that it takes the signal to traverse the path.

$$Dz(l) = L \times n_g(l, x(l), y(l)) \times \sqrt{1 + (Dx(l))^2 + (Dy(l))^2} \tag{5.4}$$

Note that the index of refraction for the neutral atmosphere is calculated using equation 2.10, which calculates the refractivity of the neutral atmosphere, and then equation 2.5, which converts the refractivity into the index of refraction. Since the index of refraction is very close to 1 (i.e. $N \times 10^{-6}$ is close to 0), numerically this conversion process (which adds 1 to a number very close to 0) causes the index of refraction to have fewer digits of precision than the refractivity. We can minimize the error this introduces into the calculation by substituting $\partial N_p(l, x(l), y(l)) \times 10^{-6}$ for $\partial n_p(l, x(l), y(l))$: precisely the same value, but with more digits of precision because the conversion process was bypassed.

5.1.3 Integrating the Initial Value Problem

Two of the five ordinary differential equations contain partial derivatives of the phase index of refraction. Since there is no explicit expression for the derivative of the phase index of refraction in either the neutral atmosphere or the ionosphere, I have implemented a numerical differentiator for these calculations.

In my ray-tracing code, the five first-order ordinary differential equations are integrated from $l = 0$ to $l = 1$ using the GNU Scientific Library (GSL) version 1.5 ordinary differential equation routines. These routines allow a number of Runge-Kutta, Burlish-Stoer and Gear integrators to be applied using the same underlying framework, allowing easy comparison of integrator performance. These routines further apply an adaptive integration method for all integrators. This method adjusts the step size dl at each integration step so that the observed error on that step does not exceed a desired error by more than 10%. Further, if the observed error is less than half the desired error, the method will increase the step size to speed up the integration. More details on these routines can be found in the *GNU Scientific Library Reference Manual*, edition 1.5[2].

Gear integrators are extremely stable, but are implicit integrators and therefore are typically less accurate and slower than an explicit integrator such as Runge-Kutta. Since the ray-tracing equations do not have problems with integration stability, I did not examine this integrator. The GSL Burlish-Stoer integrators are also implicit

[2] M. Galassi, J. Davies, J. Theiler, B. Gough, G. Jungman, M. Booth and F. Roosi. Published electronically, July 2004. URL: http://www.gnu.org/software/gsl/

(and therefore typically slower and less accurate than Runge-Kutta), and perform best when the refractivities input into equations 5.2–5.4 maintain the analyticity of these equations. Since GRAM-99 and IRI-2001 do not guarantee analytic outputs (due to their use of interpolation), the ray-tracing equations are not analytic. Of the integrators available, I therefore examined the embedded Runge-Kutta (2,3) method, the embedded Runge-Kutta-Fehlberg (4,5) method, the embedded Runge-Kutta Cash-Karp (4,5) method, and the embedded Runge-Kutta Prince-Dormand (8,9) method. Of these four, the Prince-Dormand (8,9) method was an order of magnitude slower than the rest at integrating a sample signal path. Of the remaining three, the Runge-Kutta-Fehlberg (4,5) was a factor of two faster than the other two.

I chose a desired integration precision of $5 \times 10^{-9} \times L$. Since the satellite is approximately 20,000 km from a ground user (i.e. $L \approx 20,000$), this precision works out to ≈ 0.1 m. In the worst case scenario when the receiver is at 1,700 km altitude and a satellite is at the horizon, the distance between the satellite and the receiver is approximately 30,000 km, making the integration precision ≈ 0.15 m. This value will directly affect the error on the delay calculation. More conservative precisions were examined, but they made the integrator prohibitively slower.

5.1.4 Multidimensional Root Finder

The multidimensional root-finder drives the integration of the four ordinary differential equations that determine the signal path, modifying the direction at which the signal leaves the receiver (v_x, v_y) and integrating the differential equations until the signal strikes to within a given distance of the satellite. In other words, the root-finder is trying to solve

$$\begin{aligned} x(v_x, v_y)_{l=1} &= 0 \\ y(v_x, v_y)_{l=1} &= 0 \end{aligned}$$

I examined the multidimensional root-finders provided by the GSL, version 1.5. These root-finders guess the appropriate value of $\boldsymbol{v} = (v_x, v_y)$ using a variant of Newton iteration:

$$\boldsymbol{v} \rightarrow \boldsymbol{v}' = \boldsymbol{v} - J^{-1}X$$

where X is the array $[x(\boldsymbol{v}), y(\boldsymbol{v})]$ and J is the Jacobian matrix expressed by $J_{1j} = \partial x/\partial v_j$, $J_{2j} = \partial y/\partial v_j$. Since there is no explicit expression for the derivative of the index of refraction in the neutral atmosphere or the ionosphere, I was limited to

examining root-finding algorithms that would accept derivatives approximated from finite differences. Of the four algorithms which met this criteria, I considered the Hybrid algorithm with internal scaling and the discrete Newton algorithm. I have found that the discrete Newton algorithm will often converge quicker toward the root than the Hybrid algorithm, but will not reliably reach the root to within the desired precision, instead "bouncing" around the root. Unfortunately, the Hybrid algorithm, while approaching the root more consistently than Newton and coming much closer to finding the root, will also not converge reliably.

I have therefore chosen to use a recursive grid search to solve for $\boldsymbol{v}$. This search method is initialized with a range of (v_x, v_y) values over which the satellite is likely to be located. If a previous delay calculation was made in nearly an identical geometry (for example, if we are calculating the delay to a satellite over the course of a receiver trajectory, and sample delays at a rate of 1 Hz), I search the region that encompasses 1% to either side of the previous root, broken into a 4 x 4 grid. Despite different absolute receiver and satellite positions, the magnitude and geometry of the atmosphere that the signal must travel through is roughly the same. Further, since we are only considering the *angle* from the straight line between the satellite and the receiver, this method is quite reliable. If there is no previous information about the location of the root, I pick a range that encompasses (v_x, v_y) typical for a satellite at $0° - 90°$ elevation angle for a ground-based receiver.

If a residual ($x(l = 1)$ or $y(l = 1)$, in the satellite plane) crosses zero in the course of searching for the root, and yet the root was not found to within tolerance limits, I reduce the search region to be between the v_x's or v_y's on either side of the zero crossing. For a 4 x 4 search grid, this decreases the search region by a factor of 3. If the root is not found within the chosen range, I increase the range by a factor of 9 since the root-finder will be able to recover its resolution with two subsequent reductions in the search area. These values were found to be a good balance between increasing the search region sufficiently in one step to likely include the root, and allowing the root-finder to converge quickly to a high resolution.

Since the angle at which the signal leaves the receiver is simply a function of how much refraction the signal experienced in the atmosphere, we expect most of the deviation from the straight line path to be in the direction away from the center of the Earth — in my convention, $-v_x$. For this reason, I preferentially search the grid in the v_x direction first, and shrink or expand the search range in that dimension multiple times to search for the root before modifying the value of v_y. In practice, v_y is rarely modified, and remains close to 0.

precision	$\|\boldsymbol{x}_{ghost} - \boldsymbol{x}_{true}\|$	Δangle (radians)	Δangle (deg)	max ΔL
5×10^{-7}	10 m	5×10^{-7}	3×10^{-5}	5μm
5×10^{-6}	100 m	5×10^{-6}	3×10^{-4}	500μm
5×10^{-5}	1 km	5×10^{-5}	3×10^{-3}	5 cm
5×10^{-4}	10 km	5×10^{-4}	3×10^{-2}	5 m
5×10^{-3}	100 km	5×10^{-3}	3×10^{-1}	500 m
5×10^{-2}	$1,000$ km	5×10^{-2}	3×10^{0}	50 km
5×10^{-1}	$10,000$ km	5×10^{-1}	3×10^{1}	$5,000$ km

Table 5.1: The effects of varying the precision of the root-finding algorithm for $L \approx 20,000$ km.

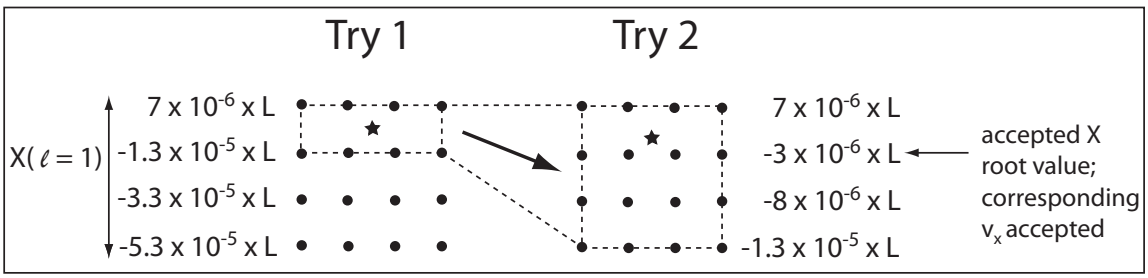

Figure 5-1: **Root-finder grid search.** Each dot represents one shot (i.e. one $\boldsymbol{v}$), and the star represents the root (the satellite). If the resulting $x(l = 1)$ from two successive shots are not within the tolerance ($5 \times 10^{-6} \times L$) but cross zero, the root-finder reduces the search region in the x-dimension and searches again.

This method, while brute force, finds the roots much faster and with fewer shots than the two GSL routines, and is 100% reliable.

I have chosen to find the root to within $5 \times 10^{-6} \times L$, or 100 m for $L \approx 20,000$ km. Since the satellite is so far away from the receiver, a signal traveling to a "ghost" satellite 100 m from the true satellite passes through essentially the same atmosphere as a signal to the true satellite, and so the delays are equal to well beyond the level of precision of the integration itself. Geometrically, the signal path has only moved 5×10^{-6} radians, or 3×10^{-4} degrees by moving the satellite 100 m. Theoretically, some care must be taken in the delay calculation itself; the line-of-sight distance subtracted in equation 2.4 must be modified according to the residuals $x(v_x, v_y)_{l=1}$ and $y(v_x, v_y)_{l=1}$. In practice, the difference between the distance to the true satellite and the ghost satellite is at most 0.05 cm for the chosen level of precision, which is well below the precision of the integration. Table 5.1 shows the results for choosing different levels of precision for the root-finder.

Trials with the manual root-finder have shown that there are actually multiple

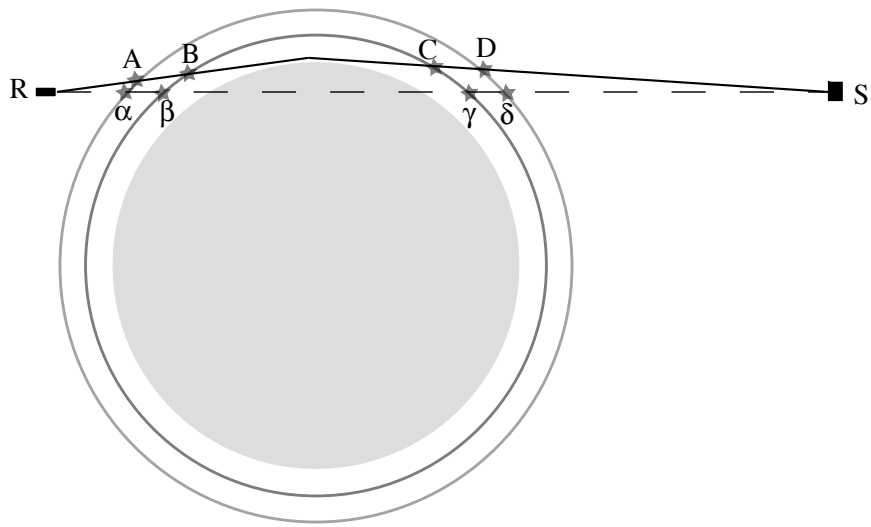

Figure 5-2: An illustration of the problems in breaking the signal delay into components. The outer ring indicates the outer edge of the ionosphere and the inner ring indicates the outer edge of the neutral atmosphere. The receiver is at R and the GPS satellite at S. Figure is not to scale.

roots to the signal delay problem. In other words, a signal can take more than one path from the receiver to the satellite. The multiplicity of these roots becomes evident with precision levels of $5 \times 10^{-7} \times L$ and smaller. These multiple paths will be essentially the same; the path length will change on the micron level. I have therefore chosen to avoid this multiplicity by using a precision level of $5 \times 10^{-6} \times L$.

5.1.5 Component Delay Calculations

We now face a seemingly easy question: what is meant by "neutral atmosphere" delay and "ionosphere" delay? The ray-tracing incorporates only the two atmosphere models (GRAM-99 and IRI-2001) — no other processes that might delay or bend the signal — and so the total delay must be the sum of the two component delays. However, distinguishing these two components is quite difficult.

Consider figure 5-2, which presents a signal from a satellite below the horizon, but that bends sufficiently in the neutral atmosphere to still be seen by the receiver. The total signal delay is defined as the time it takes to transit the path R-A-B-C-D-S (t_R^S) minus the time it would take to transit the path R-α-β-γ-δ-S in a vacuum (v_R^S). But how is this delay broken down into neutral atmosphere delay and ionosphere delay

components? One convention could be to define the ionosphere delay as $(t_A^B + t_C^D - v_\alpha^\beta - v_\gamma^\delta)$ and the neutral atmosphere delay as $(t_B^C - v_\beta^\gamma)$, but this has two problems: the delay $(t_R^A + t_D^S - v_R^\alpha - v_\delta^S)$ is unaccounted for and the neutral atmosphere delay has the distinct possibility of being negative. A second convention would be to define the ionosphere delay as $(t_R^B + t_C^S - v_R^B - v_C^S)$ and the neutral atmosphere delay as $(t_B^C - v_B^C)$. This solves the problems the first convention has with "extra delay" and a negative neutral atmosphere delay, but is attributing the extra geometrical pathlength acquired between points R and A (and points D and S) to the ionosphere, even though the bending which ultimately caused this extra pathlength occurred in the neutral atmosphere. Further, the straight line segments subtracted (v_R^B, v_B^C and v_C^S) sum to more than v_R^S, and therefore delay is being undercounted. Finally, one must know the locations of points B and C in order to calculate the values of v_R^B and v_C^S. These locations are difficult to calculate. The first problem, of the extra geometrical delay being attributed to the ionosphere instead of the neutral atmosphere, can be fixed, but the second problem cannot within this approach.

I have concluded that there are only arbitrary definitions of what constitutes "ionosphere" delay and what constitutes "neutral atmosphere" delay; we can only estimate their values using a variant of the second convention mentioned above, or by calculating each component separately assuming that the "other" part of the atmosphere is a vacuum with $n_g = n_p = 1$. With atmosphere models that do not contain specific features (such as thunderstorms), I predict that these values will almost always be lower bounds to the correct value.

Another convention presents itself if we consider the application of these ray-tracing results. Ideally, if the atmosphere matched the GRAM-99 and IRI-2001 models, a GPS receiver should be able to use the ray-traced delays to compensate for the signal delay it experiences. Therefore we might think to define the ionosphere delay as the delay predicted by the dual-frequency correction method, based upon the ray-traced transit times of an L1 signal and an L2 signal, and define the neutral atmosphere delay as all of the remaining delay (i.e. total delay minus ionosphere delay). However, this too runs into a problem. As refraction increases, the paths that the L1 and L2 signals take begin to deviate, and the assumptions that go into the dual-frequency correction algorithm break down. This convention is therefore inadequate.

One should note that choosing the appropriate convention is only relevant when the receiver is observing a satellite near (or below) the horizon, since at all other times the signal is sufficiently straight that any error due to either using the second con-

vention mentioned above or due to assuming that the "other" part of the atmosphere was a vacuum is on par with the errors from the integration routines.

This issue will be discussed further in the error analysis section of chapter 6.

5.1.6 Integration Implementation

My ray-tracing routines use neutral atmosphere indices of refraction from 0 km to 65 km altitude and ionosphere indices of refraction from 65 km to 1050 km altitude. Above 1050 km, both indices of refraction are set to 1 (i.e. the index of refraction of a vacuum).

The previous section discussed many conventions of how to define the component (neutral atmosphere and ionosphere) signal delays. I have chosen to consider the neutral atmosphere to be a vacuum when calculating the ionosphere delay, and to consider the ionosphere to be a vacuum when calculating the neutral atmosphere delay.

This approach does run into some notable problems when the signal starts to bend significantly. The most extreme example is to consider a receiver above the ionosphere, observing a satellite technically just below the horizon, such as in figure 5-2. In this geometry there is significant refraction in the neutral atmosphere, but not in the ionosphere. So, while the neutral atmosphere delay calculation will bend the signal around the horizon to connect the satellite and the receiver, the ionosphere calculation will find that the signal strikes the Earth. In real life, these situations cannot both happen.

I have therefore adopted a slight modification. I first calculate the direction $\boldsymbol{v}$ that the signal must leave the receiver in order to strike the satellite, assuming refraction in both the ionosphere and the neutral atmosphere. I then calculate the total delay using both models, the neutral atmosphere delay assuming the ionosphere is a vacuum but using the $\boldsymbol{v}$ calculated using both models, and the ionosphere delay assuming the neutral atmosphere is a vacuum but again using $\boldsymbol{v}$ calculated using both models. Since the signal mostly bends away from the Earth, this convention means that there will always be an ionosphere signal delay when there is a neutral atmosphere signal delay. This modified approach is still somewhat inaccurate when there is significant bending. Again consider a receiver above the ionosphere, observing a satellite just below the horizon. For the ionosphere calculation, the signal will not experience the significant bending experienced in the neutral atmosphere, and therefore it will be in an incorrect position (at most 6.5 km away in the F2-layer) and at an incorrect angle (at most

0.2° off from the correct angle) on its second pass through the ionosphere. With a relatively general model such as the IRI that does not have small-scale features, these differences are small enough that the resulting error in the ionosphere delay calculation should be small. Since there is very little bending in the ionosphere and the neutral atmosphere delay is not affected significantly by geographical position, this approach should provide neutral atmosphere delays with essentially no additional errors.

As was mentioned in the previous section, this method will typically provide lower bounds on both the neutral atmosphere and the ionosphere signal delay.

5.2 Implementation of the Global Reference Atmosphere Model 1999

I used GRAM-99 in its "trajectory" mode, and modified it (via file `gramtraj.f`) so that it returned temperature, pressure, and water vapor pressure values directly to the ray-tracing code. Otherwise, I made no modifications to the GRAM-99 code except to fix some file read-errors.

As mentioned in section 4.1.1, GRAM-99 will accept Range Reference Atmosphere data in order to make its atmosphere more accurate for a given location. I did not use this feature. I also did not use the feature where the user can specify initial perturbations.

The GRAM-99 program is initialized with a set of parameters from a `namelist` file. (A sample of this file can be found in the GRAM-99 documentation [40].) This file requires the user to input the Ap geomagnetic index, daily and average 10.7 cm solar flux[3], date, time and initial receiver position. (Depending on the mode the program is run in and the features selected, additional inputs may be necessary; these are all that were input in my analysis.) I obtained the Ap geomagnetic index values from the International Reference Ionosphere's code package; this data file contains values at 3-hour intervals for all days from 1960 through July 2003. Values of the 10.7 cm (2800 MHz) flux were obtained from the NOAA National Geophysical Data Center (NGDC)[4]. The daily values from 1947 through October 2004 were taken by radio telescopes near Ottowa and Penticton, B.C., Canada. Observed flux values were

[3]The 10.7 cm solar flux is a measurement of the solar radiation density that reaches the radius of the Earth's orbit at a wavelength of 10.7 cm (2800 MHz). Solar radiation near this wavelength is responsible for much of the ionization in the ionosphere.

[4]FTP: ftp://ftp.ngdc.noaa.gov/STP/SOLAR_DATA/SOLAR_RADIO/FLUX/

used, since adjusted values reflect conditions on the sun and not the flux actually reaching the Earth[5].

GRAM-99 was responsible for all refractivities between 0 km and 65 km altitude.

5.3 Implementation of the International Reference Ionosphere 2001

I used a version of the IRI-2001 developed specifically to calculate electron densities. In this version, features of the full version that do not affect electron density have been eliminated. This version is denoted "ne" in the IRI code.

As was mentioned in section 4.3.2, the IRI has many sub-model options from which to choose. I used the defaults in most cases, but selected the options to:

1. not calculate the electron or ion temperatures
2. not calculate the ion densities
3. not calculate the vertical ion drift
4. not use the foF2 storm model
5. use the URSI foF2 model
6. use the Danilov (1995) ion density model

The first four options were used to increase the run speed of the program. Of the four, testing showed that only the foF2 storm model had the potential to affect the electron density results, and then only when the Ap geomagnetic index was high. The last two options are the default, but do not follow the convention of the rest of the options in the code, where the default options are specified as "true".

I extensively modified the included file `iritestne.f` so that the IRI interfaced directly with the ray-tracing code instead of printing results to file. Otherwise, I only modified the IRI code by removing statements that printed data to an output file.

The IRI-2001 was responsible for all refractivities between 65 km and 1050 km altitude.

[5]NOAA, Nation Geophysical Data Center, Space Physics Interactive Data Resource. URL: http://spidr.ngdc.noaa.gov/spidr/

Chapter 6

Ray-Traced GPS Signal Delays

This chapter presents L1 signal delays calculated by ray-tracing through the neutral atmosphere and the ionosphere via the GRAM-99 and the IRI-2001 models as described in chapter 5. Please take careful note of sections 5.1.5 and 5.1.6 to fully understand how the neutral atmosphere and ionosphere delays were calculated and the geometries these delays represent.

I present ray-tracing results two different ways. The first method displays profiles of the delay (neutral atmosphere, ionosphere and total) experienced by signals traveling to a receiver at a given altitude for a range of satellite elevation angles. The second method displays the delay experienced by signals to a receiver aboard a space vehicle over the course of a sample space vehicle trajectory.

In the first method, I have created a control set of profiles which describe the delay experienced by signals from a satellite due east of a receiver at 28°N 80°W (approximately Kennedy Space Center) on March 21, 1980 (vernal equinox and a period of high solar activity in the Sun's 11 year cycle) at noon local time. A series of comparison profiles then follow to show the effect of latitude, date, time of day and satellite azimuth on the signal delay. The reported elevation angle is the geometric elevation angle to the satellite — not the apparent elevation angle (i.e. the elevation angle at which the signal arrived at the receiver).

In the second method, I have calculated the delay experienced by signals from all visible satellites over the course of a sample space vehicle trajectory, allowing us to see what typical signal delays and delay rates might be experienced during launch, orbit, and reentry. The space vehicle launches from 28°N 80°W (approximately Kennedy Space Center) at noon on March 21, 1980 and travels east along a constant line of latitude for approximately 36 minutes. (See figures 6-25 and 6-26.)

These ray-tracing results are intended to be average delays based on an average

atmosphere, and are not intended to predict the signal delay on a specific day. These results do not show the effects of unpredictable phenomena such as variations in water vapor concentration, ionospheric scintillation or traveling ionospheric disturbances, which could introduce bumps or noise into the observed delay over the course of a space vehicle trajectory. (See section 4.1.1 for details on the GRAM-99 water vapor distribution.)

6.1 Delay as a Function of Elevation Angle and Receiver Altitude

6.1.1 Validation of Ray-Tracing Results

In order to validate the ray-tracing routines, I compared the results from ray-tracing through the GRAM-99 and IRI-2001 models for a receiver on the surface of the Earth with results from the Altshuler, NATO and Klobuchar delay models.

To validate the neutral atmosphere ray-tracing results, I compared the Altshuler and NATO models (section 3.1.2) with the control profile for 0 km receiver altitude. The Altshuler model is a function of day of year, receiver latitude and receiver altitude; the NATO model is a function of day of year and latitude (both via the Altshuler surface refractivity equation, which was used to calculate the NATO surface refractivity input), and altitude of receiver. We can see in figure 6-1 that the delays using each of these three methods are nearly identical except at low elevation angles. The Altshuler model was only designed for angles down to 5°, explaining the discrepancy. The NATO model likewise becomes less accurate near the horizon. While conceptually the delay should increase monotonically as the signal approaches the horizon, with the signal refracting more and more in the lower atmosphere until it is abruptly blocked by the horizon, the NATO model predicts a smooth decrease in the neutral atmosphere delay until it reaches zero. The agreement of these models above the lowest elevation angles validates the neutral atmosphere ray-tracing results.

To validate the ionosphere ray-tracing results, I compared the control profile for 0 km altitude with the Klobuchar model (section 3.1.1). I used the broadcast coefficients from [1], which are reported to be actual coefficients broadcast during high solar activity, spring equinox conditions. In addition to these coefficients, the Klobuchar model is a function of local time, receiver altitude and latitude. In figure 6-2 (top), note that while ray-tracing through the IRI-2001 produced similar results to the Klobuchar model, the results are not identical. This is to be expected. First, the

Klobuchar coefficients were for specific conditions, whereas the ray-tracing ionosphere represented average conditions. Second, the Klobuchar mapping function (eqn. 3.3) is computationally simple; in figure 6-2 (bottom), note that a slightly more complex geometrical calculation of the mapping function[1] provides a curve that is closer to the shape of the ray-traced ionosphere delay. I have used this geometrical mapping function to map the ray-traced zenith delay down to the horizon in the top plot; it is much closer to the ray-traced delay. The remaining difference is likely due to the uniform scaling of all layers of the ionosphere by the geometrical mapping function. This scaling should actually vary with the altitude of the layer. Therefore, the results seen in figure 6-2 validate the ionosphere ray-tracing results.

[1]The geometrical mapping function calculates how much distance a (straight) signal must travel to go between 50 km altitude and 1050 km altitude at a given elevation angle. At the zenith, this is simply 1000 km. At the horizon, however, the signal travels a slanted path through the ionosphere, and therefore travels roughly 3,250 km to travel from 50 km to 1050 km altitude.

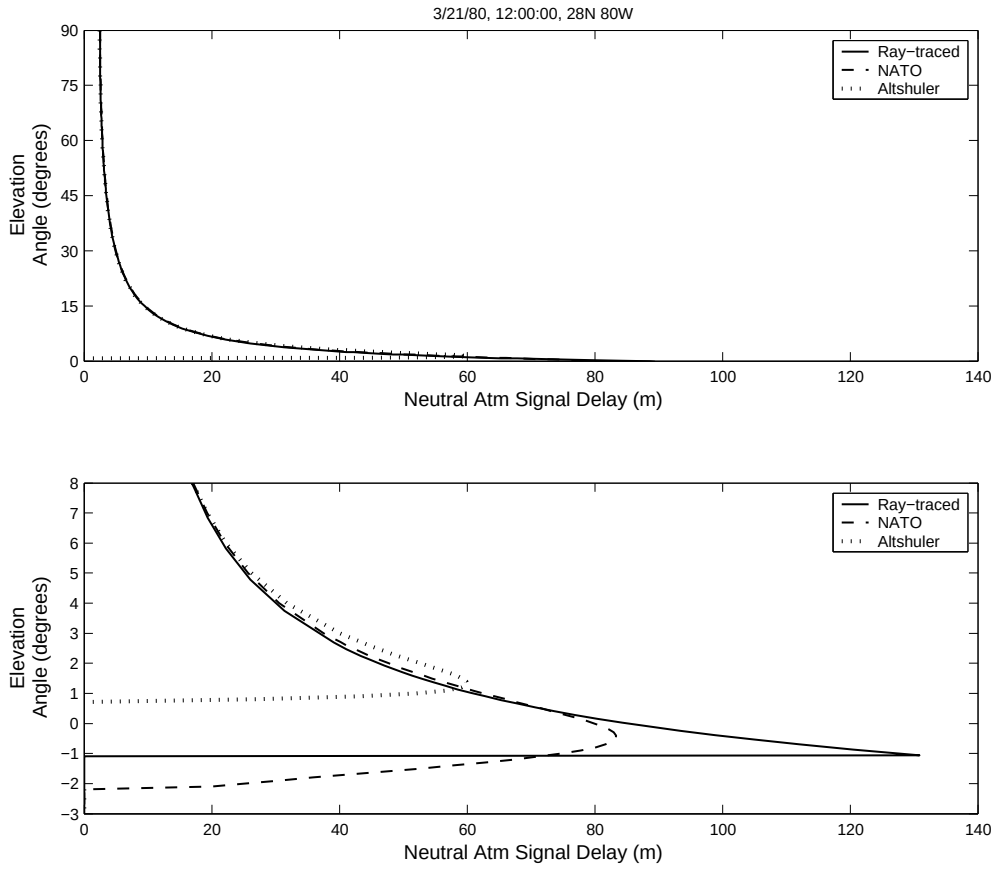

Figure 6-1: **Neutral Atmosphere Ray-Tracing Validation.** A comparison of neutral atmosphere signal delay ray-tracing results with the NATO and Altshuler delay models. Ray-tracing was done through the GRAM-99 atmosphere for a receiver on the ground at 28°N 80°W at noon on 3/21/80, and a satellite due east of the receiver. TOP: All elevation angles. BOTTOM: Low elevation angles.

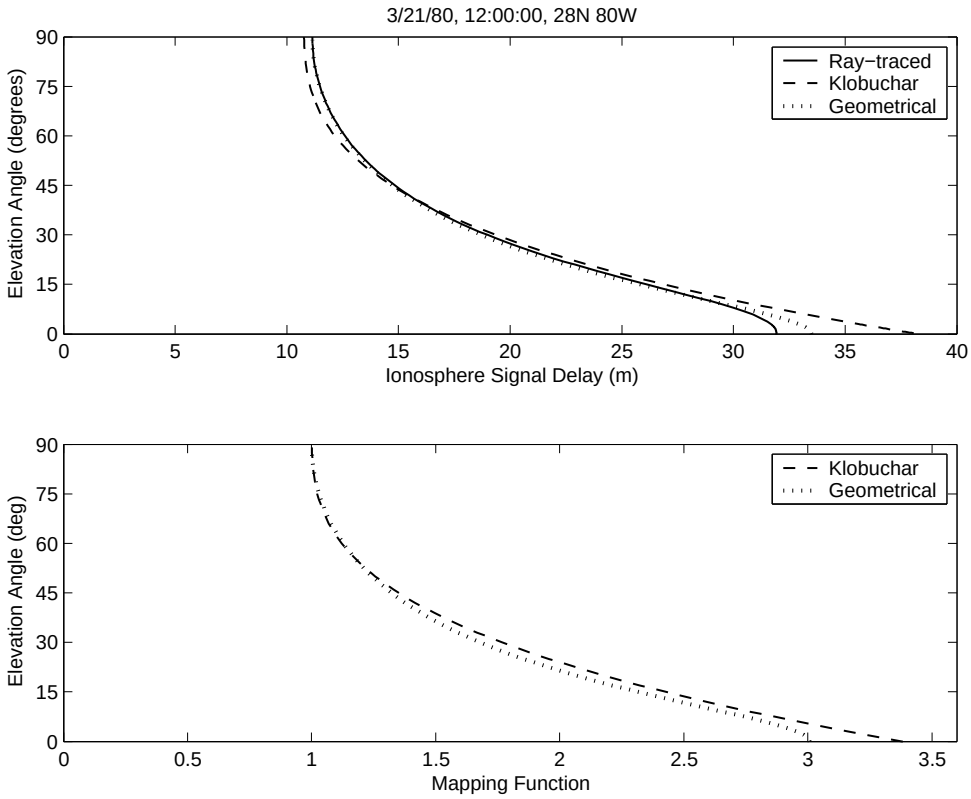

Figure 6-2: **Ionosphere Ray-Tracing Validation.** TOP: A comparison of ionosphere signal delay ray-tracing results with both the Klobuchar delay model and the ray-traced zenith delay mapped by the geometrical mapping function presented in the lower plot. Ray-tracing was done through the IRI-2001 ionosphere for a receiver on the ground at 28N 80W at noon on 3/21/80, and a satellite due east of the receiver. The Klobuchar broadcast coefficients were for a high solar activity, spring equinox day, as presented in [1]. BOTTOM: A comparison of the Klobuchar mapping function to a geometrical mapping function that calculates the linear distance a straight line signal originating from the Earth's surface spends between 50 and 1050 km altitude for a given elevation angle, and divides this distance by 1000 km. (This normalizes the zenith mapping value to 1.) The Earth is assumed to be spherical, with a radius of 6,378 km.

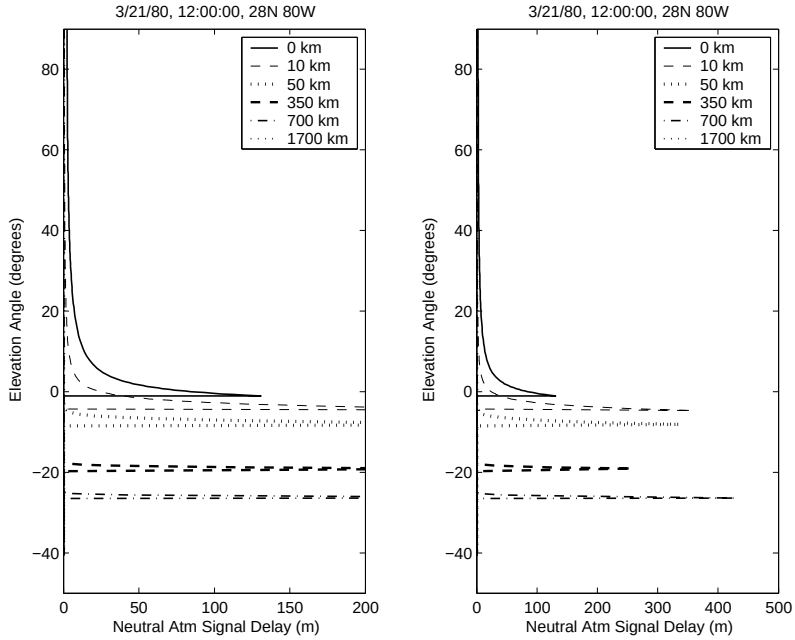

Figure 6-3: Neutral atmosphere signal delay as a function of elevation angle for control conditions at six different receiver altitudes. LEFT: Close-in view. RIGHT: Full view. The delay curve for 1700 km is missing because the range of elevation angles at which there is any delay for a receiver at this altitude is less than the resolution of the ray-tracing (0.1° apparent elevation angle).

6.1.2 Typical Features

Typical Features in Neutral Atmosphere Delay

Figure 6-3 displays the control neutral atmosphere signal delay profiles for six different receiver altitudes. These profiles agree with what we would expect geometrically (see figure 6-4), with one notable addition.

As we expect geometrically, the signal delay monotonically increases as the signal approaches the horizon since it is passing through the neutral atmosphere for a longer distance. Further, as the signal approaches the horizon it is traveling an increasing percentage of its distance in the densest regions of the atmosphere, increasing the delay. This holds true at all receiver altitudes. As we increase in altitude, the range of elevation angles that the neutral atmosphere comprises shrinks (in figure 6-4, $\gamma < \delta$) and the horizon is located at lower and lower elevation angles. (In figure 6-4, $\beta < \alpha$. β and α are elevation angles and are therefore negative.)

Note, however, that the signal delay increases drastically right near the horizon

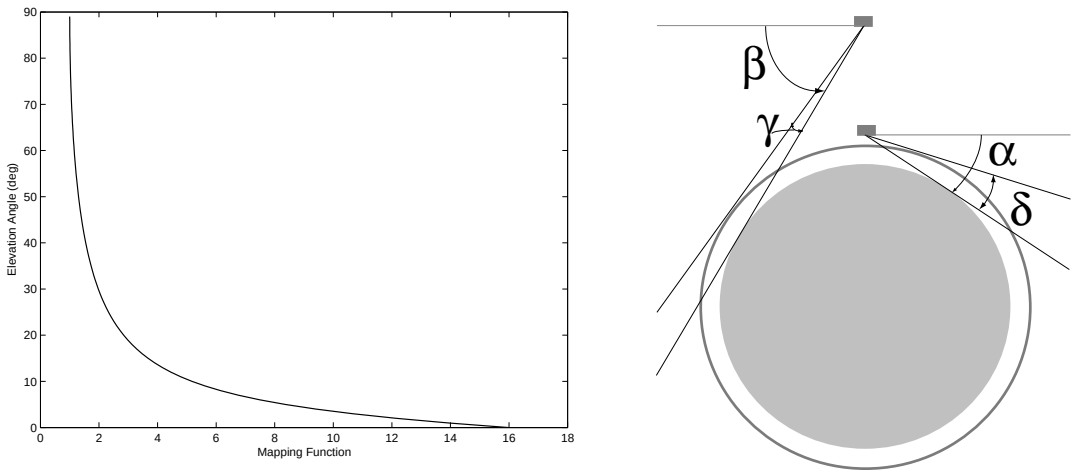

Figure 6-4: Geometrical effects on signal delay profile. LEFT: The geometrical mapping function that calculates the linear distance a straight line signal originating from the Earth's surface spends between 0 and 50 km altitude for a given elevation angle, and divides this distance by 50 km. (This normalizes the zenith mapping value to 1.) The Earth is assumed to be spherical, with a radius of 6,378 km. RIGHT: As a receiver gains in altitude, the range of elevation angles that the neutral atmosphere comprises shrinks and the horizon is located at lower and lower elevation angles. Figure is not to scale.

at all receiver altitudes. This is due to atmospheric refraction. As a signal travels through this region, which has a large refractivity gradient (see figure 2-5), it is bent around the Earth. At the lowest elevation angles the receiver is actually seeing a satellite that is below the horizon (see figure 6-5). The additional path length that the signal travels due to this bending plus the additional time spent in the neutral atmosphere causes the large delay.

Figure 6-3 implies that the maximum delay reached at each receiver altitude is different; however, I believe that the maximum delays are different due to the resolution of the ray-tracing (0.1° apparent elevation angle). Once the receiver is outside of the neutral atmosphere, all signals that pass closest to the Earth at the same altitude should have the same delay. For example, in figure 6-6 receivers A, B and C should all have the same neutral atmosphere signal delay. Changes due to temporal or latitude variations from the receiver seeing "further around the Earth" will be discussed in the following sections; they are negligible.

Also note that the delay curve for 1700 km is missing in figure 6-3. This is because the range of elevation angles at which there is any delay for a receiver at the upper altitudes is less than the resolution of the ray-tracing (0.1° apparent elevation angle). The neutral atmospheric component is evident in the total signal delay profile (figure 6-12), however.

Figure 6-7 displays the neutral atmosphere signal delay in a contour plot that encompasses all receiver altitudes and elevation angles examined.

Typical Features in Ionosphere Delay

Figure 6-8 displays the control ionosphere signal delay profiles for five different receiver altitudes. The elevation angle of the horizon is the same for both delays, but there are a few notable differences between these profiles and the neutral atmosphere signal delay profiles.

First, while the delay at first increases as elevation angle decreases, it eventually reaches a peak and then decreases until the satellite goes below the horizon. This is because the delay peaks when the signal spends the most time possible in the densest region of the ionosphere. For receivers at or above the F-layer peak of the ionosphere, this occurs when the signal travels tangent to the altitude of the F-layer peak. When the signal is tangent to a lower layer it still passes through the densest region of the ionosphere, but spends less time there and therefore accumulates less delay. (See figure 6-9.)

Second, there is no spike in the signal delay right before the signal is cut off by

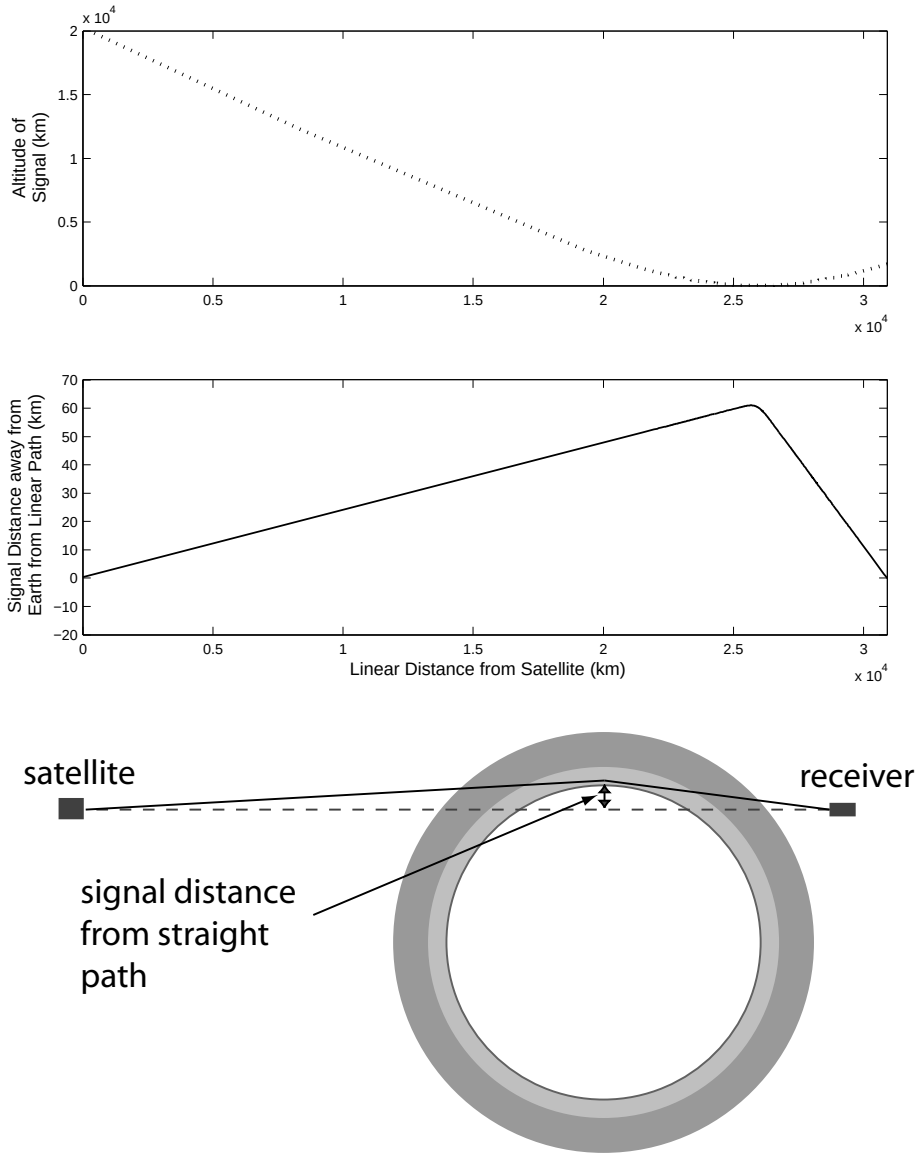

Figure 6-5: A sample horizon-grazing signal path. TOP: The altitude of the signal as it travels from the satellite to the receiver. Closest approach is ≈ 4 km. MIDDLE: The distance between the signal and the straight line connecting the satellite and the receiver in the direction toward the Earth. In order to reach the receiver, a signal must angle away from the straight line connecting the receiver and the satellite in the direction away from the center of the Earth. The bending that occurs when the signal is in the lowest layers of the atmosphere will then point the signal toward the receiver. Geometrically, this satellite is below the receiver's horizon since the signal deviates 60 km from the straight line and yet the altitude of closest approach is only 4 km. BOTTOM: Diagram of signal path. Figure is not to scale.

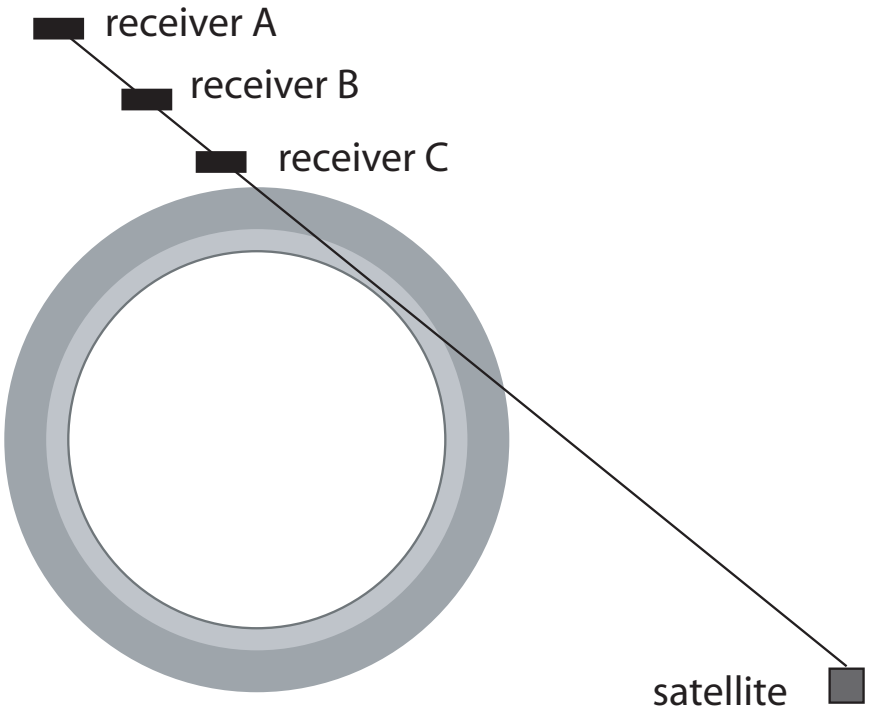

Figure 6-6: Receivers A, B and C should experience the same signal delay. Figure is not to scale.

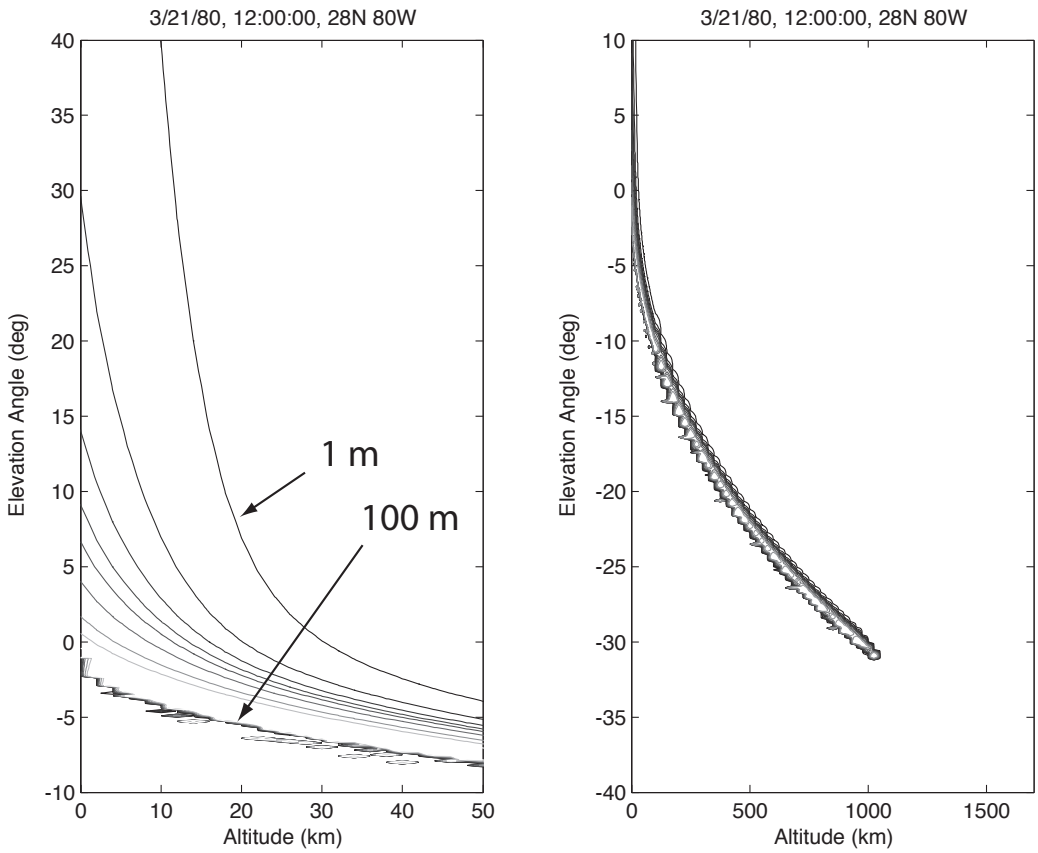

Figure 6-7: A contour plot of neutral atmosphere signal delay as a function of receiver altitude and geometric elevation angle. The plot on the left examines the contour lines at the lower altitudes. Contour lines are at 1, 5, 10, 15, 20, 30, 50, 70, 100 m delay. Note how as altitude increases, the delay is reduced to a sliver of elevation angles. Also note that the delay at high receiver altitudes is missing because the range of elevation angles for which there is delay is less than the resolution of the ray-tracing (0.1° apparent elevation angle). The curves have some distortion at the low elevation angles due to the MATLAB interpolation routines used in the contour plotting and due to the resolution of the data input (every 25 km altitude for the higher altitudes).

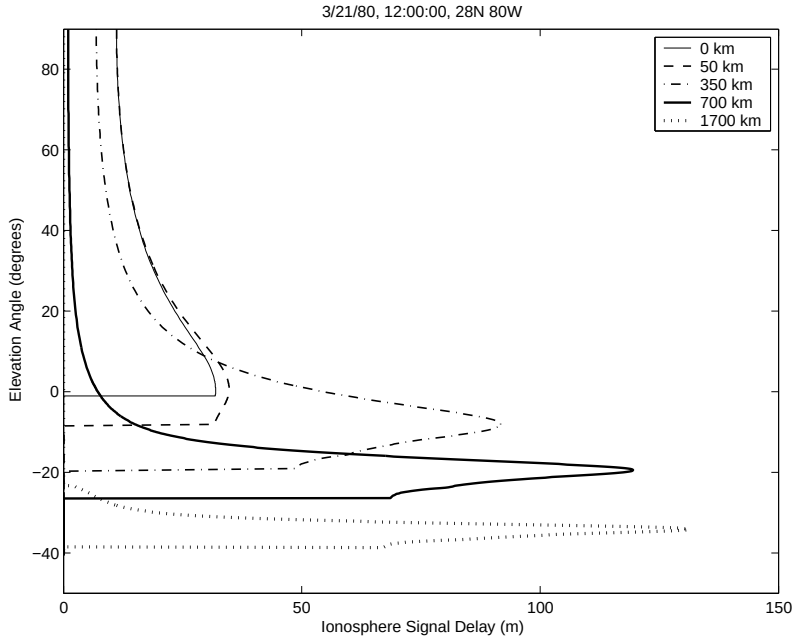

Figure 6-8: Ionosphere signal delay as a function of elevation angle for control conditions at five different receiver altitudes.

the Earth limb, as with the neutral atmosphere delay. I should first note that my ray-tracing implementation could affect behavior near the horizon. Since the neutral atmosphere is treated as a vacuum, the signal used to calculate the ionosphere delay is not bent near the horizon by the neutral atmosphere as it realistically would be. The neutral atmosphere will change the angle and point of entry of the signal into the ionosphere on the signal's second pass through the ionosphere; however, this change ($< 0.2°$ and < 60 km at 350 km altitude) should not be sufficient to significantly affect the delay accumulated in the ionosphere. Conversely, if the signal were to experience significant bending in the ionosphere, it would occur when the signal travelled tangent to $\approx$275 km altitude, where the refractivity gradient is largest. (See figure 2-5.) If we were to see a spike in the ionosphere delay, it should occur in this region and not at the horizon.

Third, for a given receiver altitude, the range of elevation angles for which there is ionosphere delay is greater than the range for which there is neutral atmosphere delay. This is due purely to geometry: the ionosphere is both higher and contains a larger range of altitudes than the neutral atmosphere. (See figure 6-10.)

Figure 6-11 displays the ionosphere signal delay in a contour plot that encompasses

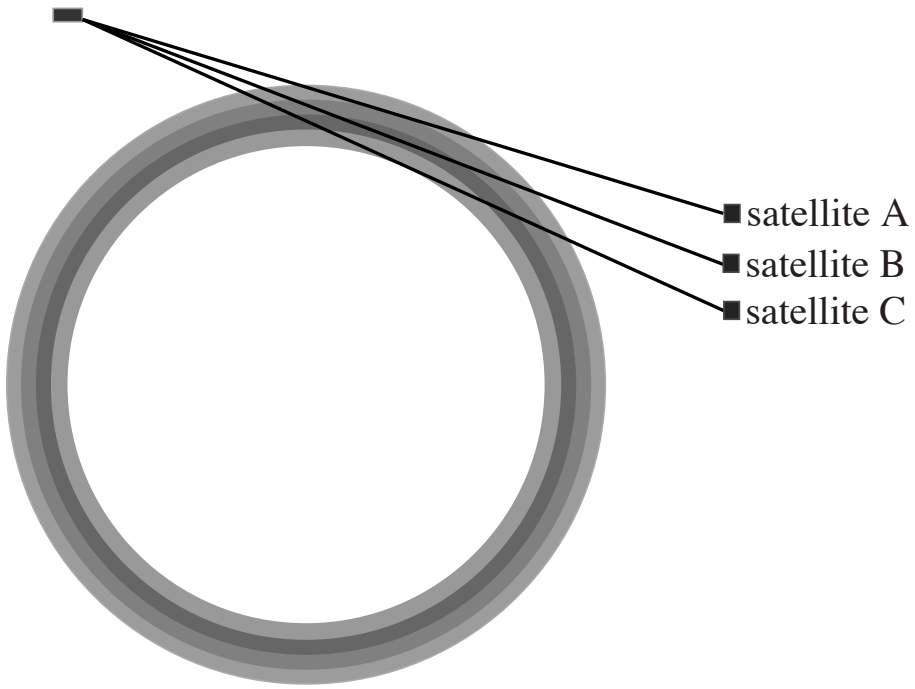

Figure 6-9: Ionosphere signal delay increases as satellite elevation angle decreases, peaks when the signal is tangent to the densest region of the ionosphere, and then decreases until the satellite goes below the horizon. In this figure, darkness indicates a higher electron density. Note that while the signal from satellite C travels the longest distance in the ionosphere, it travels much of that distance in a region of low electron density. Of the three, the signal from satellite B travels the longest distance in the densest region of the ionosphere, and therefore experiences the largest ionosphere delay. Figure is not to scale.

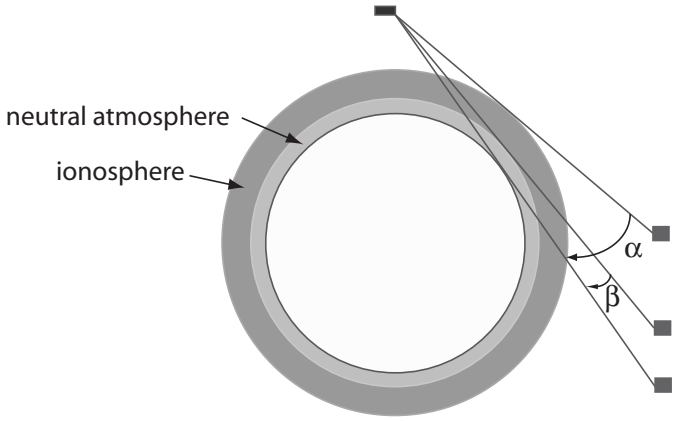

Figure 6-10: For a receiver above the ionosphere, the ionosphere will fill a larger portion of the sky than the neutral atmosphere. There will therefore be ionosphere delay over a larger range of elevation angles (α) than there will be neutral atmosphere delay (β). Figure is not to scale.

all receiver altitudes and elevation angles examined.

Typical Features in Total Delay

Figure 6-12 displays the control total signal delay profiles for four different receiver altitudes. The total delay is almost exactly a sum of the neutral atmosphere and ionosphere profiles. We do expect some differences between the sum of the two components and the total delay because of the ray-tracing implementation; this difference primarily manifests itself in the magnitude of the spike in delay right at the horizon. This is likely because the refraction in the neutral atmosphere is very sensitive to the precise angle of the signal entering, and by incorporating the ionosphere we have changed how the signal travels through the neutral atmosphere. The neutral atmosphere is only dominant near the horizon, even for a receiver on the surface of the Earth.

Figure 6-13 displays the total signal delay in a contour plot that encompasses all receiver altitudes and elevation angles examined.

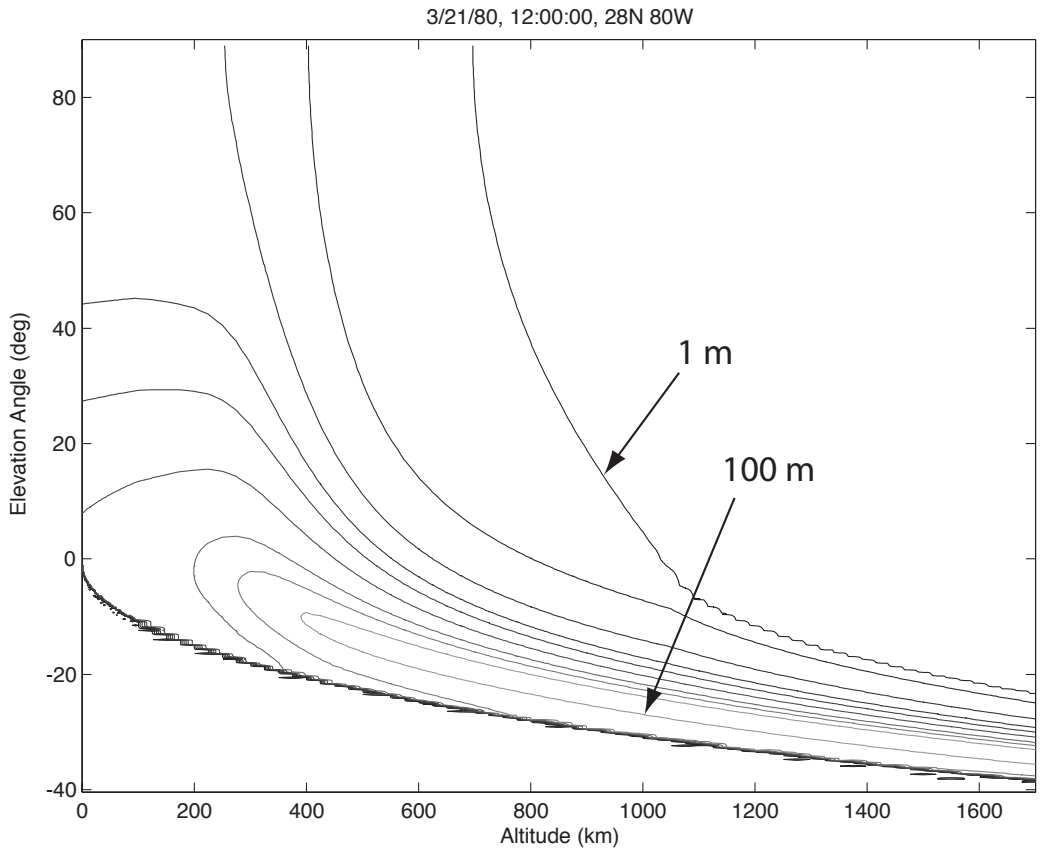

Figure 6-11: A contour plot of ionosphere signal delay as a function of altitude and geometric elevation angle. Contour lines are at 1, 5, 10, 15, 20, 30, 50, 70, 100 m delay. Note that at higher altitudes, the data resolution decreases, causing the bumpiness of the 1 m contour line. As with the neutral atmosphere delay, the curves have some distortion at the low elevation angles due to the MATLAB interpolation routines used in the contour plotting.

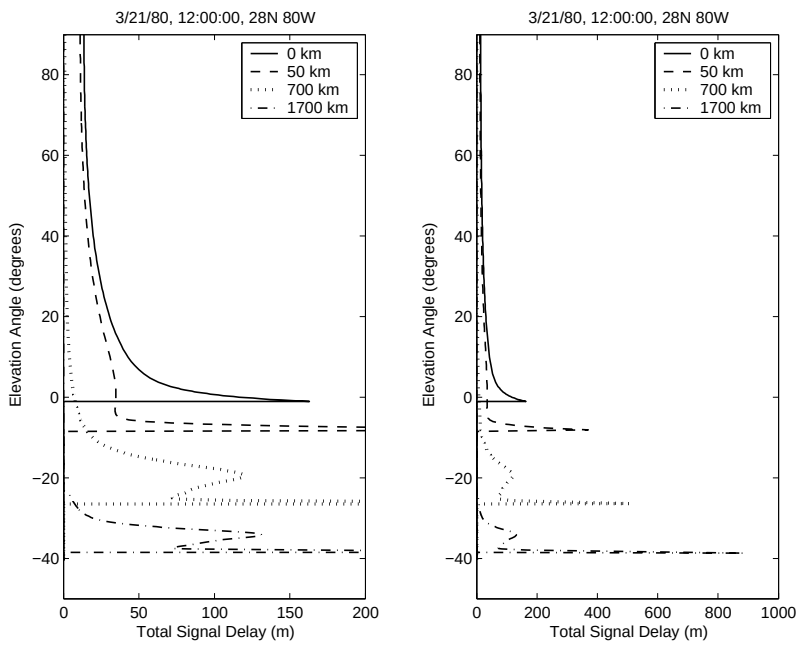

Figure 6-12: Total signal delay as a function of elevation angle for four different receiver altitudes and control conditions. LEFT: Close-in view. RIGHT: Full view.

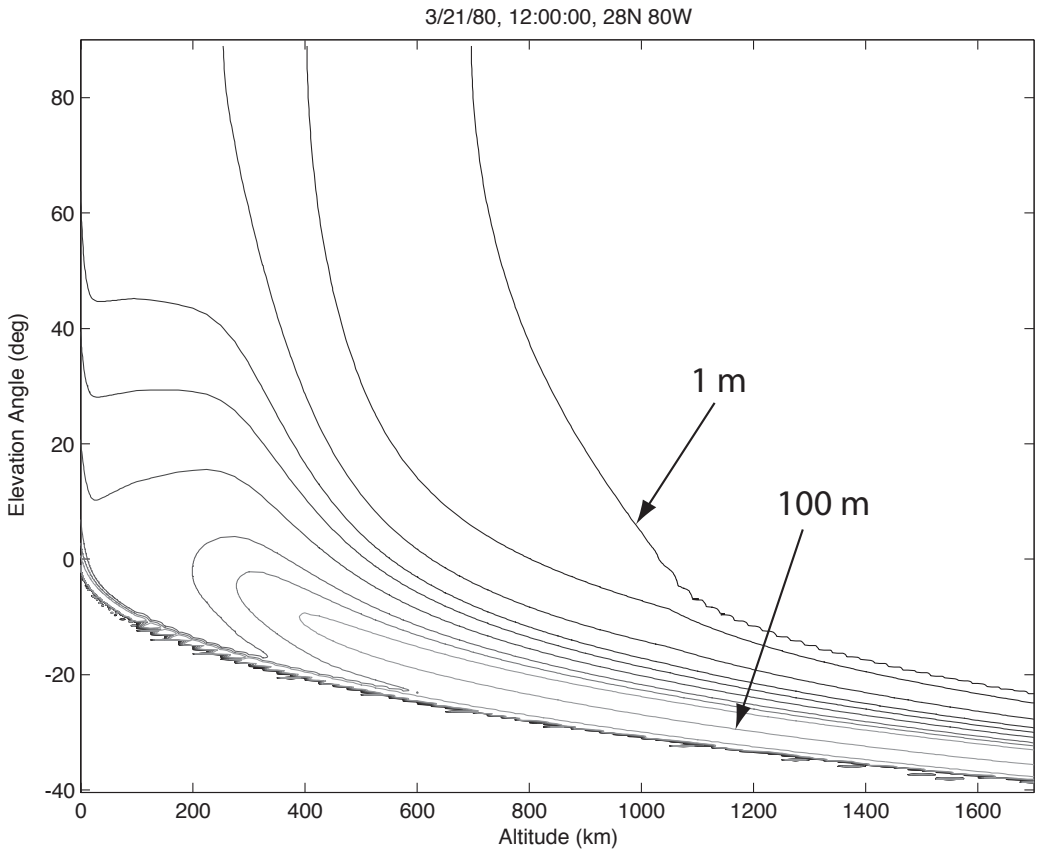

Figure 6-13: A contour plot of total signal delay as a function of altitude and geometric elevation angle. Contour lines are at 1, 5, 10, 15, 20, 30, 50, 70, 100 m delay. Note that this plot looks very similar to the ionosphere delay plot except at low receiver altitudes, where there is a kink in each of the otherwise smooth contour lines. Note that this plot also has the effects of resolution and interpolation visible.

6.1.3 Conditions Examined

In section 2.5.2 we saw that the temperature and pressure profiles in the neutral atmosphere vary as a function of season, time of day and latitude. In section 2.6.2 we saw that the electron density in the ionosphere varies as a function of season, time of day, geomagnetic latitude, solar activity level and geomagnetic storm conditions. For these reasons, section 6.1 examines signal delay as a function of season, time of day, latitude, solar activity level and satellite azimuth. The presented delays are all variations on the control profiles. For example, if azimuth is varied the rest of the parameters (date, time of day, solar activity, latitude) match those of the control profiles. Delays are presented at four different receiver altitudes: 0 km (the surface of the Earth), 350 km (the approximate altitude of peak electron density), 700 km (above the majority of the electron density) and 1700 km (the maximum altitude examined).

The last condition examined, satellite azimuth, is a geometrical rather than a strictly physical consideration. As the receiver gains in altitude, the relevant atmosphere and ionosphere that the signal passes through is farther ahead or behind in local time (if the satellite is to the east or west), or at a more distant latitude (if the satellite is to the north or south).

Understanding how these conditions affect the signal delay to a GPS receiver aboard a spacecraft allows us to determine the factors that need to be incorporated into any delay compensation model. It also tells us what series of conditions a receiver must be tested over in order to verify that the receiver signal delay compensation performs adequately.

In the following figures, there is no neutral atmosphere signal delay recorded at 1700 km, although this delay should be present (as is evident from the curve for the total delay at 1700 km in figure 6-12). As was mentioned in section 6.1.2, this is because the range of elevation angles at which there is any delay for a receiver at the upper altitudes is less than the resolution of the ray-tracing (0.1° apparent elevation angle).

6.1.4 Delay as a Function of Season

Figures 6-14 and 6-15 show how the signal delay varies as a function of season. Signal delays were compared on four dates: 3/21/80, 6/21/80, 9/21/80, 12/21/80 — roughly the equinoxes and the solstices, which are the approximate starting dates of the four seasons. Since the control profiles were made for a receiver in the Northern

hemisphere, 3/21/80 is the start of spring, 6/21/80 is the start of summer, 9/21/80 is the start of fall and 12/21/80 the start of winter.

There is a distinct ionosphere delay change with season. Delay is highest in the early spring and lowest in the early summer, just as electron density is highest in the early spring and lowest in the early summer, although the precise ordering of which season has the most delay depends on elevation angle. The elevation angle of peak delay is due both to the change in altitude of the F2-layer peak with season and to how the relative densities of the layers of the ionosphere differ with season. For example, the F2-layer peak is at its lowest altitude in the winter, and we see that the elevation angle at which the early winter delay peaks is accordingly lower. In the early summer the F2-layer density decreases proportionally more than the other layers, and we see that the elevation angle of the peak ionosphere delay lowers in the early summer.

The neutral atmosphere is only marginally affected by changes in season.

6.1.5 Delay as a Function of Solar Activity

Figures 6-16 and 6-17 show how the signal delay varies as a function of solar activity level, as defined by the twelve-month running mean Zurich sunspot number (Rz) and ionospheric index (IG). These values were obtained from the IRI-2001 code. Signal delays are compared on three dates: 3/21/80, a period of high solar activity in the Sun's 11 year cycle (IG = 159.3; Rz = 160.9), 3/21/78, a period of medium solar activity (IG = 79.9; Rz = 69.6), and 3/12/86, a period of low solar activity (IG = 7.6; Rz = 13.0).

There are two features to note. First, solar activity increases the magnitude of the ionosphere signal delay. This is because the electron density in the ionosphere increases with solar activity levels. Second, solar activity affects the elevation angle at which the signal delay peaks. The altitude of the F2-layer of the ionosphere increases with solar activity levels (see figure 2-4), thus raising the elevation angle at which a signal will pass tangent to this layer for a given receiver altitude. (The signal experiences peak ionosphere delay when the signal passes tangent to this layer, as seen in figure 6-9.)

The neutral atmosphere delay is negligibly affected by solar activity level.

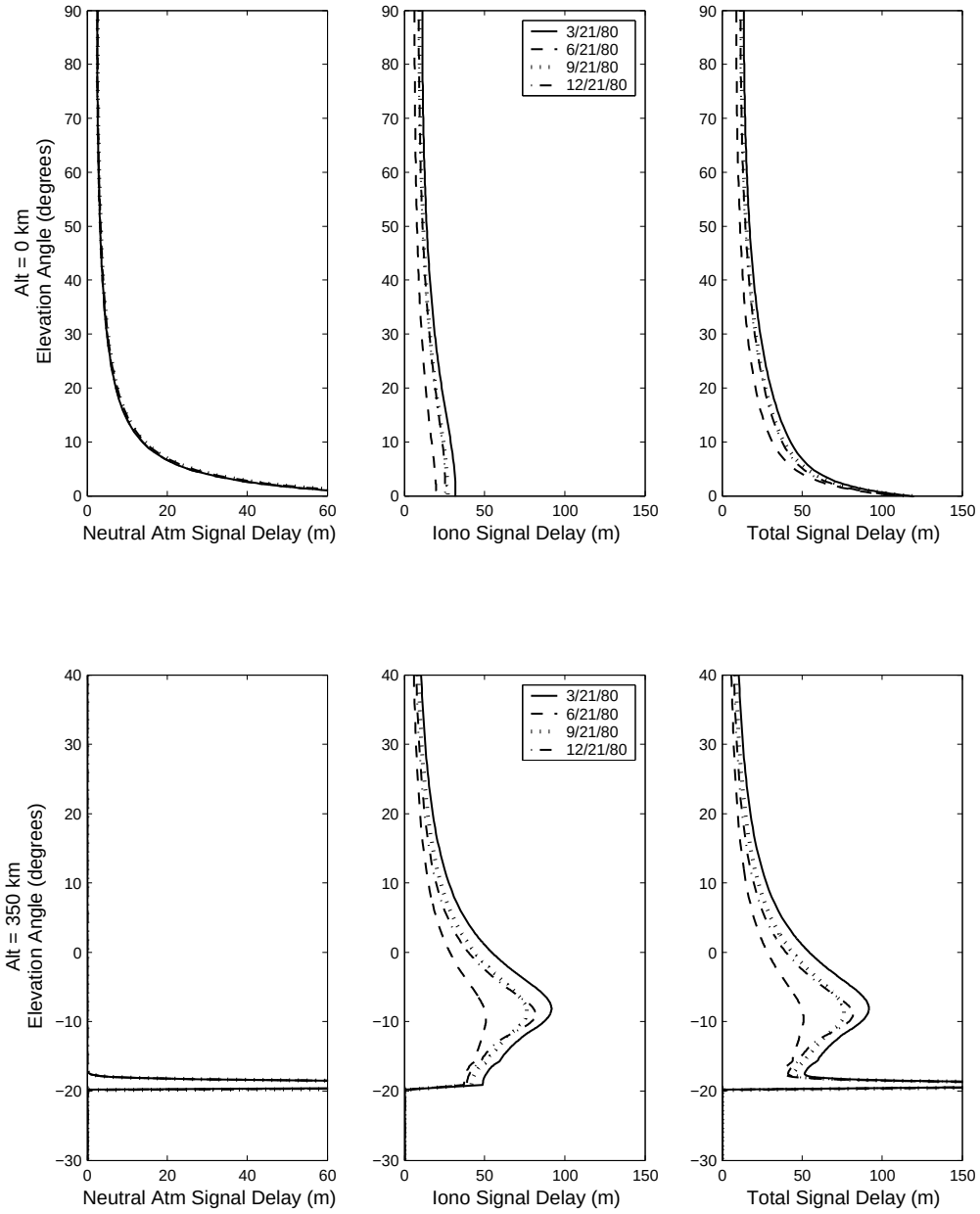

Figure 6-14: Neutral atmosphere, ionosphere and total signal delay as a function of **season** for 0 km and 350 km receiver altitude. The delay during the summer solstice (6/21/80) is indicated with a dashed line, and the delay during the winter solstice (12/21/80) is indicated with a dash-dotted line.

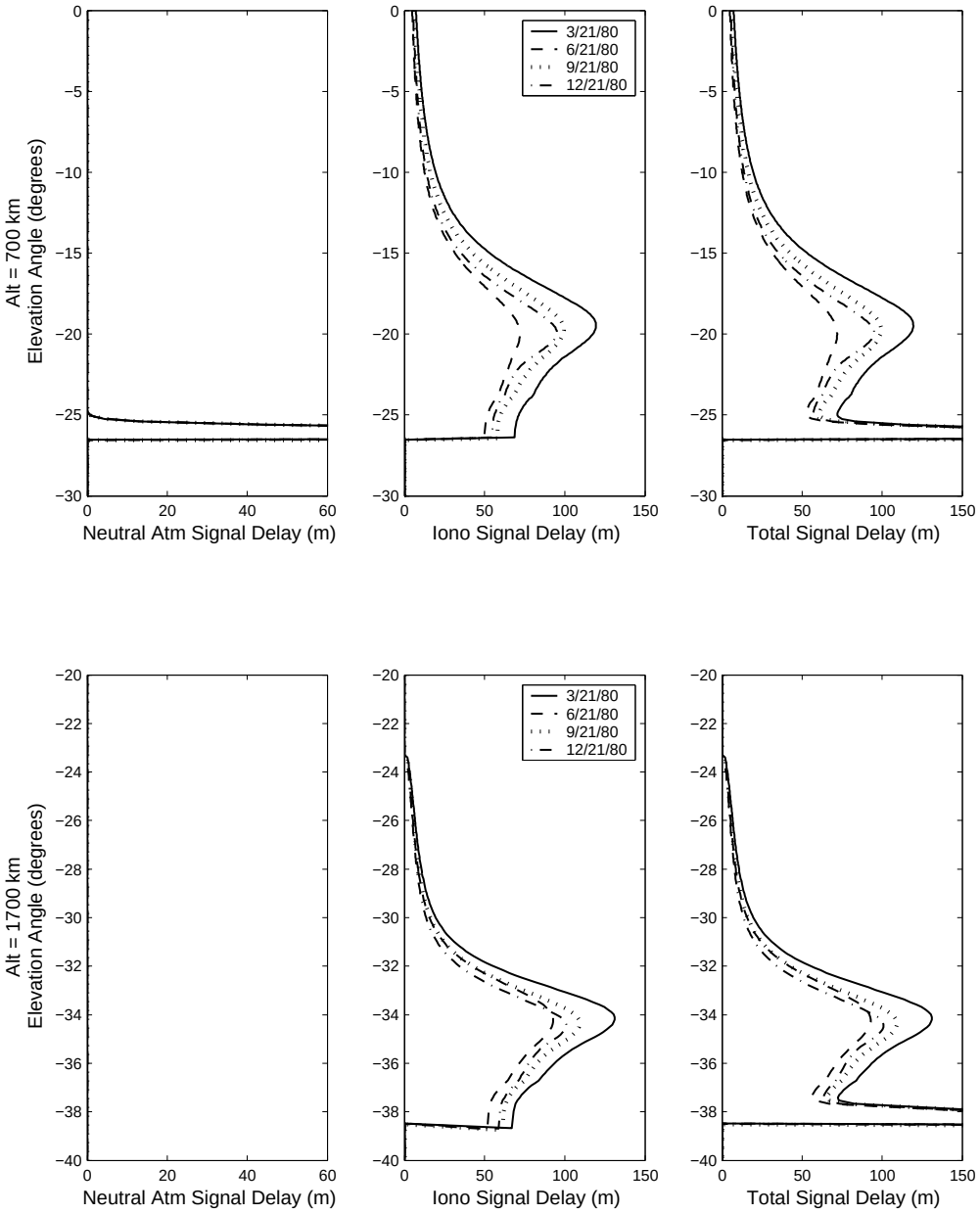

Figure 6-15: Neutral atmosphere, ionosphere and total signal delay as a function of **season** for 700 km and 1700 km receiver altitude. The delay during the summer solstice (6/21/80) is indicated with a dashed line, and the delay during the winter solstice (12/21/80) is indicated with a dash-dotted line.

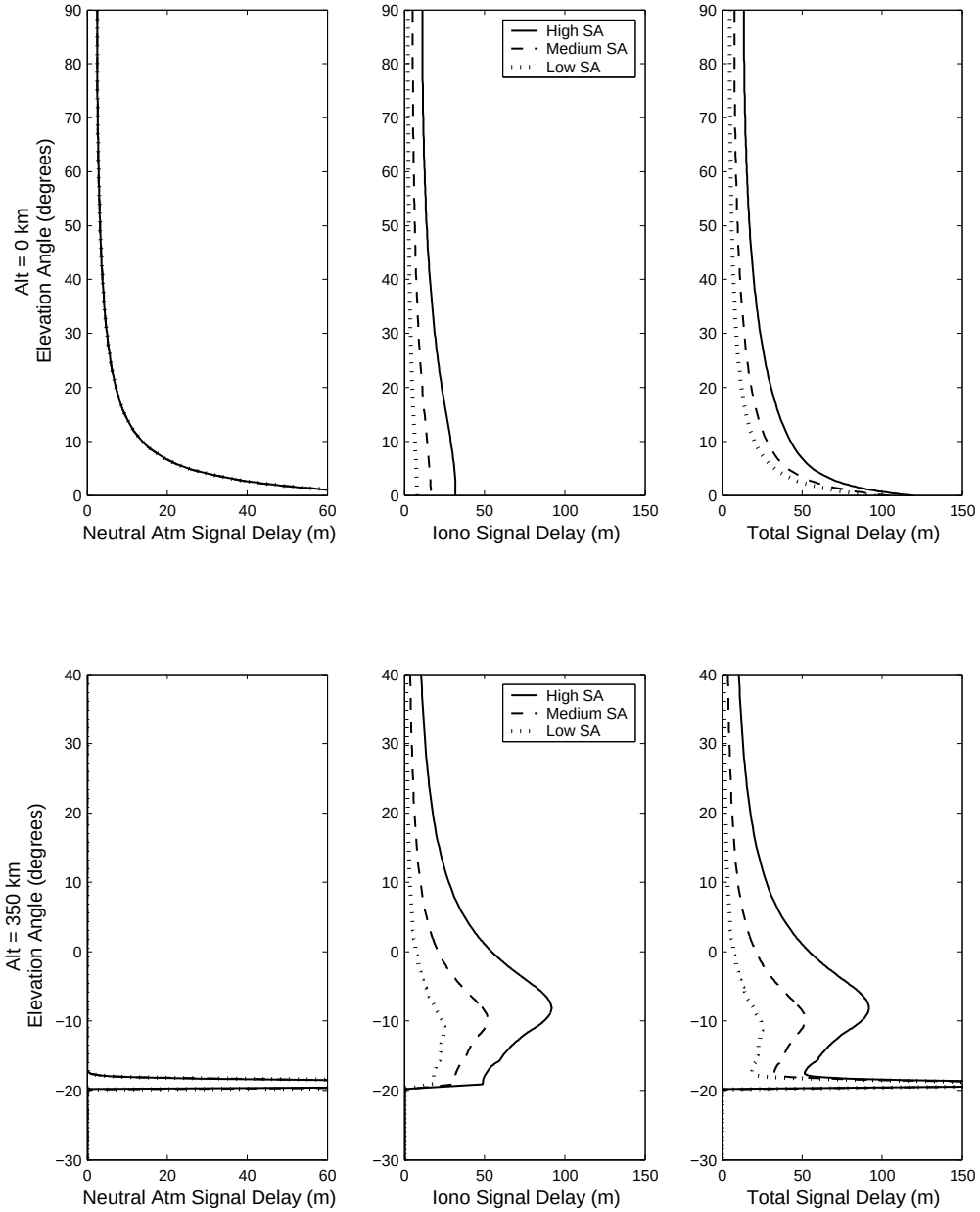

Figure 6-16: Neutral atmosphere, ionosphere and total signal delay as a function of **solar activity** at 0 km and 350 km receiver altitude.

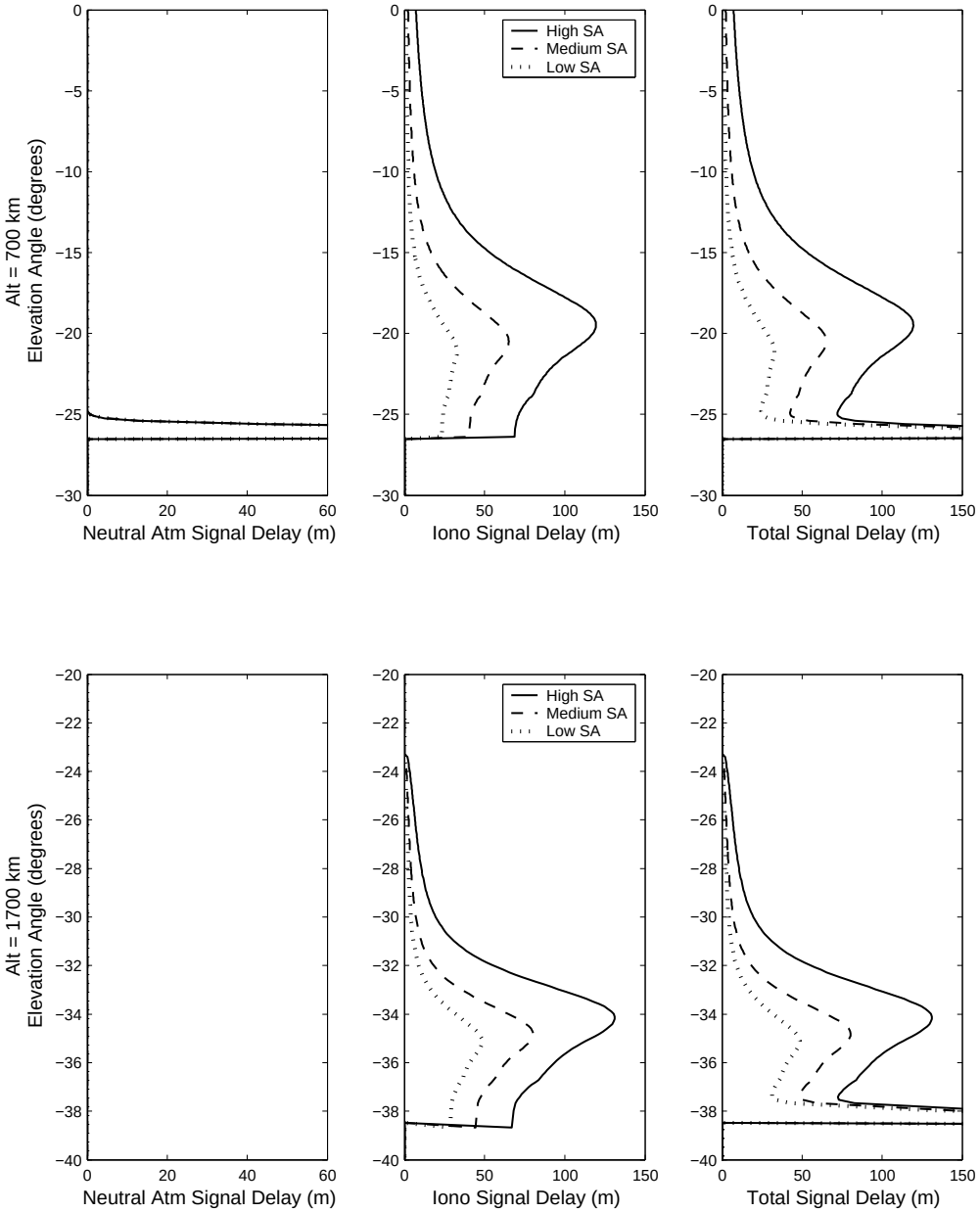

Figure 6-17: Neutral atmosphere, ionosphere and total signal delay as a function of **solar activity** at 700 km and 1700 km receiver altitude.

6.1.6 Delay as a Function of Time of Day

Figures 6-18 and 6-19 show how the signal delay varies as a function of time of day. Signal delays are compared at six different times throughout the day, spaced at four-hour intervals starting at midnight. Recall that since the control signals are received from a satellite at 90° azimuth, as the receiver altitude increases the receiver will be looking further "ahead" in time — the relevant ionosphere will have a local time later in the day than the receiver.

The neutral atmosphere is only marginally affected by changes in the time of day. The ionosphere, however, does have a dependence on time of day. Recall from section 2.6.2 that the electron density in the ionosphere is lowest at dawn, builds up throughout the day, and then begins to decrease in the afternoon when the ion recombination rate exceeds new electron and ion creation from solar radiation. We see this pattern in the ionosphere signal delay: delay is greatest at noon local time and decreases the later in the day one gets, with the minimum occurring at 4 am local time, right before dawn. Also note that the elevation angle of the peak delay is lower than average at 8 am local time for receivers at 350 km and 700 km altitude, and at 4 am local time for receivers at 700 and 1700 km altitude. This is because the altitude of the F-layer drops in the early morning. (See figure 2-4.)

6.1.7 Delay as a Function of Latitude

Figures 6-20 and 6-21 show how signal delay varies as a function of receiver latitude, comparing delays at 0°N, 15°N, 28°N, 45°N, 60°N and 75°N.

Receiver altitude has a strong effect on how much the latitude curves differentiate themselves. For a receiver on the surface of the Earth, the delay curves have exactly the same shape but different magnitudes of ionosphere and therefore total delay. This follows directly from the results in figure 2-4; electron density roughly decreases with latitude.

At 350 km receiver altitude, the curves continue to have mostly the same shape. Since the receiver is in the thickest part of the ionosphere, variations in latitude still have little effect on how the delay changes with elevation angle. One noticeable difference is that the equatorial curve begins to peak at a higher elevation angle. This is because the F-layer peak is at a higher altitude at the equator than at other latitudes.

Once the receiver is above the bulk of the ionosphere we begin to notice the effects of the differently shaped electron density profiles at the different latitudes

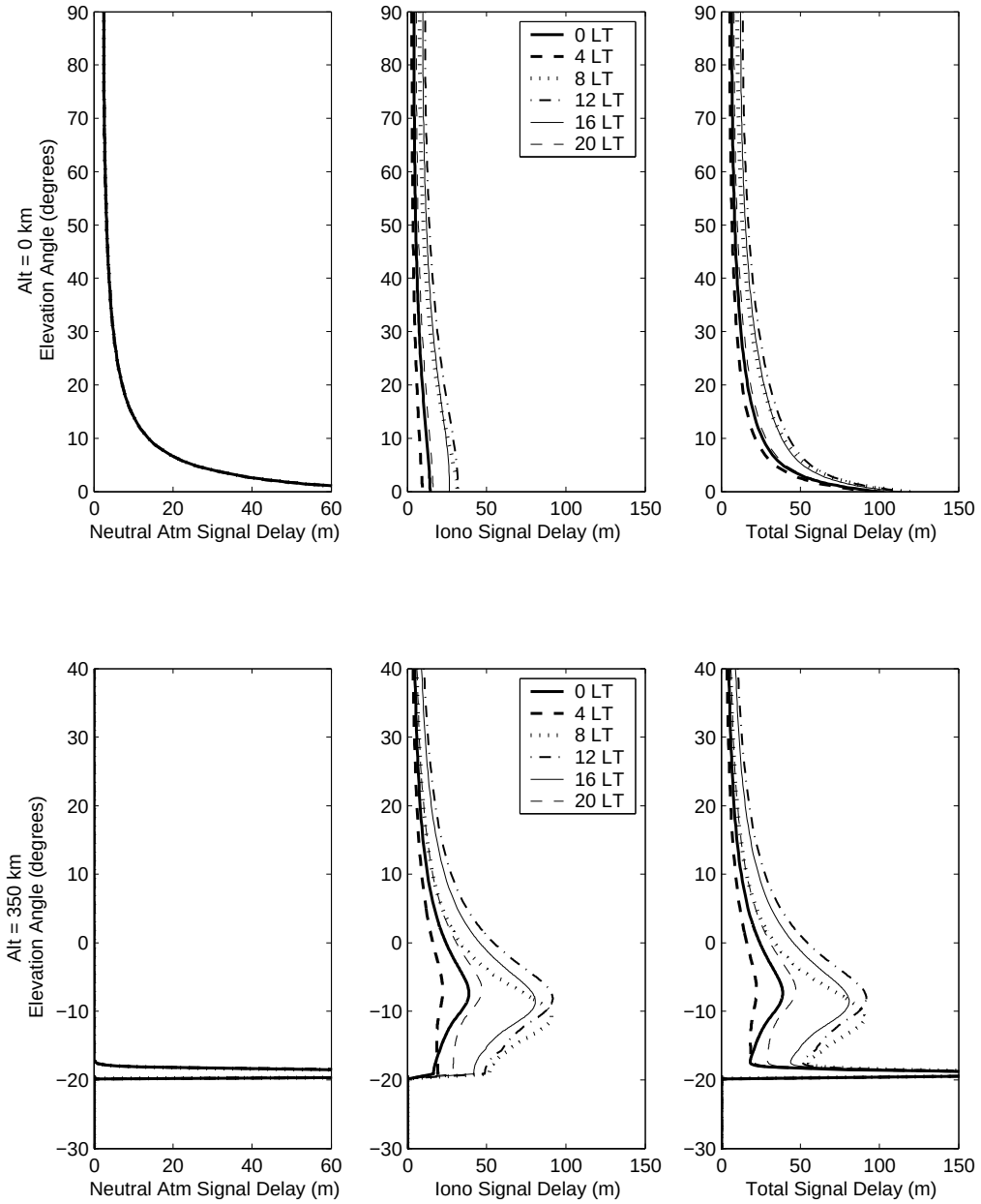

Figure 6-18: Neutral atmosphere, ionosphere and total signal delay as a function of **time of day** at the receiver for 0 km and 350 km receiver altitude.

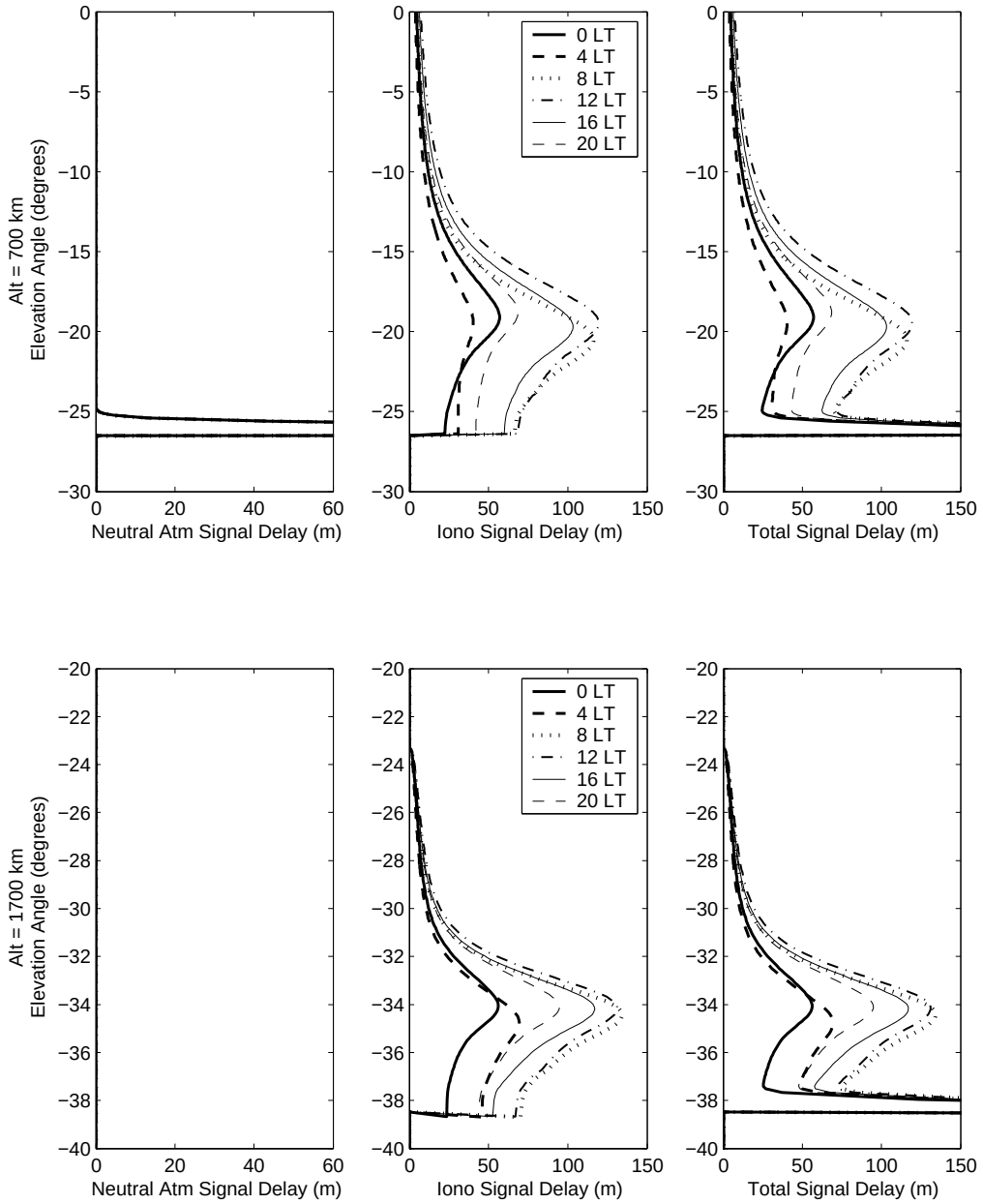

Figure 6-19: Neutral atmosphere, ionosphere and total signal delay as a function of **time of day** at the receiver for 700 km and 1700 km receiver altitude.

(figure 2-4). The profiles at the polar latitudes are flatter — the electron density does not decrease drastically with altitude. We therefore see a larger ionosphere delay at higher elevation angles. However, since the F-layer electron density is smaller at these latitudes, the delay at the peak elevation angle remains lower. We again see that the equatorial curve peaks at a higher elevation angle than the other latitudes.

The neutral atmosphere delay does vary with latitude. The neutral atmosphere delay is typically greatest at the equator and smallest at 75°N, although this trend does not always hold. The delay at 28°N is roughly the average delay over all latitudes. The neutral atmosphere delay as a function of latitude will be discussed further in section 7.1.2.

6.1.8 Delay as a Function of Satellite Azimuth

Figures 6-23 and 6-24 show how signal delay varies as a function of GPS satellite azimuth. At the lower altitudes there is some variation in the ionosphere delay with azimuth, but the differences do not become significant until the receiver reaches higher altitudes. This is because at higher altitudes the relevant ionosphere (the point where the signal crosses the densest layer of the ionosphere) is farther away from the receiver site. For a receiver on the surface of the Earth, the relevant ionosphere for a signal on the horizon is $\approx 14°$ away. For a receiver at 1700 km altitude, the signal is tangent to the thickest layer of the ionosphere $\approx 34°$ away from the receiver. At altitude, therefore, time of day considerations (for satellites to the east or west) and latitude considerations (for satellites to the north or south) become important.

The electron density results of figure 2-4 explain the variations we expect to see. For example, at 350 km receiver altitude the ionosphere signal delay increases when receiving a signal from the south (of 28°N); the electron density is largest at the equatorial latitudes. At 1700 km receiver altitude, the signal from the south has a peak in ionosphere delay at a higher elevation angle than the other three; the equatorial F-layer is at a much higher altitude than the F-layer at other latitudes. The signal from the north has a significantly different profile with respect to elevation angle than the other signals at this altitude; the electron density profile at the polar latitudes has significantly less structure, with no discernible decrease in electron density with altitude, than the profiles at the middle and equatorial latitudes. The signal from the west is passing through the ionosphere before noon, local time, whereas the signal from the east is passing through the ionosphere after noon, local time, causing a lower and higher elevation angle of peak delay, and a lower and higher magnitude of peak

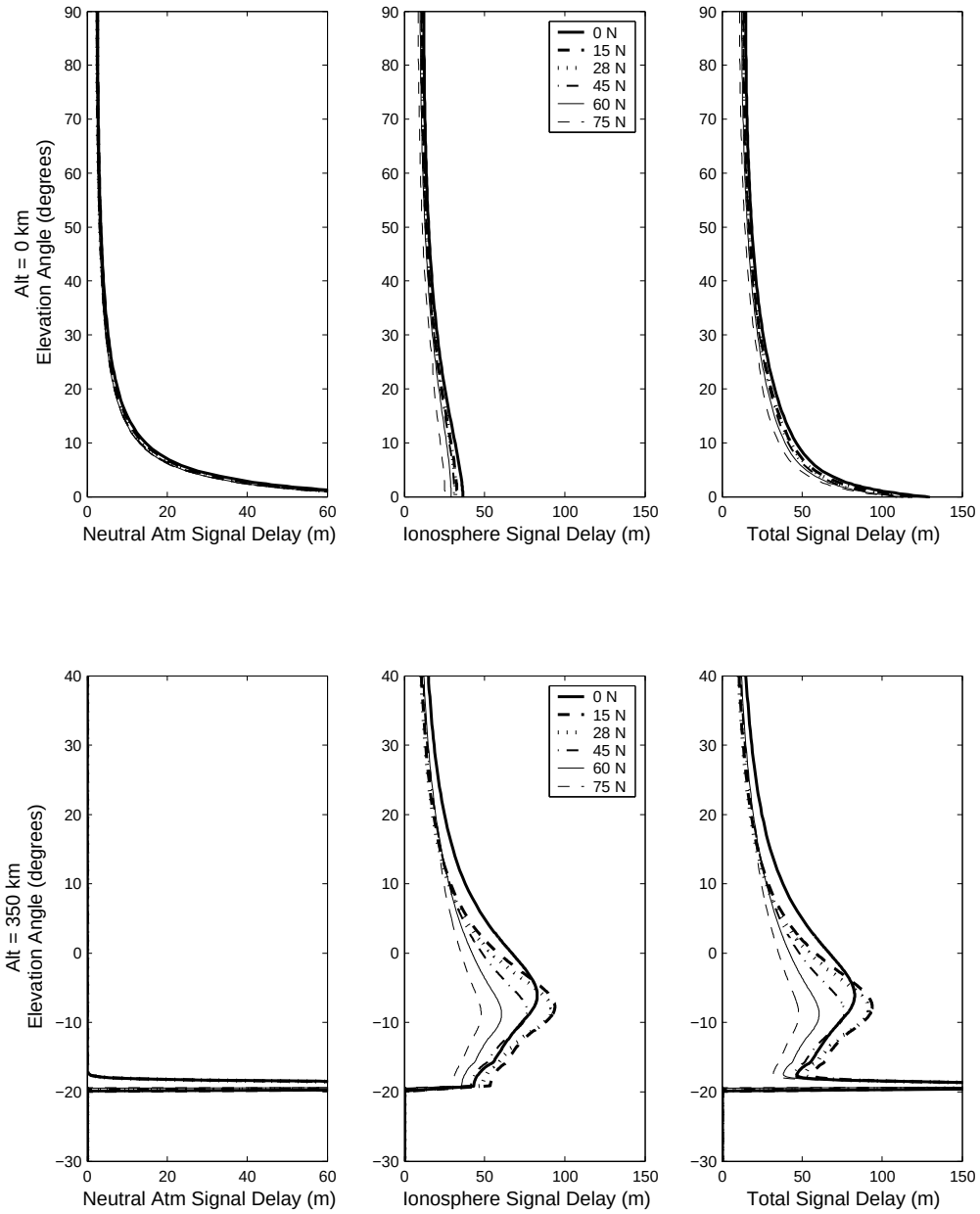

Figure 6-20: Neutral atmosphere, ionosphere and total signal delay as a function of receiver **latitude** at 0 km and 350 km receiver altitude. Delay at 15°N is indicated by a thick dashed line; delay at 45°N by a thick dash-dotted line; and delay at 75°N by a thin dashed line.

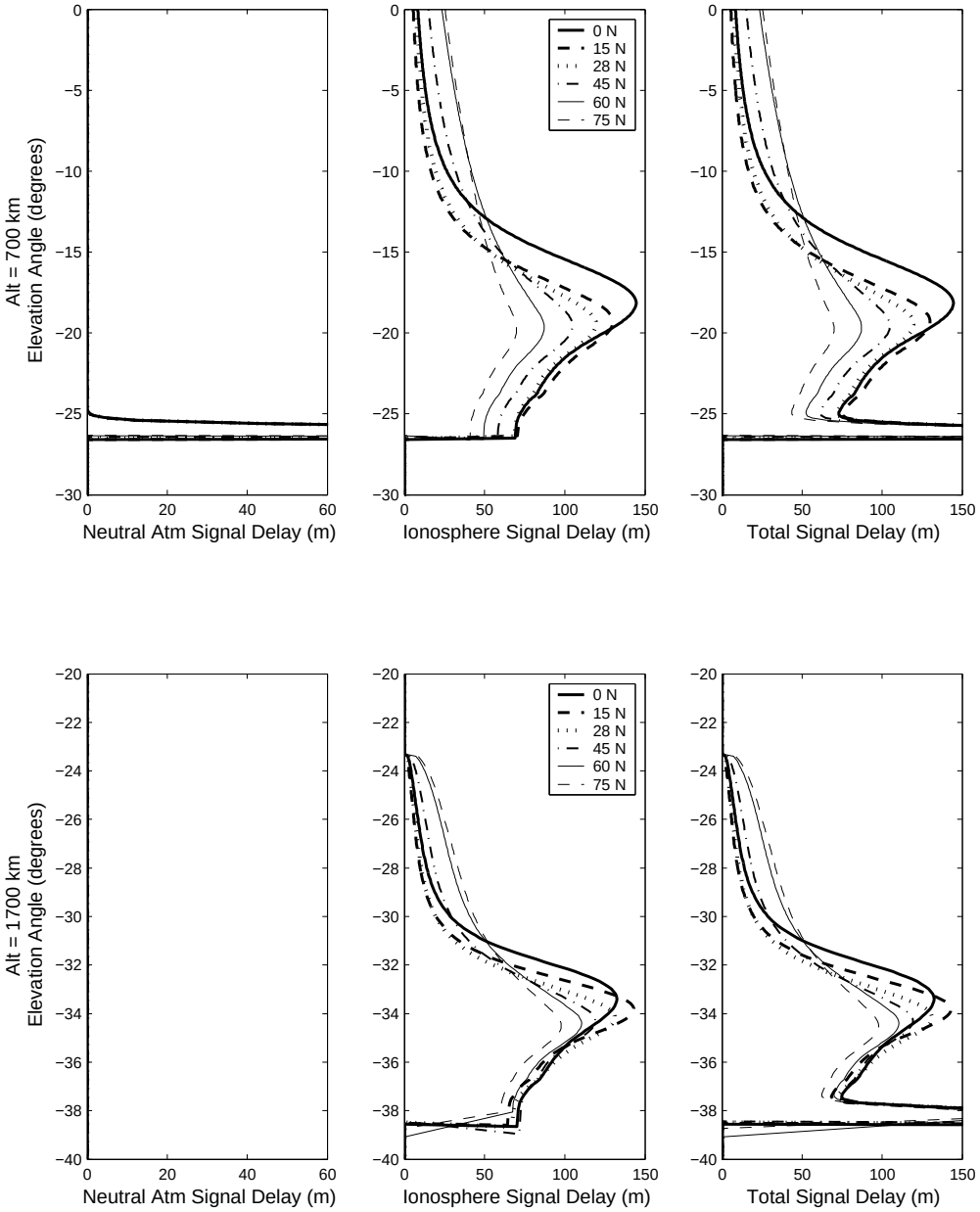

Figure 6-21: Neutral atmosphere, ionosphere and total signal delay as a function of receiver **latitude** at 700 km and 1700 km receiver altitude. Delay at 15°N is indicated by a thick dashed line; delay at 45°N by a thick dash-dotted line; and delay at 75°N by a thin dashed line.

delay, respectively. The altitude of the peak electron density is higher after noon than before noon, and that the electron density is greater after noon than before noon.

The neutral atmosphere delay is not noticeably affected by variations in azimuth. This is in part because at lower receiver altitudes the signal passes through neutral atmosphere close to the receiver regardless of azimuth, and because the neutral atmosphere delay did not show much variation with time of day or receiver latitude.

The "relevant" neutral atmosphere is the neutral atmosphere with the highest refractivity that the signal passes through, almost always located at the point of the signal's closest approach to the Earth. Likewise, the "relevant" ionosphere is where the signal passes through the densest layer of the ionosphere, the F2-layer peak. As a rough rule of thumb, for a spacecraft above the neutral atmosphere the relevant atmosphere is the same number of degrees away from the receiver as the magnitude of the elevation angle to the satellite. (See figure 6-22.) There is one problem with extending this approximation to the ionosphere: when the signal passes closer to the Earth than the altitude of the F2-layer peak, there will be two relevant points. The point of closest approach will be midway between these two points and will generally be a good approximation. However, since the electron density does not change linearly with latitude or time of day, this approximation will not always be accurate.

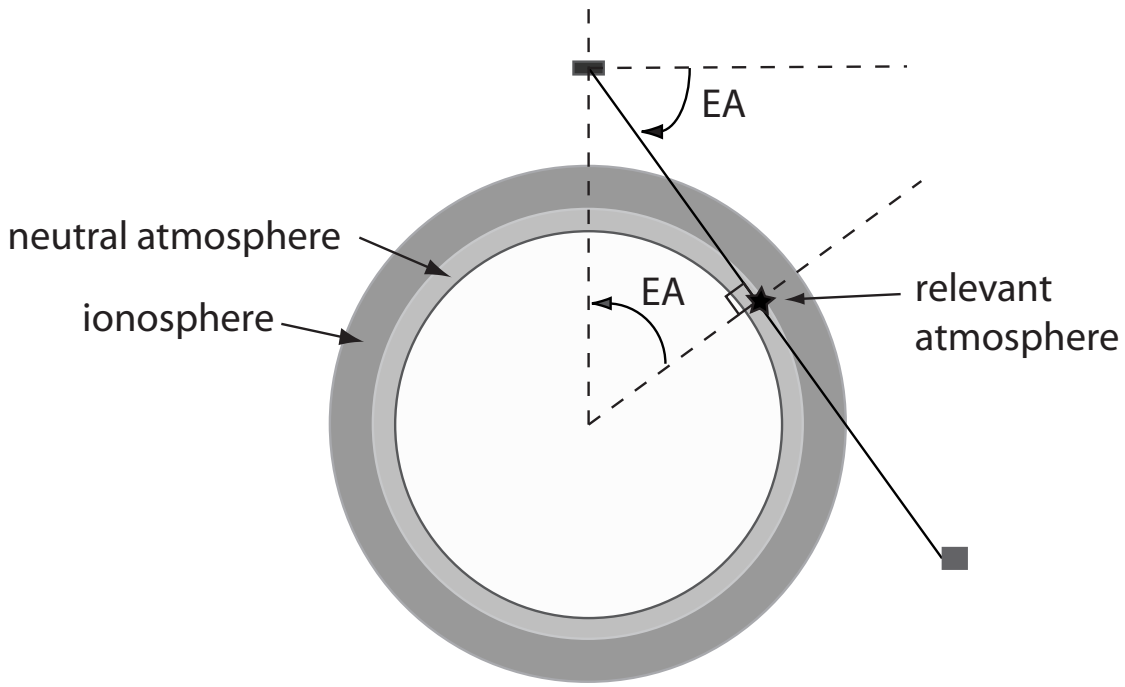

Figure 6-22: The "relevant" atmosphere for a spacecraft above the atmosphere is approximately the same number of degrees away from the receiver as the magnitude of the elevation angle to the satellite.

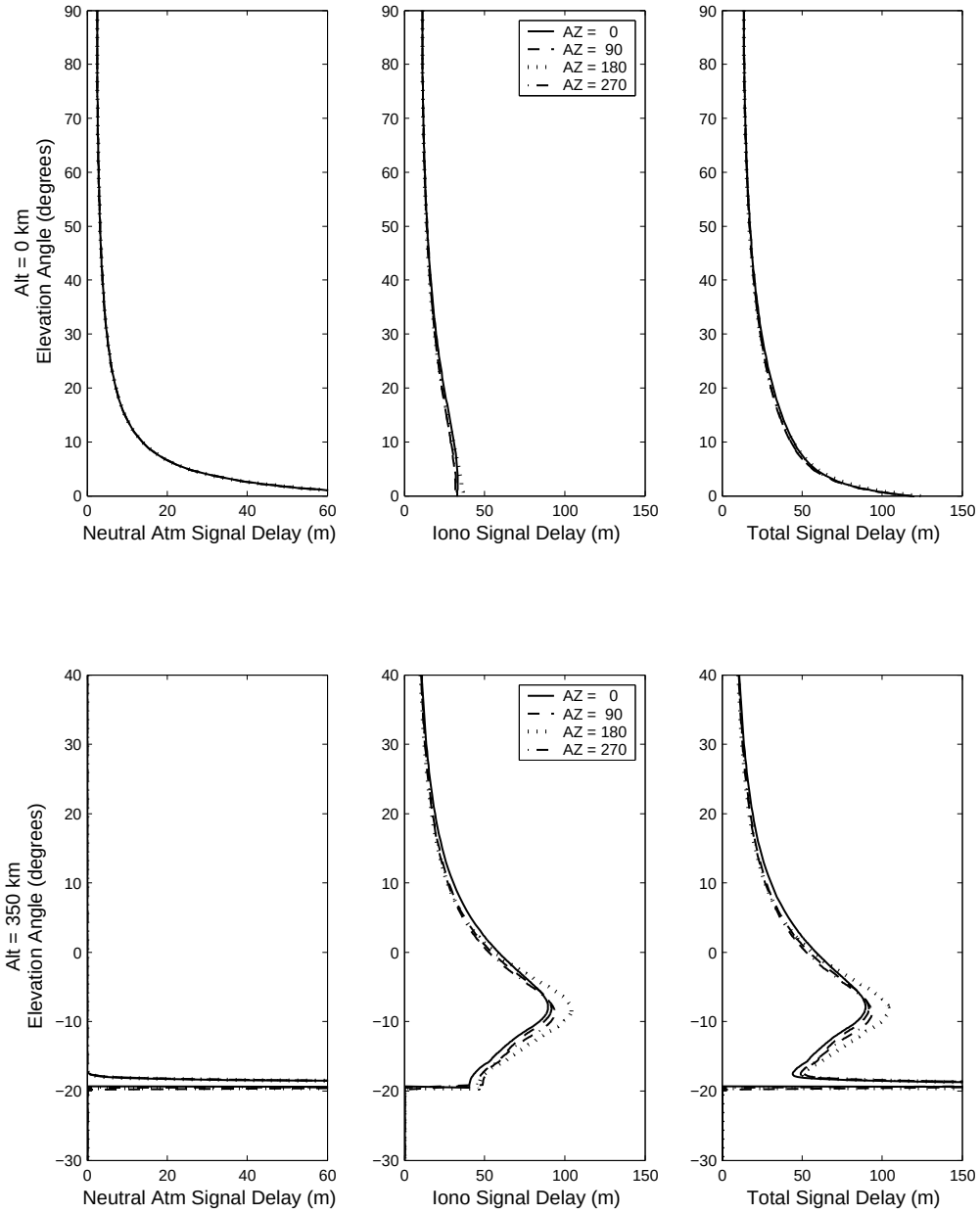

Figure 6-23: Neutral atmosphere, ionosphere and total signal delay as a function of satellite **azimuth** for 0 km and 350 km receiver altitude. Delay at 90° azimuth is indicated by a dashed line, and delay at 270° azimuth is indicated by a dash-dotted line.

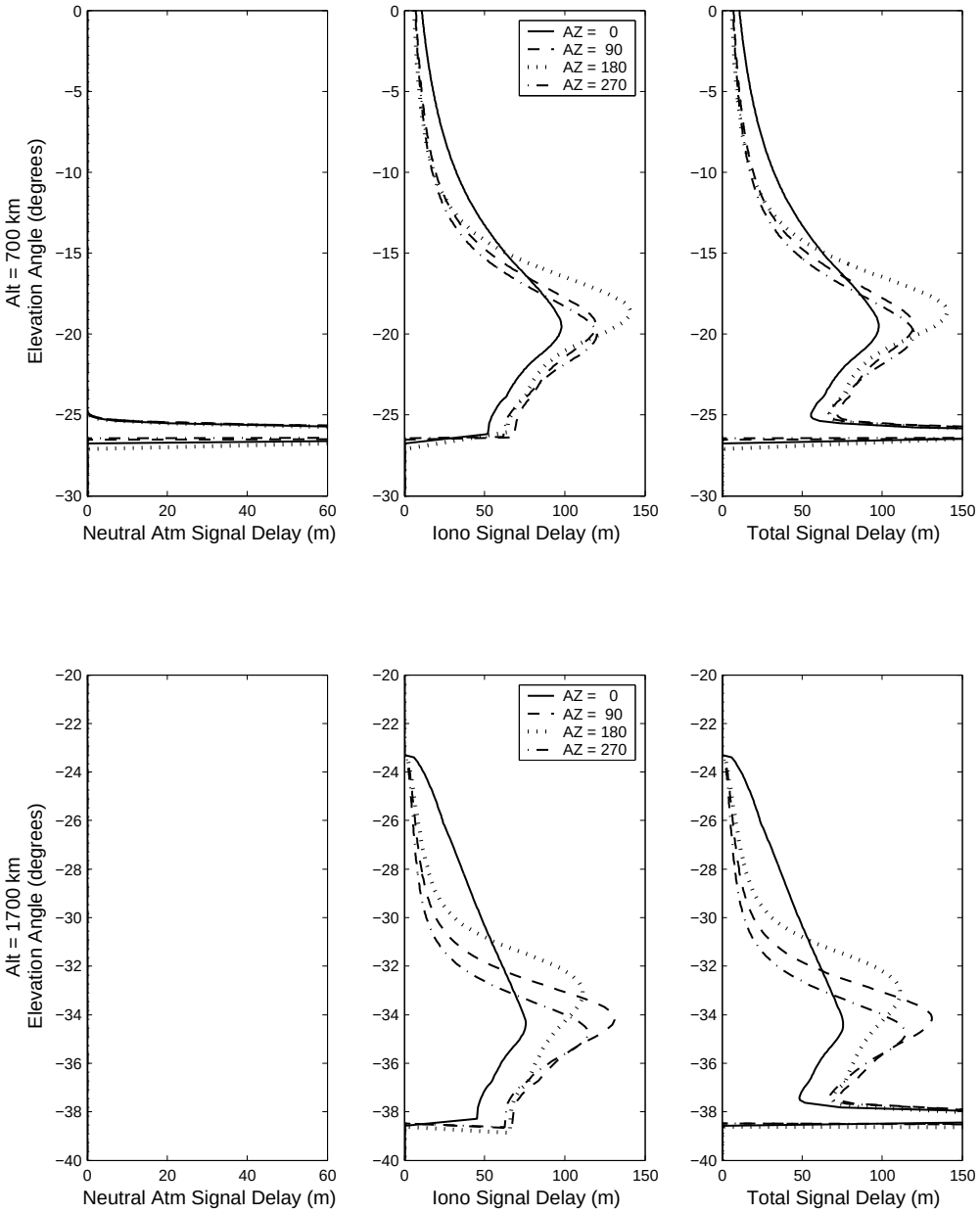

Figure 6-24: Neutral atmosphere, ionosphere and total signal delay as a function of satellite **azimuth** for 700 km and 1700 km receiver altitude. Delay at 90° azimuth is indicated by a dashed line, and delay at 270° azimuth is indicated by a dash-dotted line.

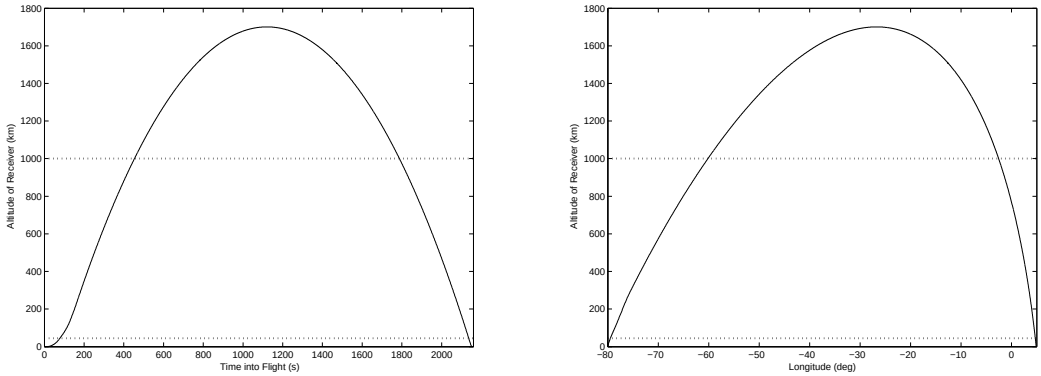

Figure 6-25: LEFT: Receiver altitude as a function of time into flight. RIGHT: Receiver altitude as a function of longitude. Lines indicate the top of the neutral atmosphere and the top of the ionosphere in both plots.

6.2 Delay as Experienced by a GPS Receiver Aboard a Space Vehicle

To examine the delays that signals to a space vehicle receiver would experience in flight, I flew a receiver through a sample trajectory covering launch, low Earth orbit and reentry altitudes, peaking at 1700 km altitude. This trajectory began at 28°N 80°W (roughly Kennedy Space Center, Florida) and travelled east along the 28°N line of latitude for approximately 36 minutes. Figure 6-25 displays receiver altitude as a function of time into flight and longitude.

I used a real distribution of GPS satellites taken from a GPS ephemeris for GPS week 191, time of week 0. Since the receiver flight went nearly a quarter of the way around the world, twenty-two GPS satellites were visible during flight. Figure 6-26 shows the latitude and longitude of satellites that are visible during the receiver flight, and has horizon lines to show which satellites will be visible during the three main parts of the receiver flight.

Figures 6-27, 6-28 and 6-29 show the geometrical elevation angle to each satellite during the receiver flight, as well as the geometric elevation angle of the horizon (in the spherical approximation). This data is presented in three figures with six to eight satellites each for clarity. The elevation angle to a satellite will not tell us everything about the signal delay during flight. The receiver will be traveling through many time zones, and so while the flight starts at noon, local time, it ends in the evening, local time. The azimuth to each satellite and the altitude of the receiver will also

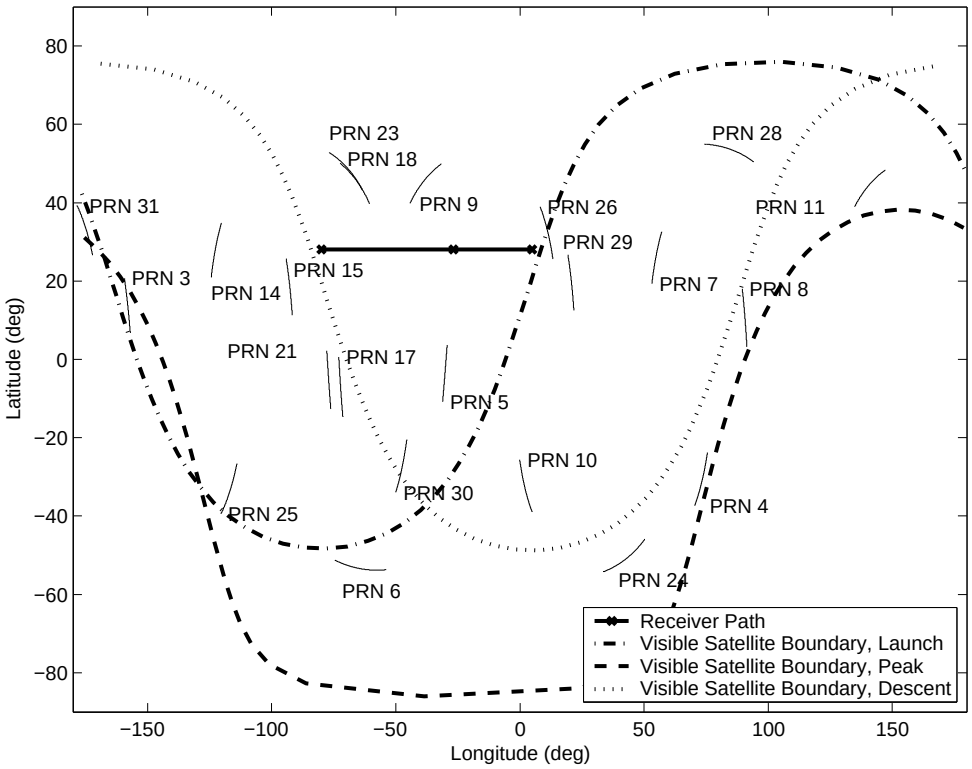

Figure 6-26: The trajectory of the receiver and the distribution of GPS satellites that will be visible during the flight. Labeling of satellite PRN number is at the starting latitude and longitude of the satellite path during the 36 minutes of receiver flight. The horizon lines are not at the latitude and longitude of the horizon, but at the latitude and longitude of a satellite at GPS satellite orbital altitude that will be visible at the geometrical horizon.

change constantly in flight. These three variables, in addition to elevation angle, will determine the variations observed in signal delay between satellites. However, elevation angle will remain the primary indicator of general signal delay levels.

6.2.1 Signal Delay

Figures 6-27, 6-28 and 6-29 show the neutral atmosphere, ionosphere and total signal delay during flight. There are three features to note.

First, the receiver is above the neutral atmosphere for the vast majority of the flight, from less than 100 seconds into flight until less than a minute before landing. The neutral atmosphere signal delay therefore only contributes a very sharp spike of signal delay right as a satellite rises above the horizon or right before a satellite sets below the horizon, except during the brief periods of launch and reentry. The delay spikes of satellites that rise or set during the main orbital portion of the flight are not all fully visible in figures 6-27, 6-28 and 6-29; the same resolution problem affecting the neutral atmosphere delay profiles discussed in section 6.1.2 for receivers at higher altitudes affects these plots as well. During the launch and reentry periods satellites that rise and set still experience a large increase in neutral atmosphere signal delay, but the peak is not as sharp, with the build up or drop down period extended over tens of seconds.

Second, the receiver is above the ionosphere from roughly 500 s into flight to 1800 s into flight. Ionosphere signal delay in this region therefore only occurs when a satellite is at negative elevation angles; you can see the ionosphere delay for many satellites approach zero at 500 s and reemerge around 1800 s. It is possible for a satellite to be visible at negative elevation angles for much of the flight; note PRN 8, PRN 11 and PRN 31 in figure 6-28. Signals from these satellites will always be delayed by the ionosphere, and experience a peak in ionosphere delay as the satellite is rising or setting. These peaks exhibit similar characteristics to the profiles of the previous section — they occur as the signal is tangent to the densest layer of the ionosphere, and therefore the magnitude of the signal delay is lower before (if the satellite is rising) or after (if the satellite is setting) the signal reaches this tangental point. Signals from most satellites will experience no ionosphere signal delay in this region between 500 s to 1800 s, or if they do are quickly rising or setting and will not experience signal delay for long.

Third, while the receiver is within or below the ionosphere, the potential ionosphere delay will be quite low compared to the delay potentially experienced in other

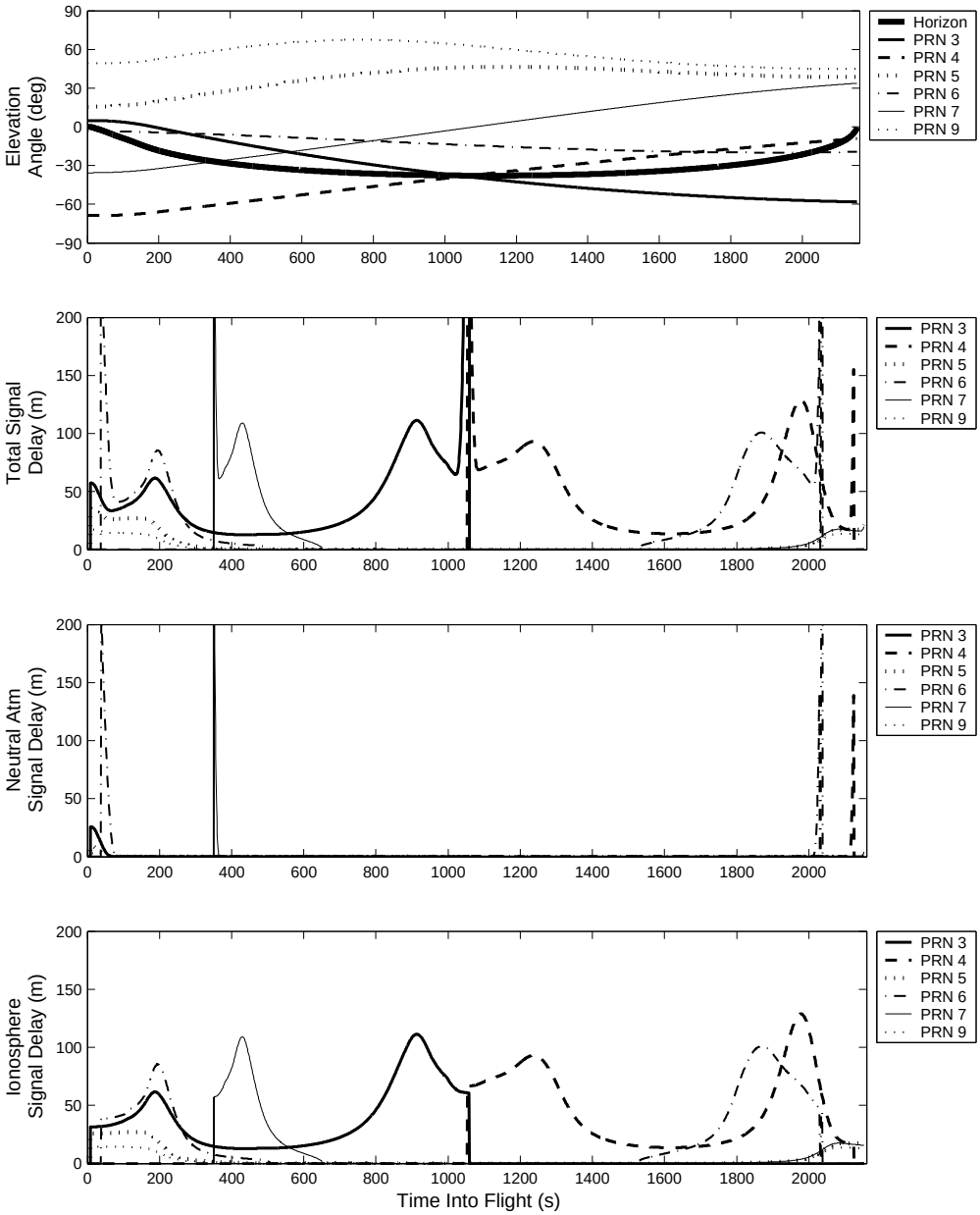

Figure 6-27: The neutral atmosphere, ionosphere and total signal delay experienced by signals from 6 satellites during the sample receiver flight. The top plot gives the elevation to each satellite for reference.

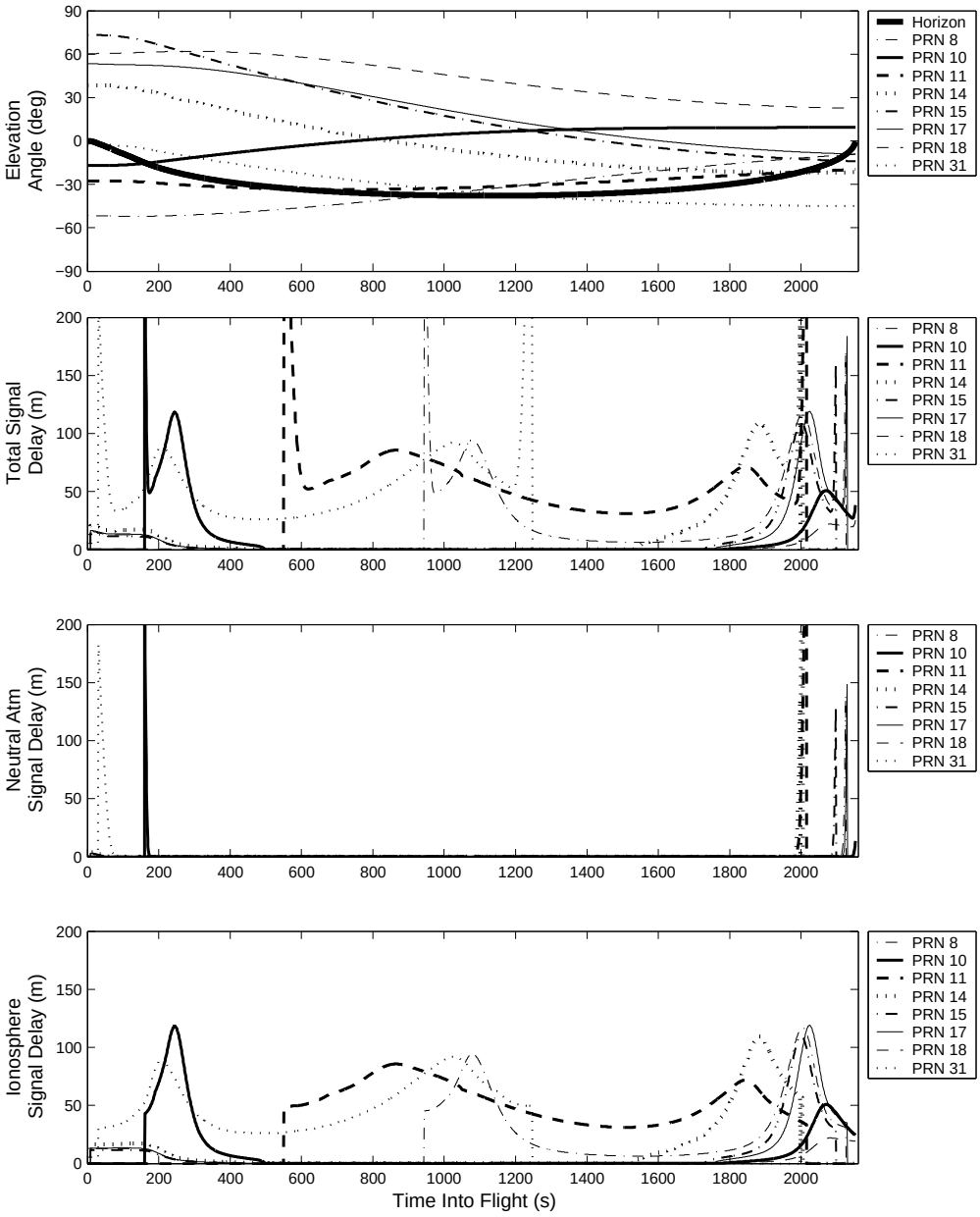

Figure 6-28: The neutral atmosphere, ionosphere and total signal delay experienced by signals from 8 satellites during the sample receiver flight. The top plot gives the elevation to each satellite for reference.

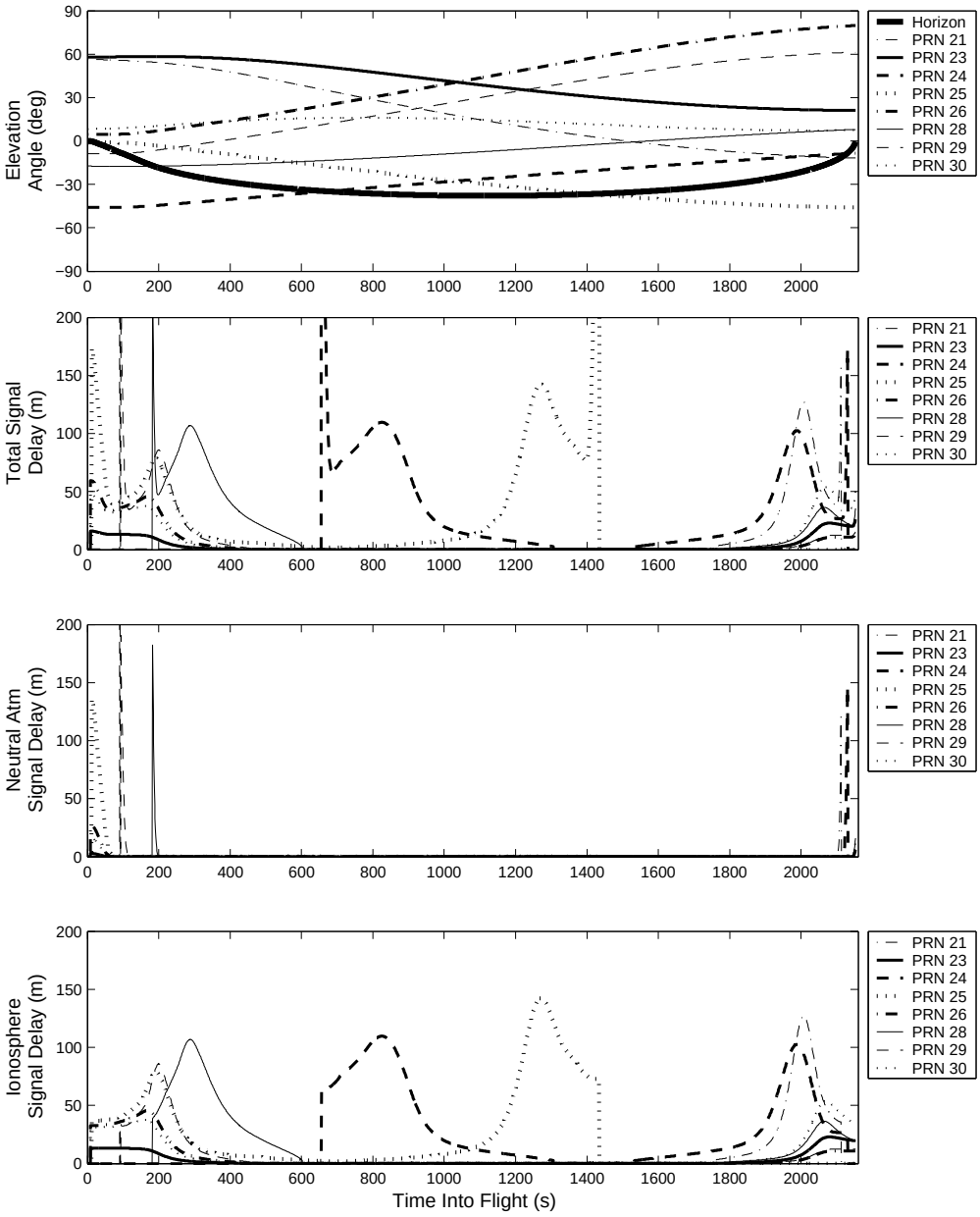

Figure 6-29: The neutral atmosphere, ionosphere and total signal delay experienced by signals from 8 satellites during the sample receiver flight. The top plot gives the elevation to each satellite for reference.

periods of the flight. For example, in figure 6-31 the ionosphere delay is less than 50 m during launch, even when a satellite rises, whereas during flight signals from the same satellites experience over 100 m of ionosphere delay (figure 6-27).

Signal Delay During Launch

Figures 6-31, 6-32 and 6-33 highlight the signal delays during the launch period of the receiver flight. Figure 6-30 provides the altitude of the receiver during this period. During launch, since the receiver spends time below both the neutral atmosphere and the ionosphere, signals can be delayed even if the satellite they originate from is not rising or setting. The delay on these signals is low compared to that possible during the orbital segment of the flight, where the originating satellite is typically rising or setting when any delay occurs. Also, when there is delay during the orbital segment of the flight, the signal must have passed "twice" through the atmosphere as was described in figure 1-2.

In general, neutral atmosphere delays start low and quickly go to zero as the receiver travels above the neutral atmosphere. A few satellites rise in the first 100 seconds of flight and exhibit relatively long periods of significant neutral atmosphere delay (compared to satellites that rise in the middle of the flight) since the receiver is still at low altitudes. The later into flight the satellite rises, the less time the signal from that satellite experiences neutral atmosphere delay (compare PRN 25 to PRN 29 in figure 6-33).

Ionosphere delays likewise start below 50 m, and tend to slowly rise as receiver altitude increases and the geometry allows signals to spend more time in the ionosphere. The ionosphere delays do not show spikes in delay when a satellite rises; instead, the delay starts at a given value, where it remains relatively constant.

Satellites only rise in this period — none set. These satellites are not rising because of their motion, but rather due to expanding range of elevation angles visible to the receiver.

Signal Delay During Reentry

Figures 6-35, 6-36 and 6-37 highlight the signal delays during the reentry period of the receiver flight. Figure 6-34 provides the altitude of the receiver during this period. The features of the delays during this period are similar to those during launch. In general, there is virtually no neutral atmosphere delay except for during the last 15 seconds, when the receiver travels below 50 km. Even then, delays are small. The

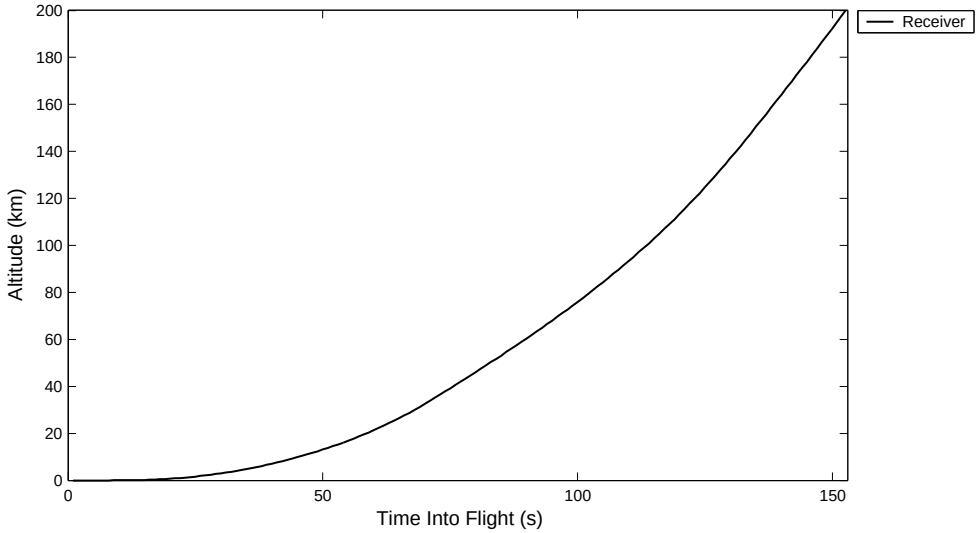

Figure 6-30: The altitude of the receiver during the launch period of the receiver flight.

delays from the satellites that set are similar in shape and magnitude to those of satellites that rise during launch.

Ionosphere delays, as during launch, remain below 50 m, but tend to slowly decrease as receiver altitude decreases and the geometry allows signals to spend less time in the ionosphere. The ionosphere delays do not show spikes in delay when a satellite sets; instead, the delay just drops to zero.

There are six satellites that set during this period. These satellites set not because of the satellites' motion, but rather because the visible region of the receiver is shrinking rapidly.

6.2.2 Delay Rates

Figures 6-38, 6-39 and 6-40 give the delay rates that would be introduced if the neutral atmosphere and ionosphere signal delay were not corrected for. These delay rates are the first time derivatives of the delays in figures 6-27, 6-28 and 6-29. Delay rates are significant right as a satellite sets or rises, from the neutral atmosphere delay spike and the subsequent drop to zero delay (or the start at zero delay, spike, and settle to normal values). It is unlikely that a receiver will acquire a satellite immediately after it becomes visible, and therefore the delay rates of rising satellites are less important than those of setting satellites.

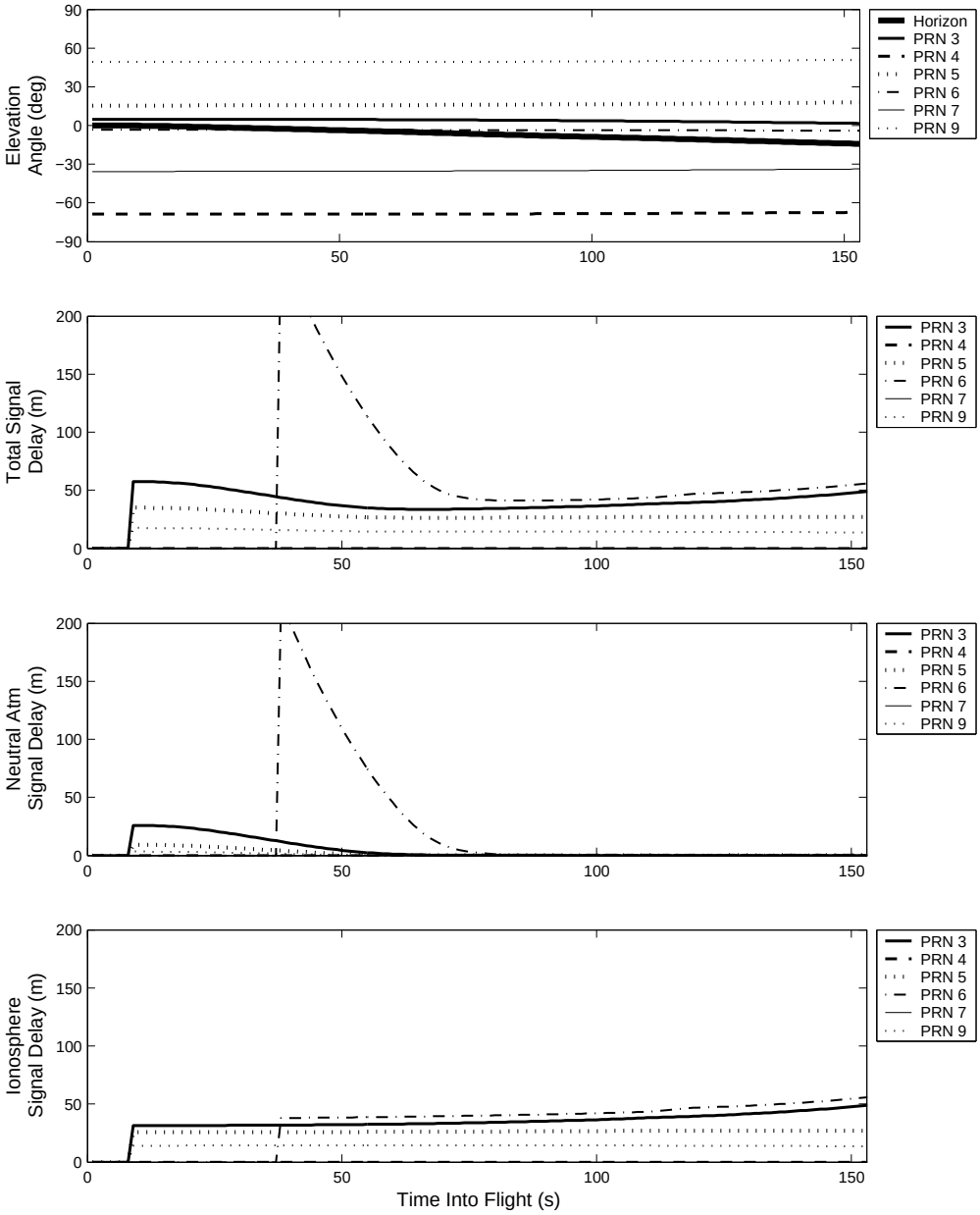

Figure 6-31: The neutral atmosphere, ionosphere and total signal delay experienced by signals from 6 satellites during the launch period of the sample receiver flight. Note that the receiver does not go above local sea level until 10 s into flight, and thus the ray-tracing program did not report any delay.

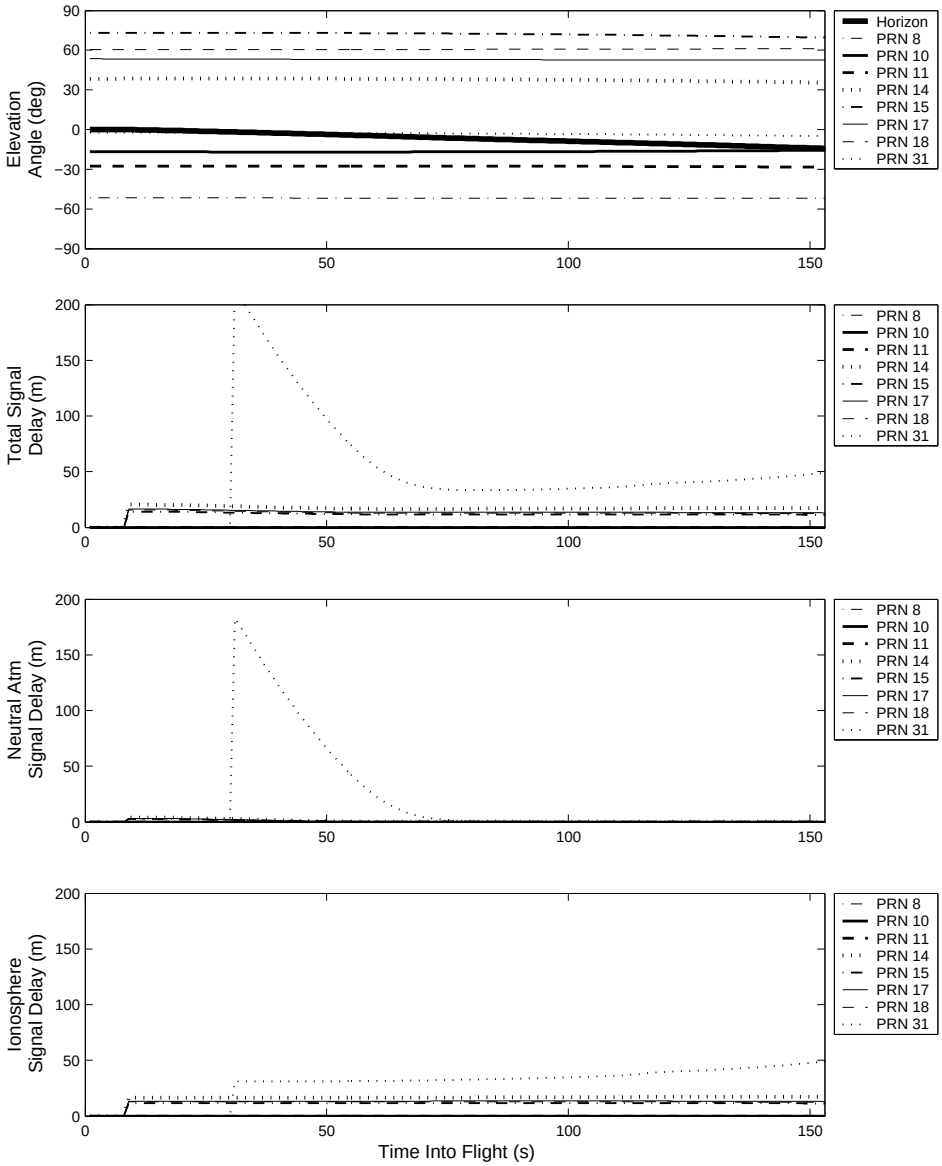

Figure 6-32: The neutral atmosphere, ionosphere and total signal delay experienced by signals from 8 satellites during the launch period of the sample receiver flight. Note that the receiver does not go above local sea level until 10 s into flight, and thus the ray-tracing program did not report any delay.

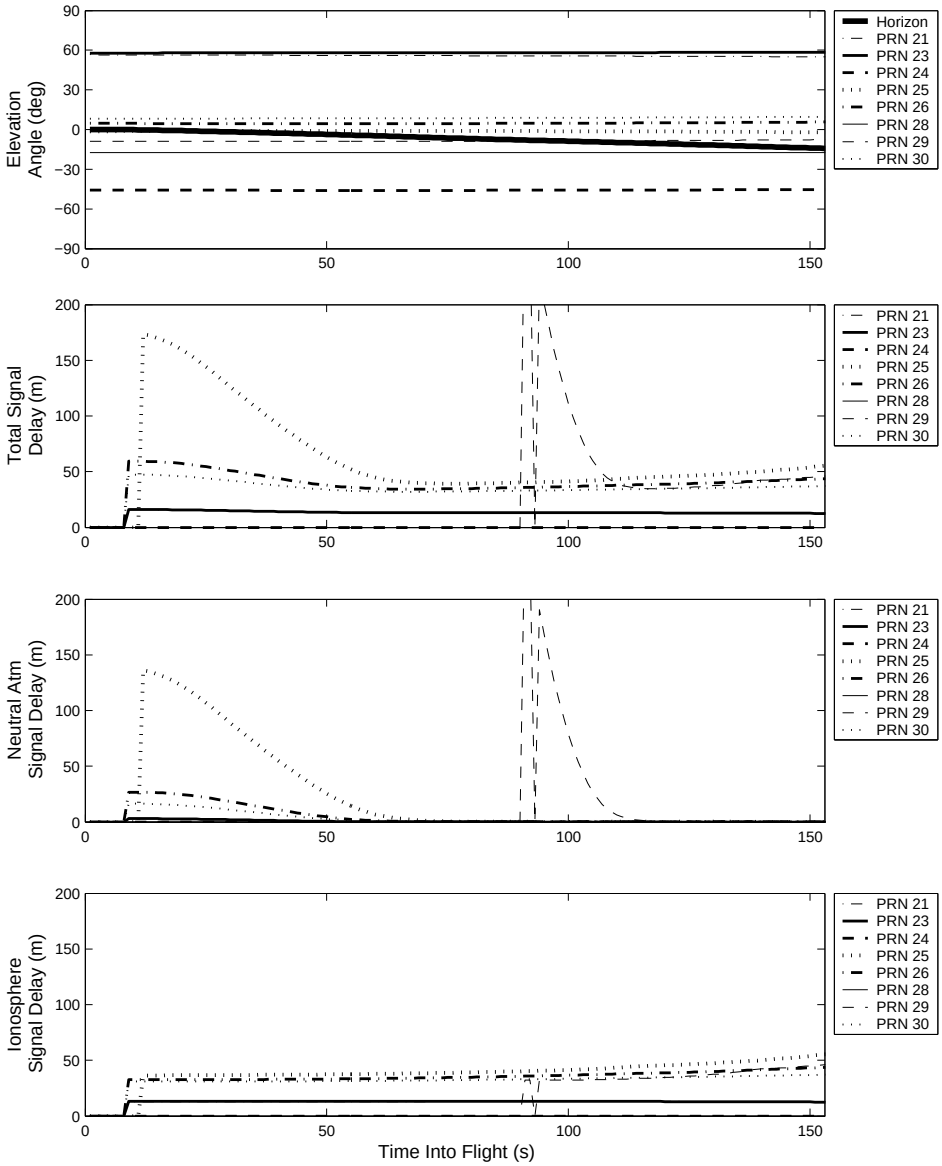

Figure 6-33: The neutral atmosphere, ionosphere and total signal delay experienced by signals from 8 satellites during the launch period of the sample receiver flight. Note that the receiver does not go above local sea level until 10 s into flight, and thus the ray-tracing program did not report any delay.

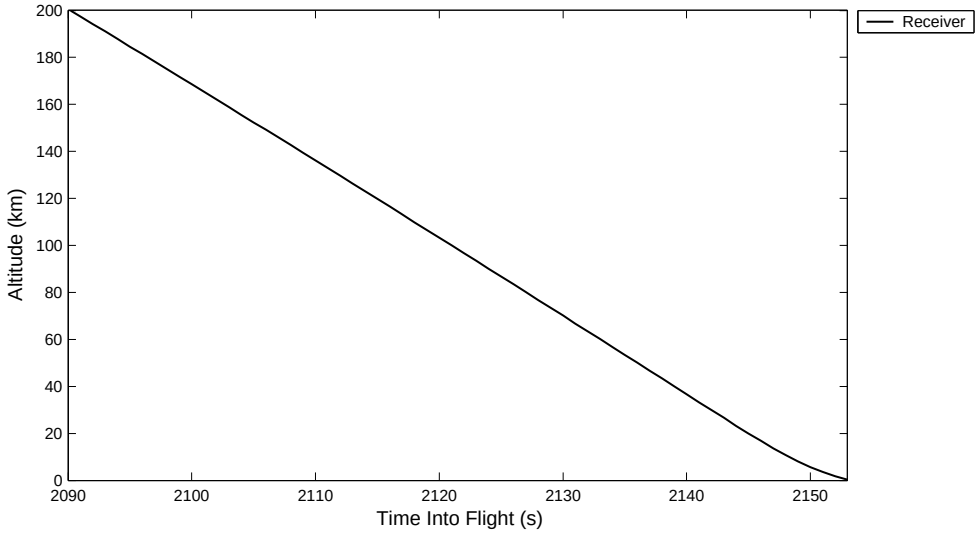

Figure 6-34: The altitude of the receiver during the reentry period of the receiver flight.

While delay rates are quite large as a satellite sets, they are otherwise typically less than 1 m/s. The delay rates are mostly caused by ionosphere signal delay during the period leading up to a satellite setting or right after a satellite rising, as the delay ramps up to its peak value and then decreases until the satellite sets.

6.2.3 Effect of Variables on Signal Delay Features

The delay curves in the previous sections could be predicted mostly on the basis of elevation angle. The magnitude of the peak delay and the time at which the peak delay occurred, however, varied with the local time at the receiver, satellite azimuth and satellite altitude. The magnitude of the peak and the time at which the peak delay occurred were also a function of the conditions that were constant for all satellites over the course of the receiver trajectory: day of year and solar activity level. Figure 6-41 displays the effect of the solar activity level on the signal delay during the trajectory. Not only does the magnitude of the delay drop due to the lower electron densities, but the time at which the peak delay occurs shifts. This shift is equivalent to the change in elevation angle of peak delay that we observed in section 6.1.5. A similar shift will occur due to changes in the other variables (time of day, season, latitude and azimuth) that affected the elevation angle at which the peak delay occurred.

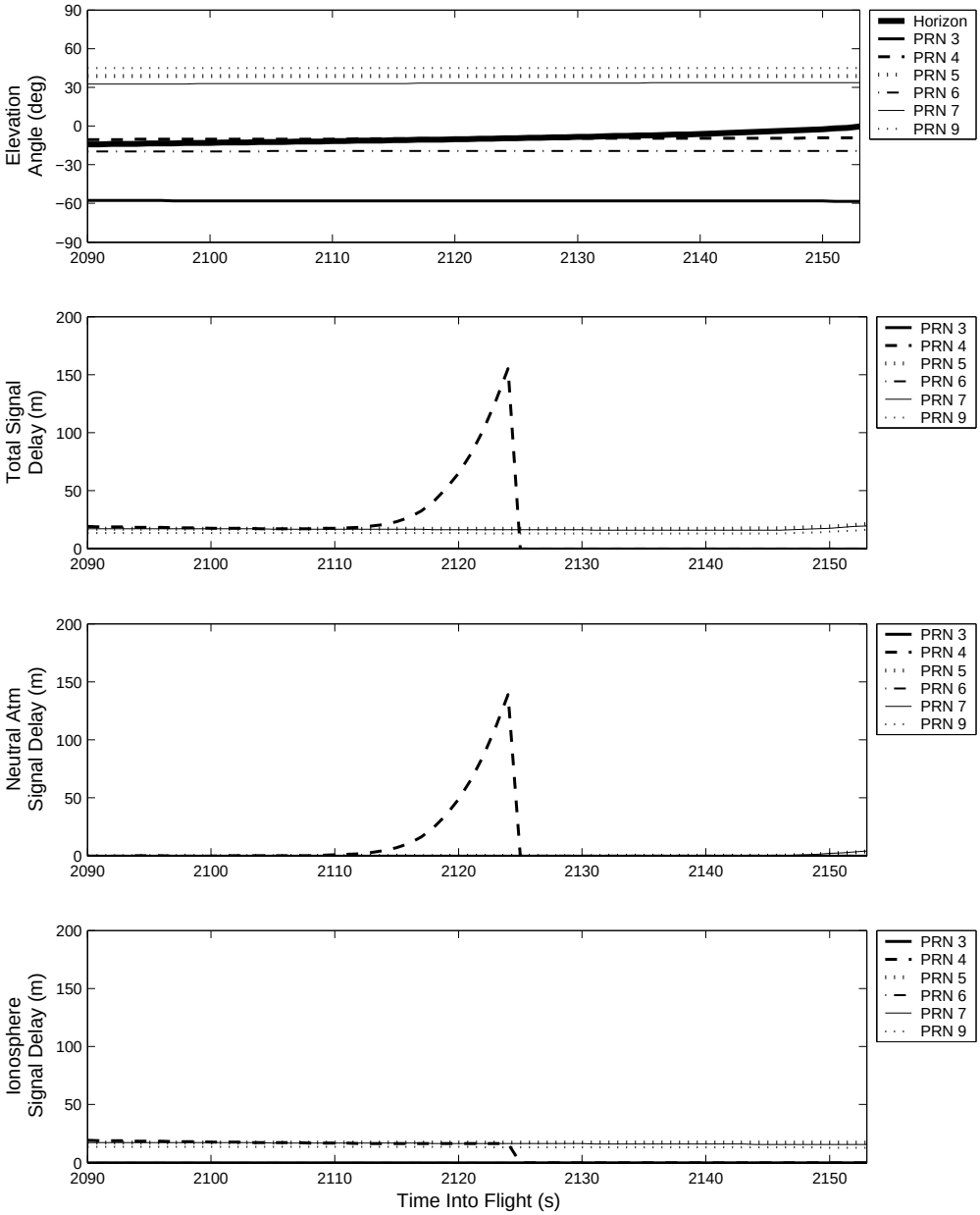

Figure 6-35: The neutral atmosphere, ionosphere and total signal delay experienced by signals from 6 satellites during reentry of the sample receiver flight.

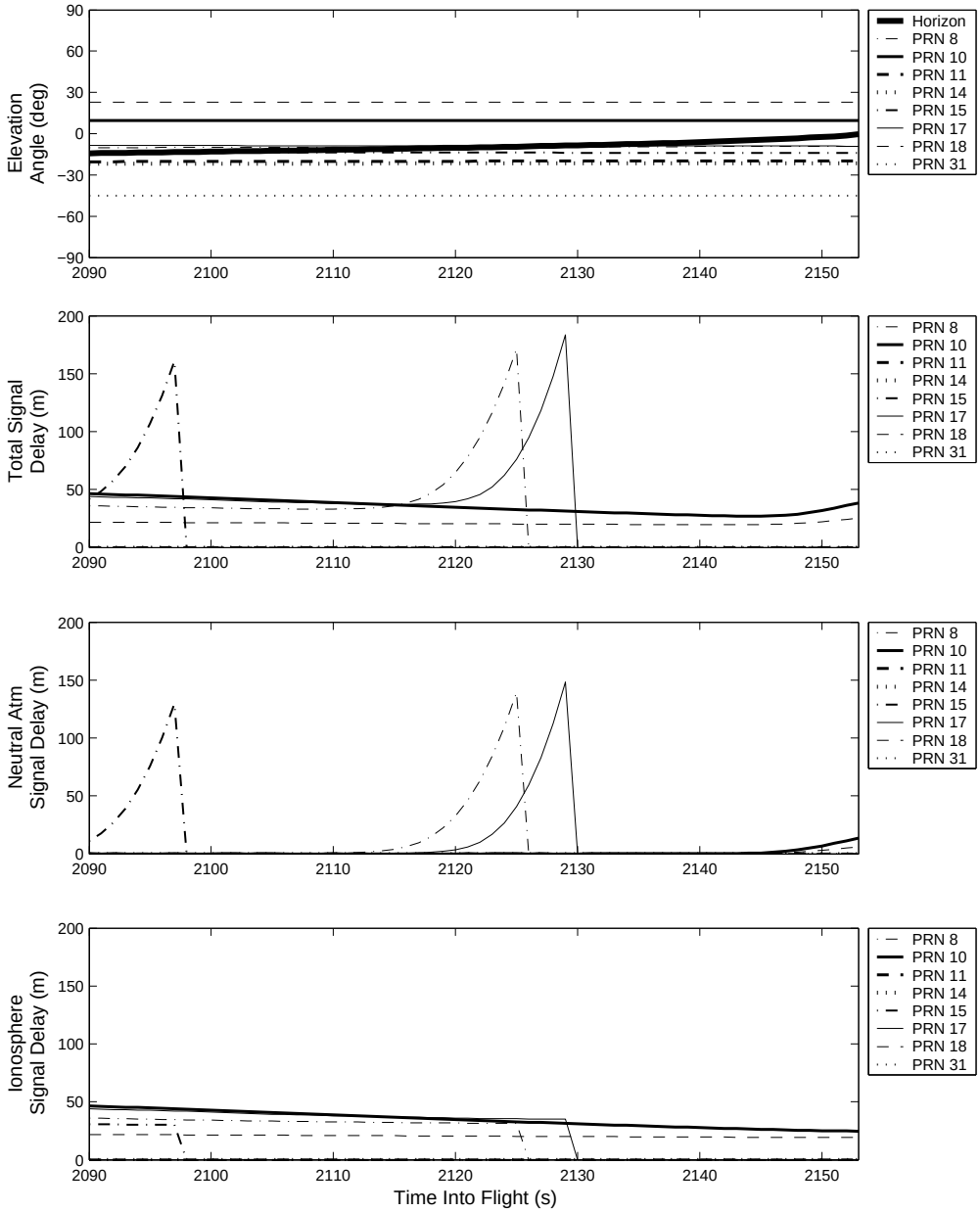

Figure 6-36: The neutral atmosphere, ionosphere and total signal delay experienced by signals from 8 satellites during reentry of the sample receiver flight.

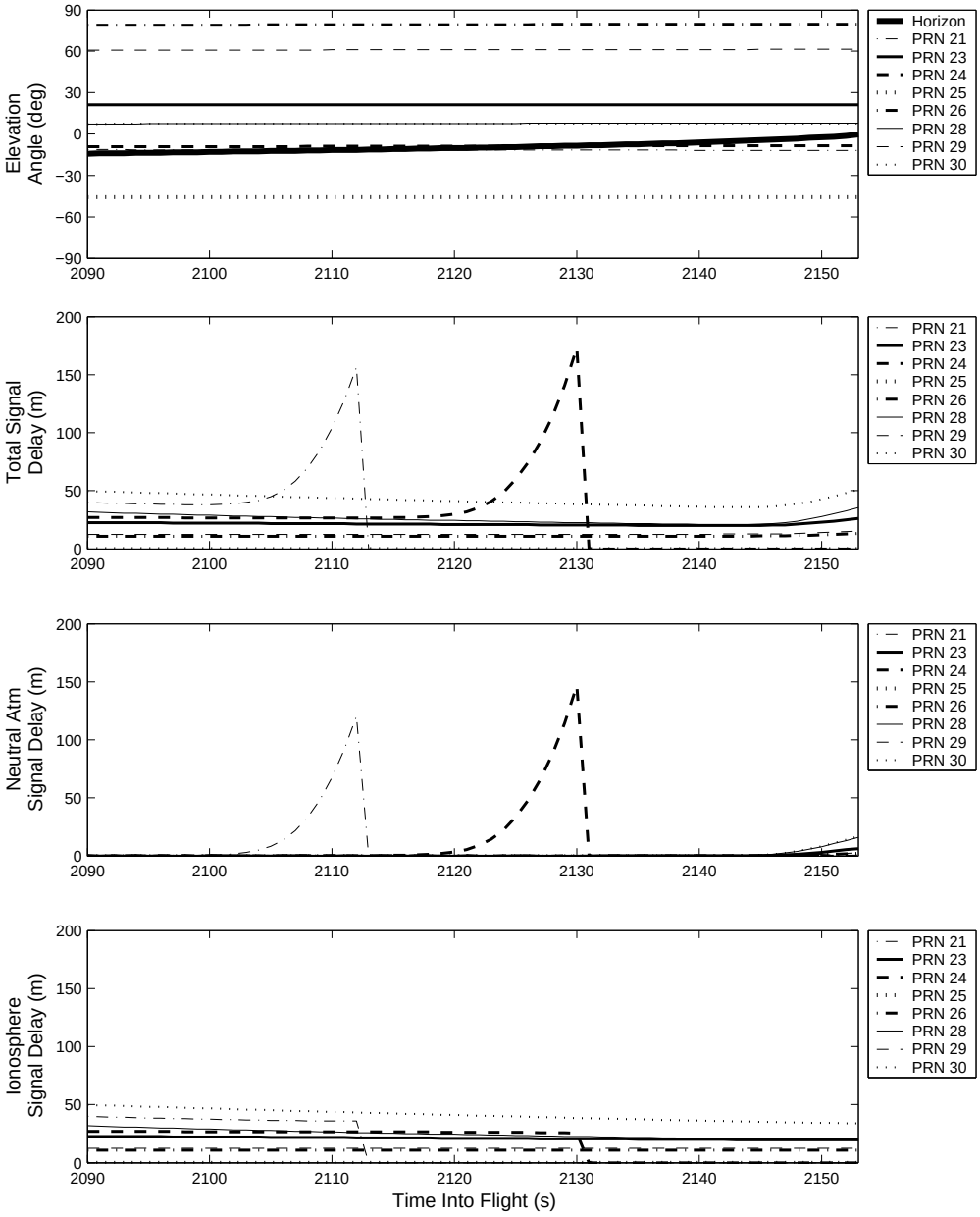

Figure 6-37: The neutral atmosphere, ionosphere and total signal delay experienced by signals from 8 satellites during reentry of the sample receiver flight.

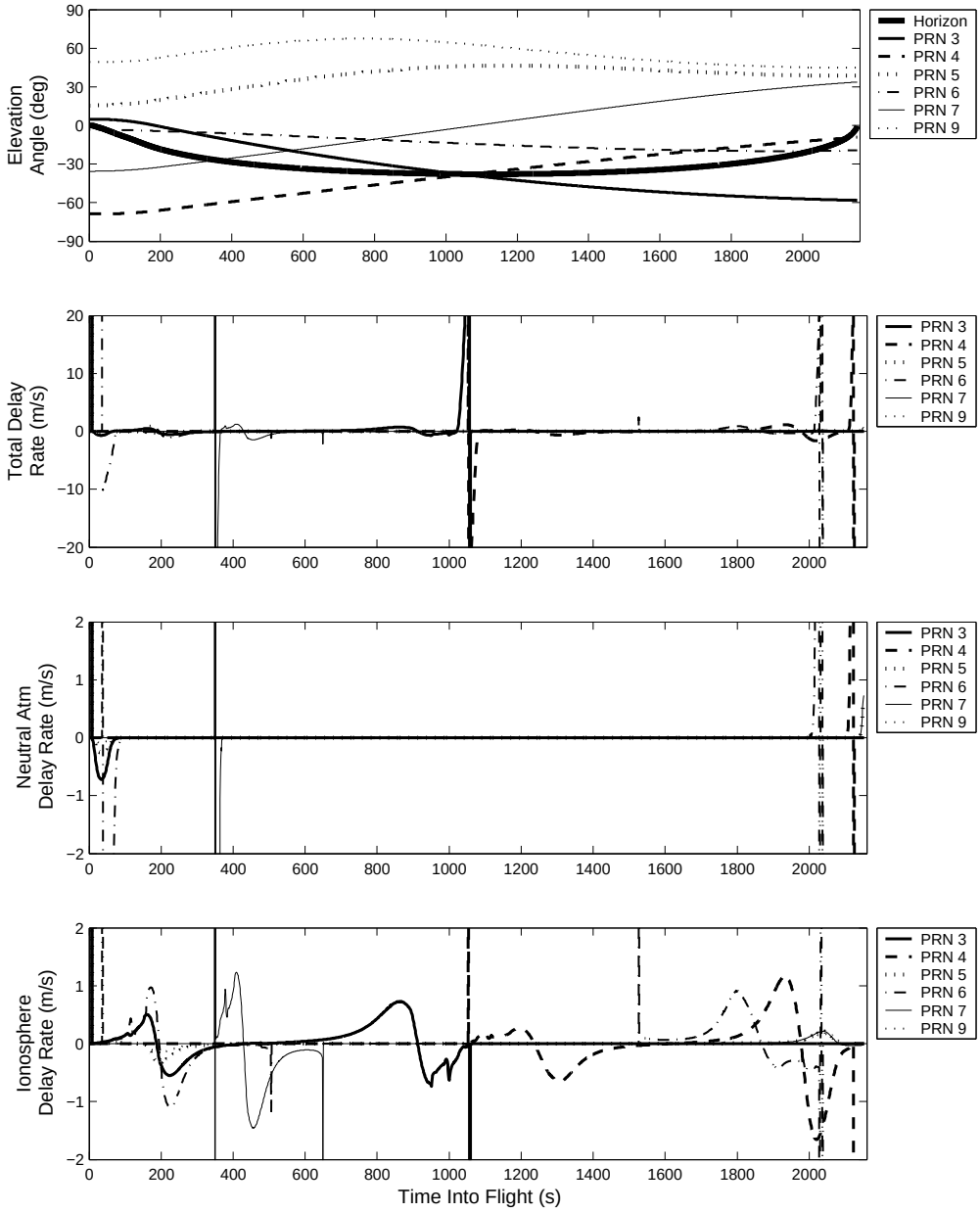

Figure 6-38: Delay rates introduced by uncorrected total, neutral atmosphere and ionosphere signal delay experienced by signals from 6 satellites during the sample receiver flight.

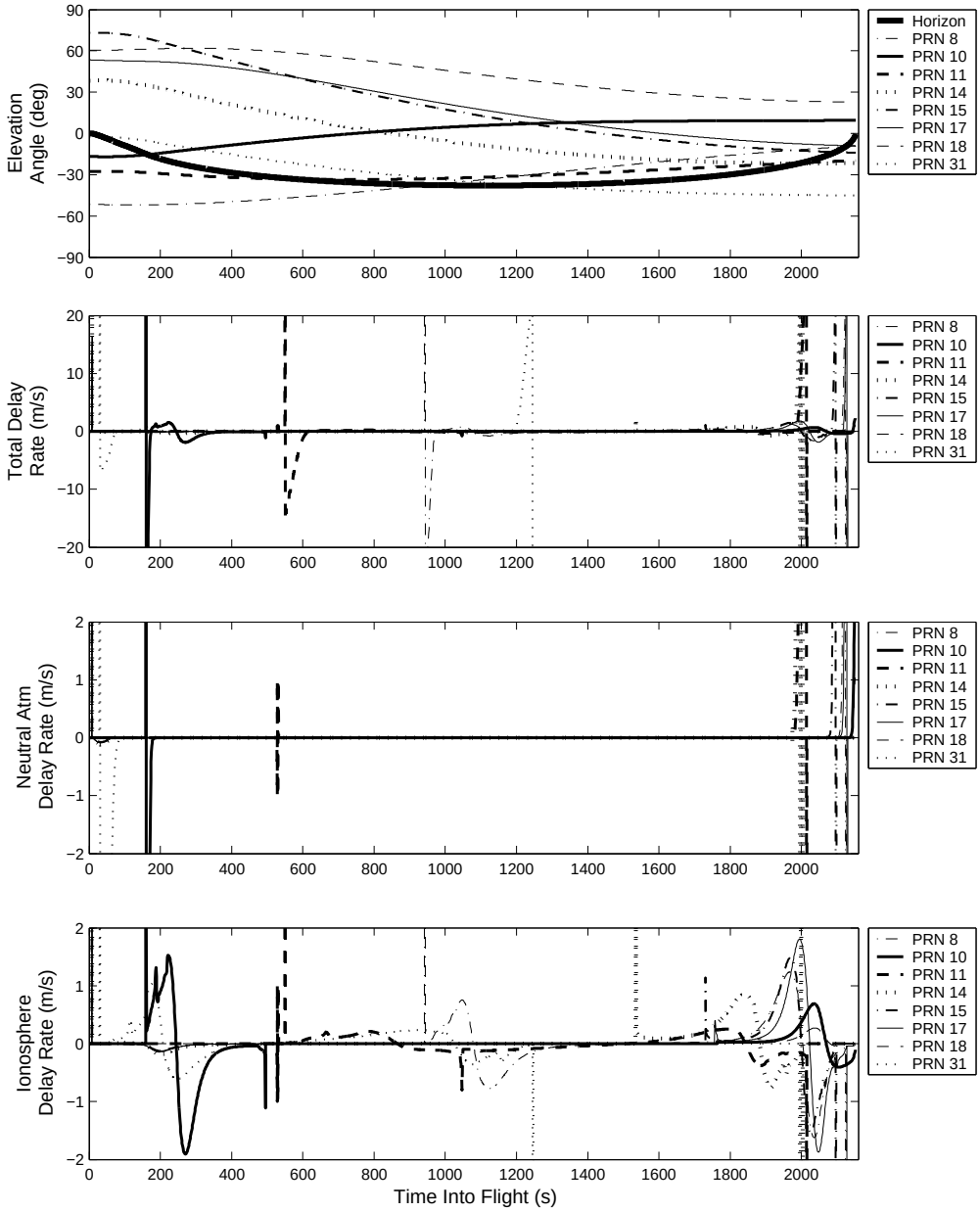

Figure 6-39: Delay rates introduced by uncorrected total, neutral atmosphere and ionosphere signal delay experienced by signals from 8 satellites during the sample receiver flight.

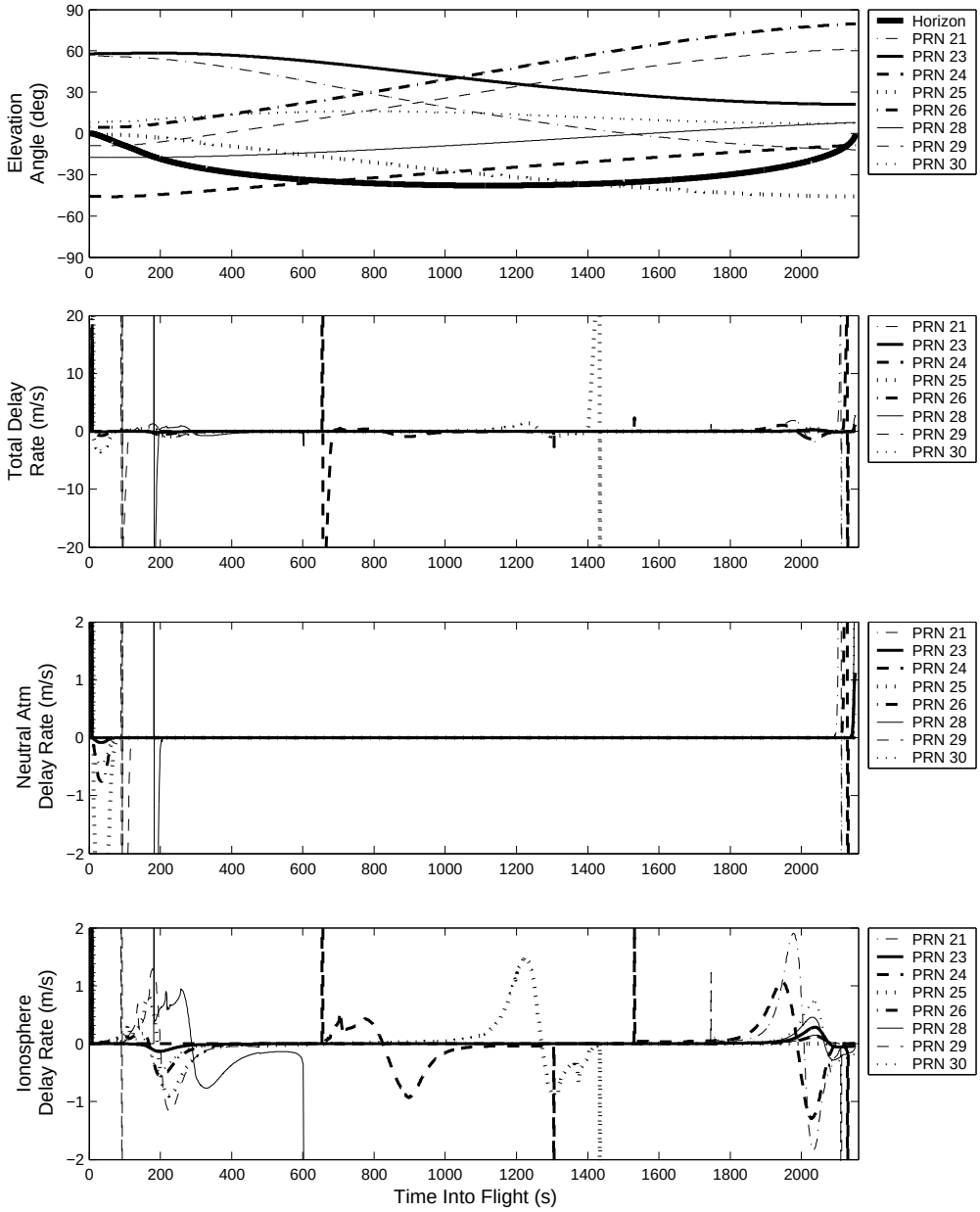

Figure 6-40: Delay rates introduced by uncorrected total, neutral atmosphere and ionosphere signal delay experienced by signals from 8 satellites during the sample receiver flight.

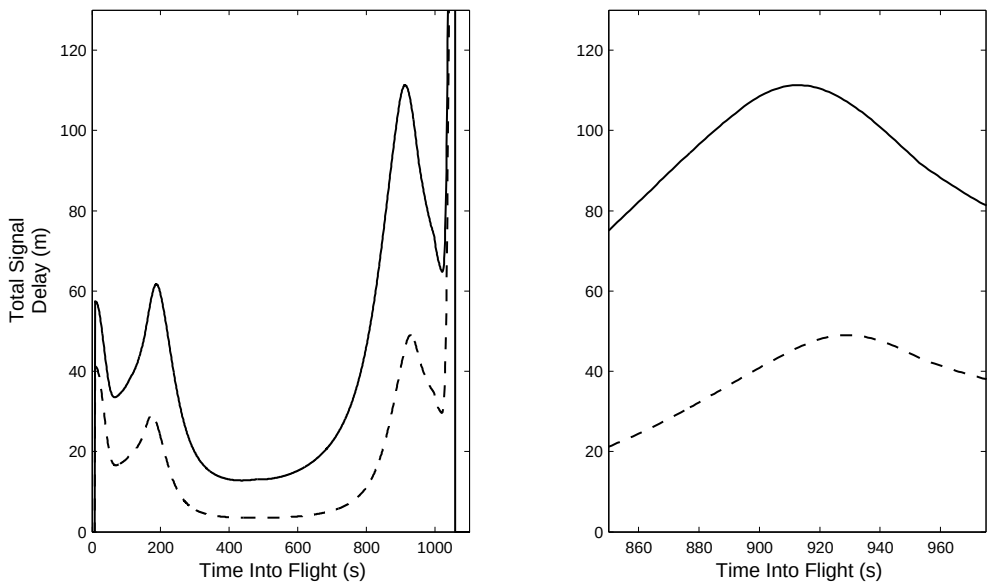

Figure 6-41: Signal delays for the same trajectory flight with the same satellite positions on two different dates: 3/21/80 (high solar activity; solid) and 3/21/78 (low solar activity; dashed). Note that the time at which the peak delay occurs has shifted.

6.3 Error Analysis

There are two types of error in the ray-tracing results. The first is error due to the approximations present in the ray-tracing implementation. The second is prediction error, since these models represent monthly averaged conditions and do not contain the features of a specific day.

Errors Introduced through Ray-tracing Implementation

Error could be introduced four ways in the ray-tracing implementation.

First, the ray-tracing integration is good to $5 \times 10^{-9} \times L$, where L is the distance between the satellite and the receiver. (See section 5.1.3.) This value is at most 0.15 m. This means that each point along the signal path will be at most 0.15 m away from the true signal path. Further, this means that the accumulated travel time of the signal (in meters) (eqn. 5.4), and therefore the calculated total delay, will at most be off by 0.15 m. In practice, I have found that restricting equation 5.4 to be within 0.15 m of the correct value forces the signal path to be much closer than 0.15 m to the true signal path.

Second, error is introduced into the component delay calculations (but not the

total delay calculation) by treating the "other" part of the atmosphere as a vacuum. (See section 5.1.6.) One measure of this error is to compare the sum of the neutral atmosphere and ionosphere delays with the total signal delay. Figure 6-42 shows the difference of these two values. The difference between these two quantities is at the same level as the integration error for low receiver altitudes ($\lessapprox$ 0.15 m), and is most severe right at the horizon. At higher receiver altitudes, we see that the ray-tracing does not accurately model the horizon elevation angles. These results indicate one of two things: that the ionosphere also experiences an increase in signal delay at the horizon (note that the error is only 10–20 m, far less than a typical neutral atmosphere spike), or that the ionosphere refracts a signal at low elevation angles sufficiently that it travels a longer path in the neutral atmosphere than it would have if we considered the ionosphere to be a vacuum. I believe that this second explanation is more likely. We have already noted that at high receiver altitudes the neutral atmosphere signal delay is sensitive to the precise angle of entry into the neutral atmosphere, and therefore it seems reasonable that the small amount of ionosphere refraction could significantly alter the neutral atmosphere signal path. The large error at 1700 km receiver altitude is evidence of this: the neutral atmosphere delay spike was missing in the neutral atmosphere ray-tracing, but evident in the total delay ray-tracing. If it is in fact the case that the neutral atmosphere signal delay is extremely sensitive to the entry angle of the signal into the neutral atmosphere, the aforementioned errors may also fall into the category of "predictive" error. Since we cannot predict the ionosphere conditions well enough to predict the precise entry angle into the neutral atmosphere, the neutral atmosphere delay at the horizon will be unpredictable.

Note that in figure 6-42, all of the errors are positive. This suggests that, as mentioned in section 5.1.5, the neutral atmosphere and ionosphere delays are lower bounds on their true values.

Third, the multidimensional root-finder used in the trajectory ray-tracing implementation requires that the signal strike to within 100 m of the true satellite, but not to strike the satellite precisely. Geometrically, this only moves the path 3×10^{-4} degrees. If the neutral atmosphere delay is *extremely* sensitive to entry angle this could affect the neutral atmosphere delay, although I find it to be highly unlikely. Note that since I only used the root-finder in the trajectory implementation, this only affects the delays in section 6.2, and not the profile delays in section 6.1.

Fourth, the ray-tracing considers the neutral atmosphere to only contribute to the refractivity between 0 and 65 km, and the ionosphere to only contribute to the

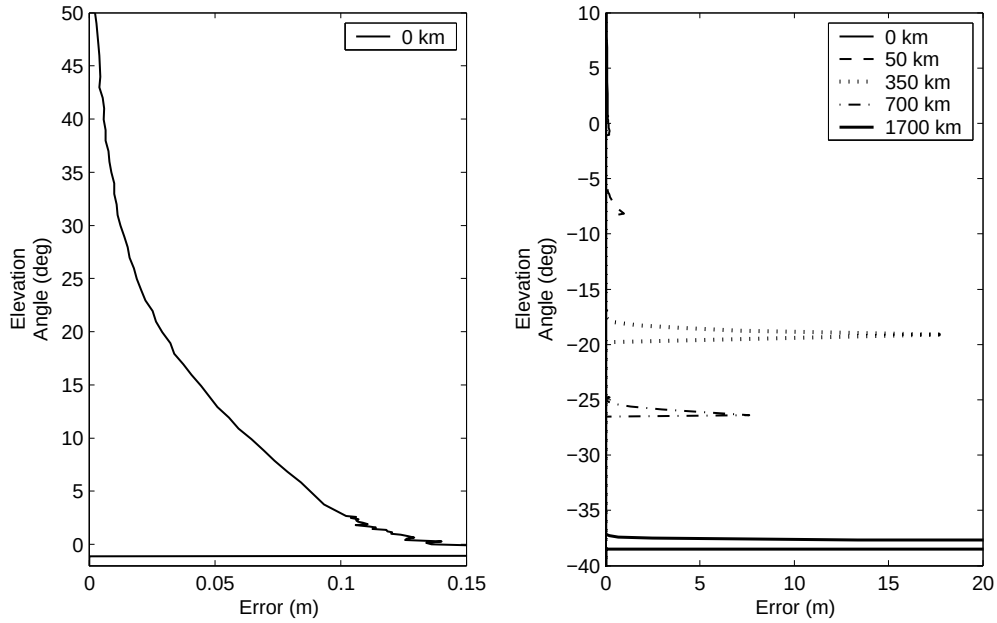

Figure 6-42: Total delay minus the sum of neutral atmosphere delay and ionosphere delay: one measure of error in delay calculations due to ray-tracing implementation.

refractivity between 65 and 1050 km. These hard boundaries obviously do not exist in real life, and will introduce some error. Limiting the neutral atmosphere to be below and the ionosphere to be above 65 km should add negligible error to the neutral atmosphere and ionosphere delay calculations; the error from limiting the ionosphere to be below 1050 km is less clear since in figure 2-4, none of the electron density profiles go to zero at 1050 km. The error will be most significant at the higher latitudes.

Prediction Error

Prediction error is the difference between real delays measured on a given day and the delays that ray-tracing would have predicted if those conditions were input into the IRI-2001 and GRAM-99 models.

The ionosphere is highly unpredictable. Not only do electron densities deviate by 25% from the monthly mean at the 1 sigma level, local phenomena such as traveling ionosphere disturbances and scintillation cause local electron densities to vary drastically. As a result, the ray-traced ionosphere delays are only accurate to within approximately 25% of the measured ionosphere delay on a given day with the same general conditions. When looking at the delays over the course of a receiver trajec-

tory, this prediction error could manifest itself as an overall magnitude change or as local magnitude changes, which would appear as bumps or noise in the otherwise smooth curves.

The neutral atmosphere is much more predictable than the ionosphere — note how in section 6.1, it was nearly impossible to discern if changing the neutral atmosphere conditions changed the neutral atmosphere delays. I am mostly concerned with how varying the GRAM-99 water vapor distribution model would affect signal refraction near the surface of the Earth. Local water vapor distributions can significantly affect refraction near the surface of the Earth, especially over oceans. The GRAM-99 water vapor distribution will introduce a prediction error, but it has not been quantified.

If refraction in the neutral atmosphere is sensitive to entry angle as suggested above, the unpredictability of the ionosphere will prevent us from accurately predicting neutral atmosphere delays near the horizon.

6.4 Comparison of Ray-Traced Delay to Receiver Compensation Methods

Figures 6-43 and 6-44 show how the receiver-based neutral atmosphere and ionosphere delay models discussed in sections 3.1.2 and 3.1.1 perform at space vehicle altitudes.

As displayed in figure 6-43, the Altushuler and NATO models perform relatively well at altitudes where there is still significant structure to the neutral atmosphere delay — until the signal goes below 0° elevation angle. These two models will not predict the delay for a signal from a satellite at most negative elevation angles.

As displayed in figure 6-44, the Klobuchar model performs very poorly at space vehicle altitudes, particularly at the lower elevation angles. The Klobuchar model has a singularity at $-19.8°$ elevation angle and does not predict the elevation angle of the horizon.

6.4.1 Model Performance on Reentry

A space vehicle is most likely to rely on satellites at low elevation angles when reentering the atmosphere due to the geometries involved. The spacecraft will block nearly half the sky and the Earth will block a quarter of the sky. Signals from some of the visible satellites will have to pass through the atmosphere twice (see figure 1-3 for an exaggerated representation of this geometry), but since the spacecraft is relatively close to the surface of the Earth, most satellites will be at positive elevation angles.

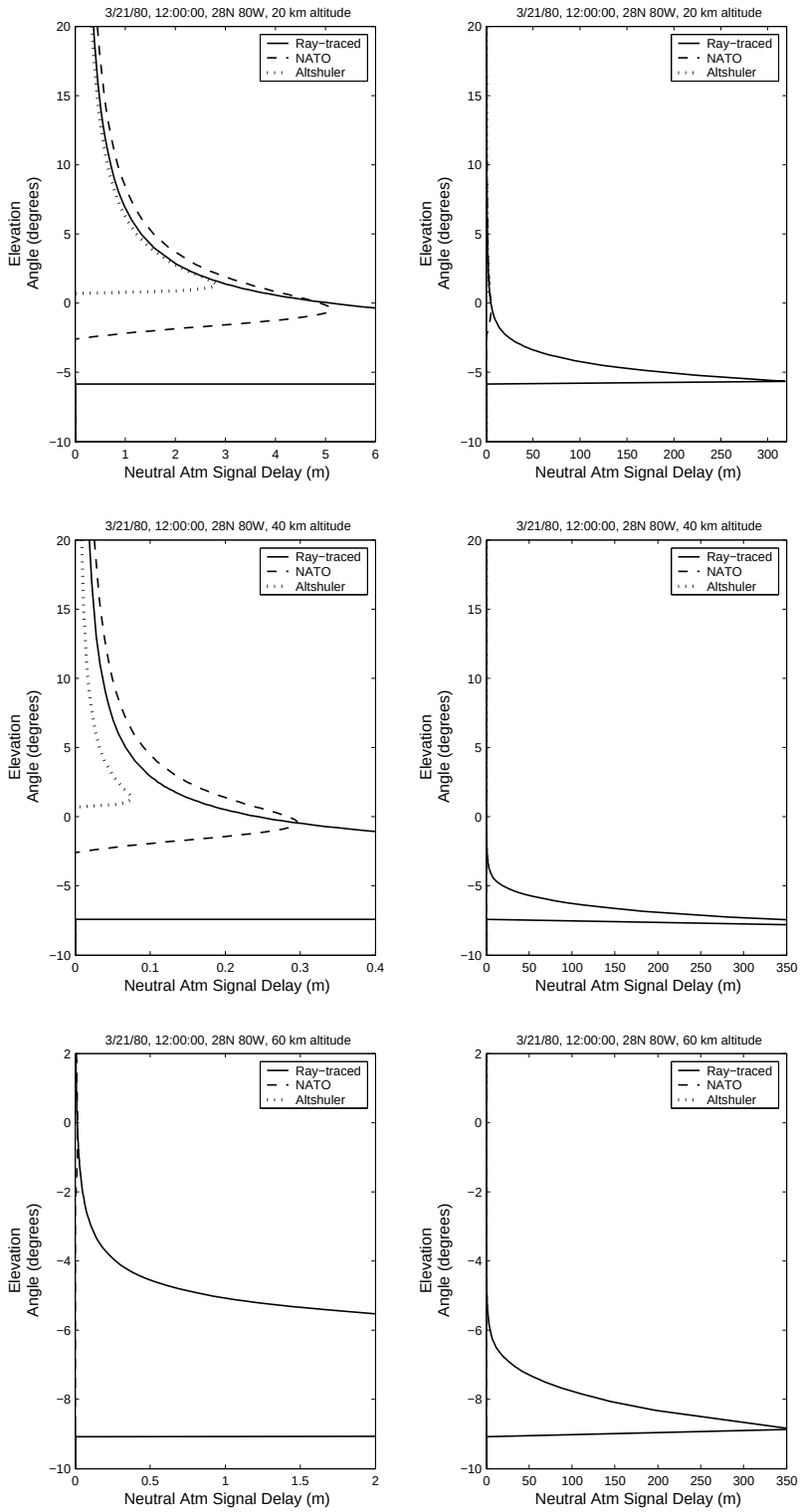

Figure 6-43: Performance of the Altshuler and NATO delay models at 20, 40 and 60 km receiver altitude compared to neutral atmosphere ray-tracing results. Note that the Altshuler and NATO models do not cover negative elevation angles, and so predict no delay at 60 km receiver altitude.

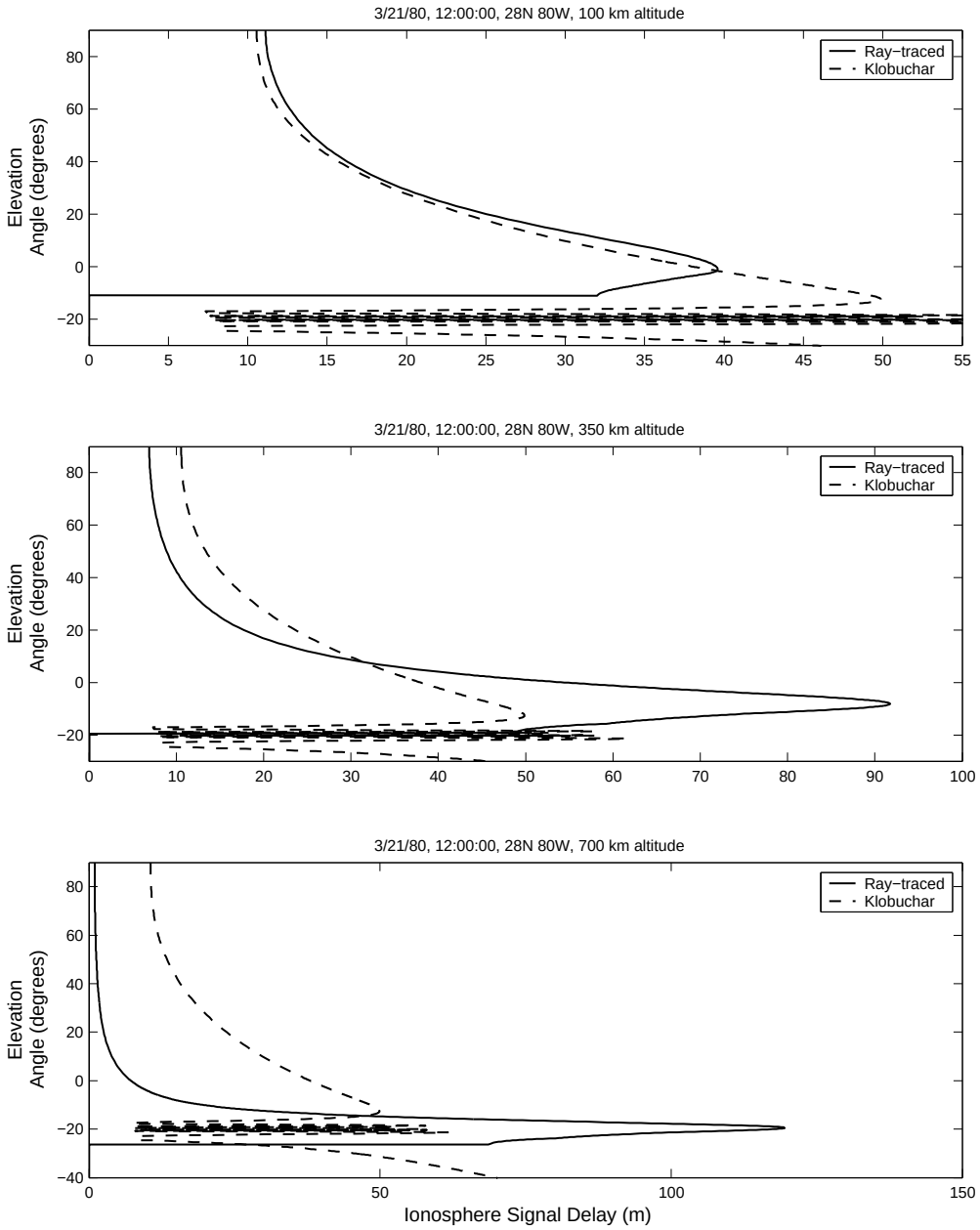

Figure 6-44: Performance of the Klobuchar delay model at 100, 350 and 700 km altitude compared to ionosphere ray-tracing results. Note that the Klobuchar algorithm did not determine at what elevation angle the signal would no longer be visible.

During reentry, all signals will pass through part of the ionosphere at least once. Since the spacecraft is relatively close to the surface of the Earth and few satellites will be at extremely negative elevation angles, it is conceivable that the onboard delay models might perform adequately during reentry.

Figures 6-45, 6-46 and 6-47 show how well the NATO neutral atmosphere delay model performs during reentry. These figures show the neutral atmosphere delay as the spacecraft descends from 200 km altitude at a roughly constant rate. The NATO model performs poorly for satellites that set during reentry since these satellites set while at negative elevation angles, which are not covered well by the model. Once the spacecraft enters the neutral atmosphere, the NATO model begins to report delays on signals from satellites at positive (or very small negative) elevation angles. During this period, when the spacecraft is within the neutral atmosphere, the model is off by at most 0.75 m for visible satellites. (The NATO model continues to predict delays for satellites after they have set. I have removed from figures 6-45, 6-46 and 6-47 the satellites which have set before the reentry period, but for satellites such as PRN 4, PRN 17 and PRN 24 (and others) that set during reentry, these figures report NATO delay even after the satellite has set. These satellites therefore have a large error in the bottom tile, even though there should be no error since the satellite has set.)

Figures 6-48, 6-49 and 6-50 show how well the Klobuchar ionosphere delay model performs during reentry. Note that the error is large for satellites that set during flight since they are at negative elevation angles, which the Klobuchar model is not designed for. The Klobuchar model will deviate from the ray-traced results by anywhere from 10–15 m for these satellites. The Klobuchar model also performs poorly for satellites such as PRN 10, which skirt the horizon for much of reentry. The model is off by as much as 20 m in these circumstances. Otherwise, errors typically remain below 10 m, and are often below 5 m.

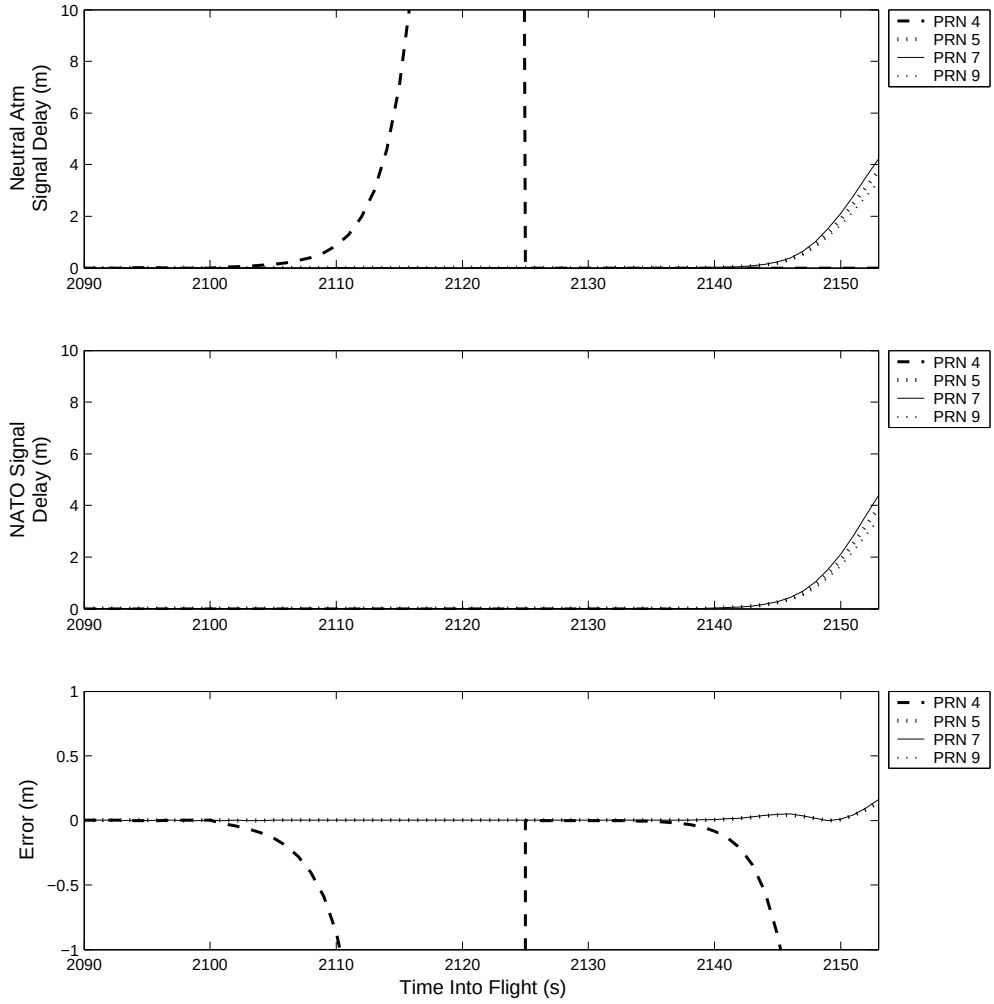

Figure 6-45: A comparison of ray-traced neutral atmosphere signal delays with delays predicted by the NATO model during reentry for 6 satellites. Note that the NATO model continues to predict delays even once a satellite has set. I have removed the NATO predictions for satellites that are not visible at any point during the reentry period.

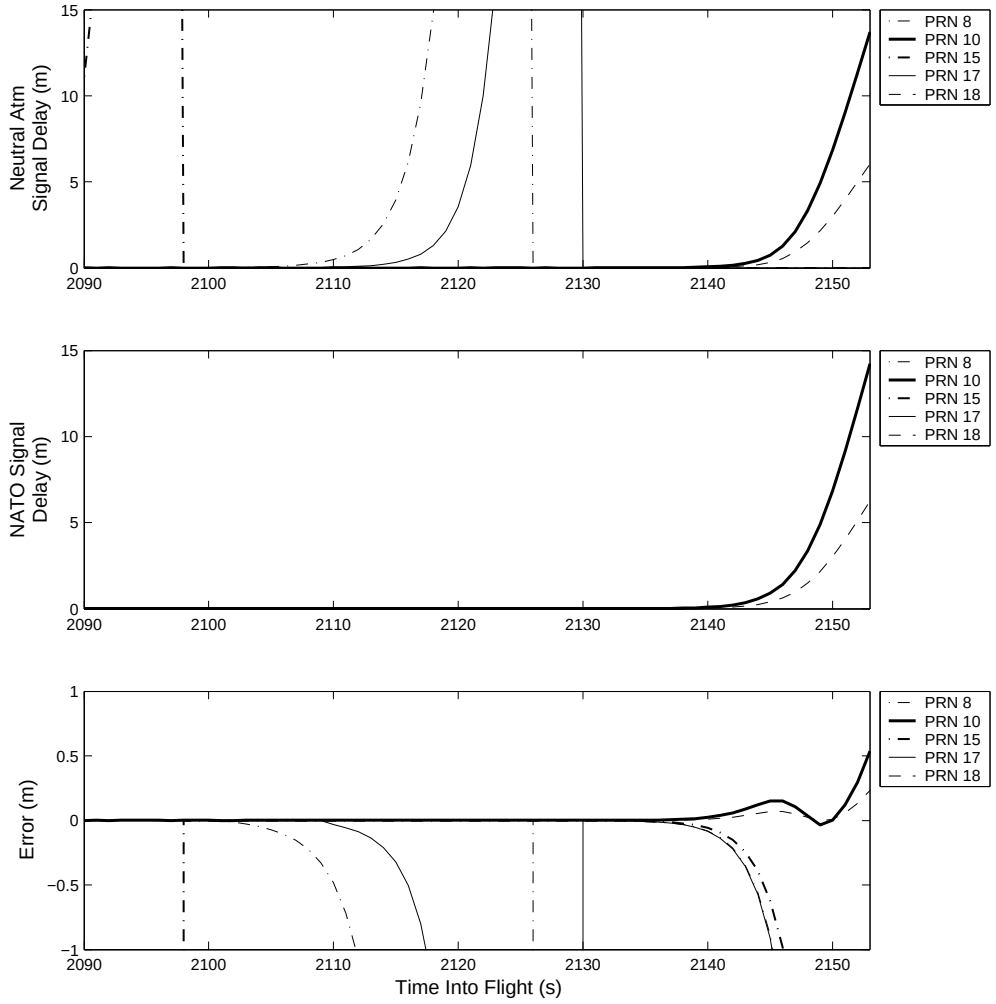

Figure 6-46: A comparison of ray-traced neutral atmosphere signal delays with delays predicted by the NATO model during reentry for 8 satellites. Note that the NATO model continues to predict delays even once a satellite has set. I have removed the NATO predictions for satellites that are not visible at any point during the reentry period.

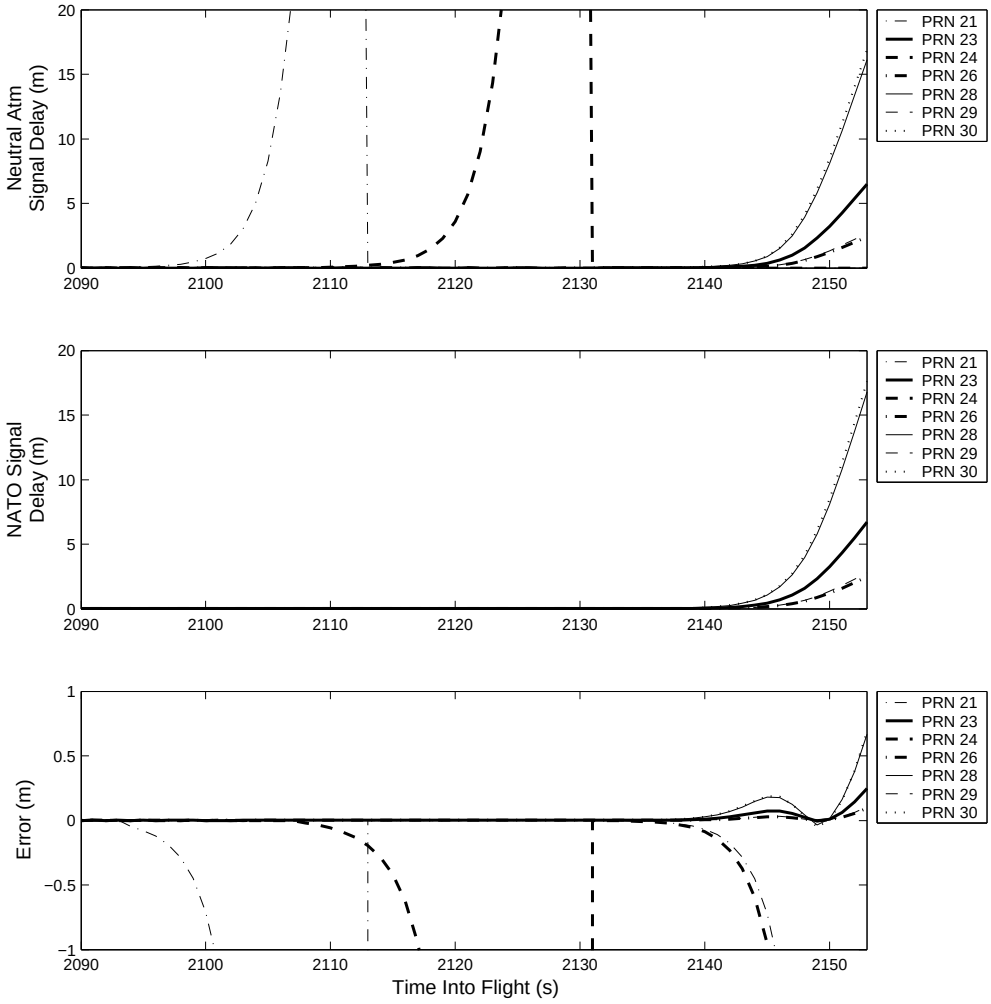

Figure 6-47: A comparison of ray-traced neutral atmosphere signal delays with delays predicted by the NATO model during reentry for 8 satellites. Note that the NATO model continues to predict delays even once a satellite has set. I have removed the NATO predictions for satellites that are not visible at any point during the reentry period.

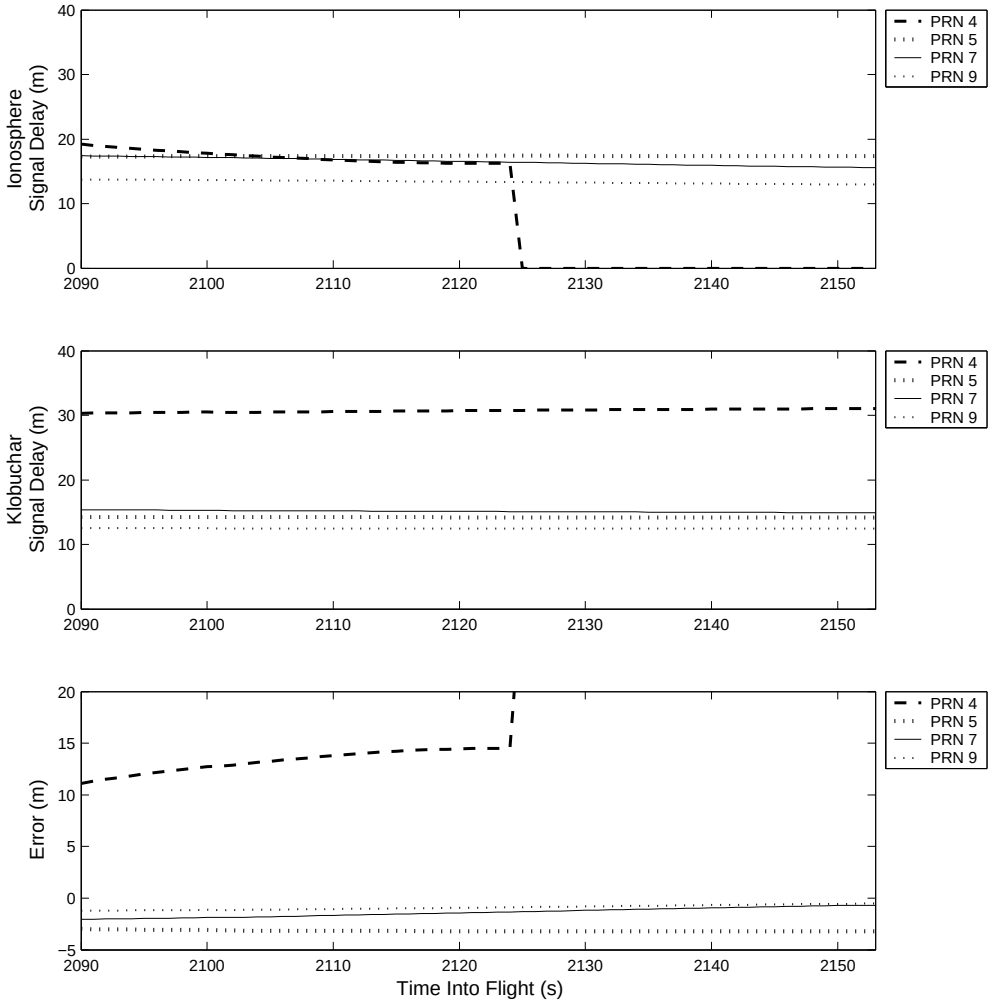

Figure 6-48: A comparison of ray-traced ionosphere signal delays with the Klobuchar delay prediction during reentry for 6 satellites. Note that the Klobuchar model continues to predict delay even once a satellite has set. I have removed the Klobuchar predictions for satellites that are not visible at any point during the reentry period.

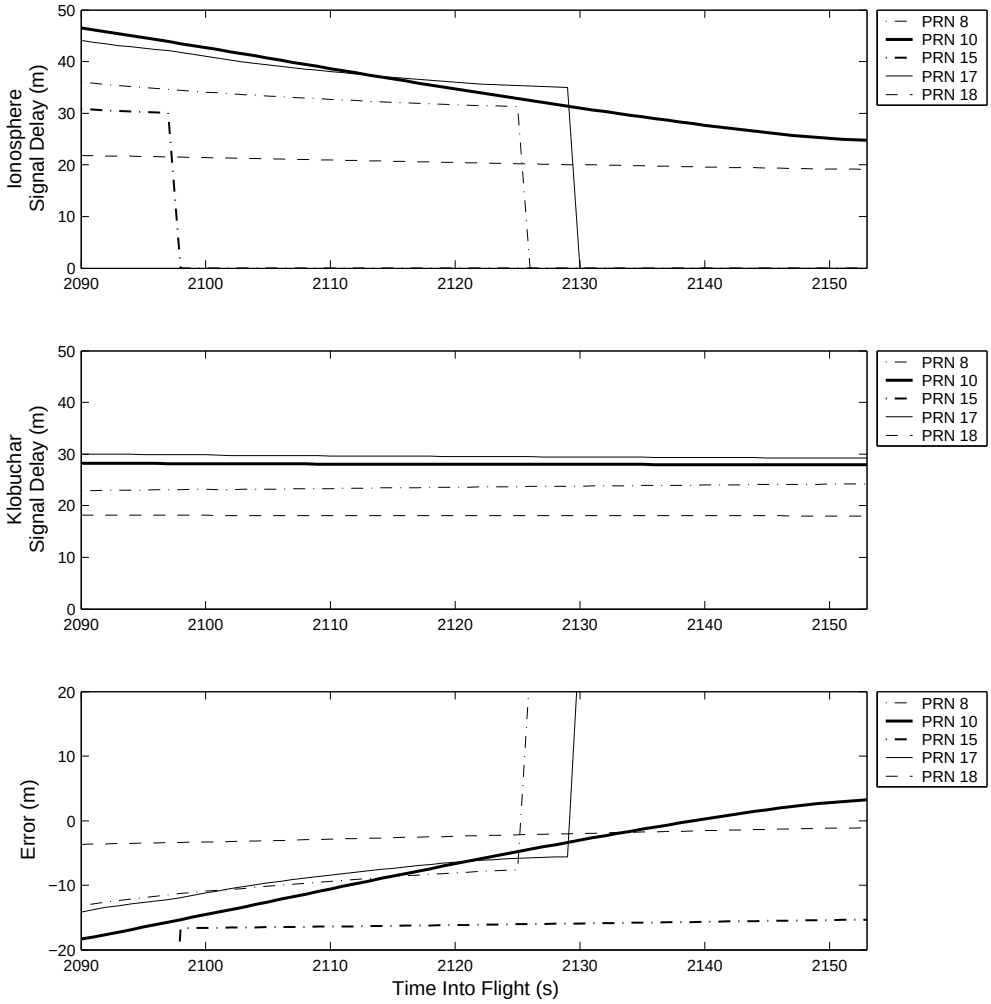

Figure 6-49: A comparison of ray-traced ionosphere signal delays with the Klobuchar delay prediction during reentry for 8 satellites. Note that the Klobuchar model continues to predict delay even once a satellite has set. I have removed the Klobuchar predictions for satellites that are not visible at any point during the reentry period.

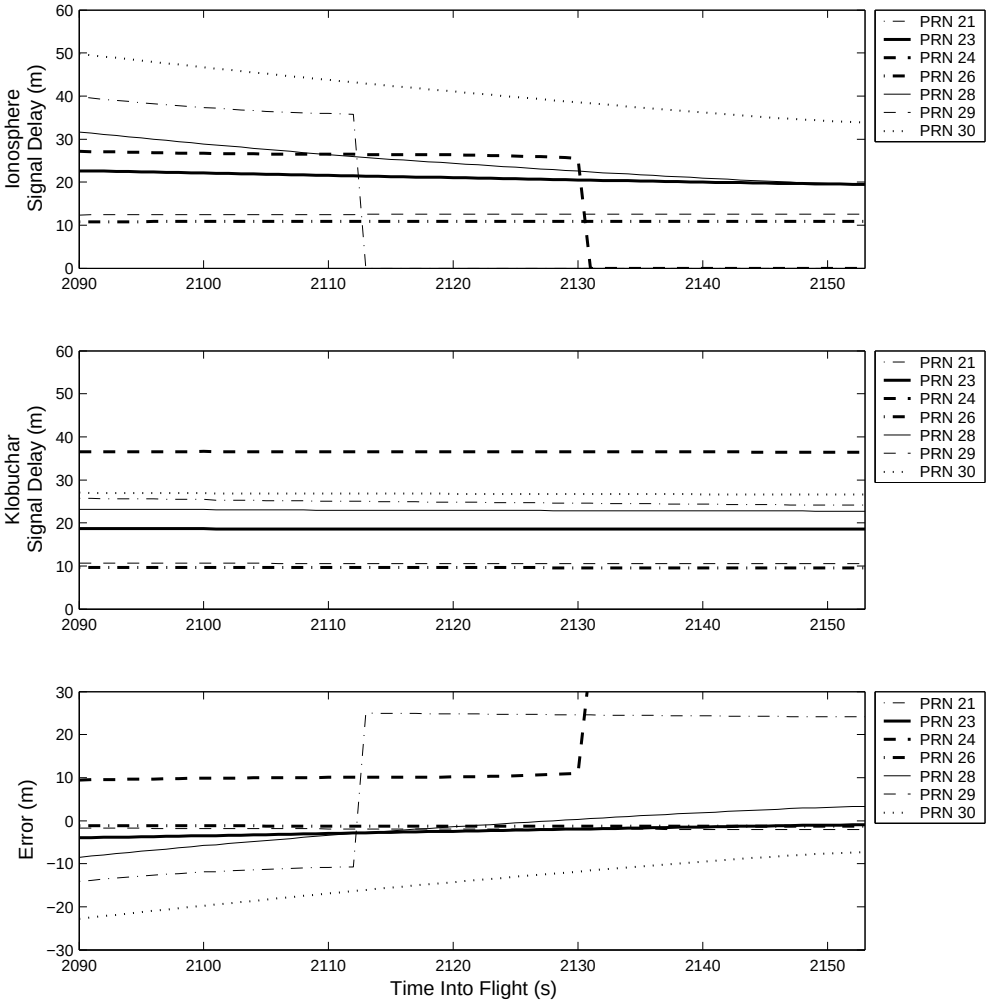

Figure 6-50: A comparison of ray-traced ionosphere signal delays with the Klobuchar delay prediction during reentry for 8 satellites. Note that the Klobuchar model continues to predict delay even once a satellite has set. I have removed the Klobuchar predictions for satellites that are not visible at any point during the reentry period.

Chapter 7

Correcting Signal Delay Aboard a Space Vehicle

In this chapter I discuss how one might use the results of the previous chapter to better correct for the neutral atmosphere and ionosphere signal delay aboard a space vehicle.

7.1 Correcting the Neutral Atmosphere Delay

We saw in section 6.4 that the neutral atmosphere signal delay models discussed in section 3.1.2 are neither valid at most space vehicles altitudes nor valid for signals from satellites at negative elevation angles. These models do perform relatively well during reentry, when the space vehicle is at low altitudes and observing satellites at positive (or only slightly negative) elevation angles, however. We are therefore left with three options: accept the existing neutral atmosphere delay models, use them during launch and reentry, and do not observe signals that pass through the neutral atmosphere at other times; develop a new neutral atmosphere signal delay model; or assume a large error in the range calculation when neutral atmosphere signal delay is possible. Since the neutral atmosphere is not dispersive, dual frequency correction is not an option.

7.1.1 Accept Existing Models

If we accept the existing neutral atmosphere delay models, the receiver will only be able to correct for neutral atmosphere delay when the space vehicle is below ~ 50 km altitude. Even then, the receiver will not be able to correct for the delay on signals

from satellites as they set during reentry, since the setting satellites will be at negative elevation angles.

To continue to use the existing models, the receiver must be programmed with the elevation angle for a given receiver altitude at which to no longer track GPS satellites. I fit a curve to the points of 0.01 m neutral atmosphere delay from 50 km to 1050 km altitude from the ray-tracing control profile results. (I used 0.01 m instead of 0 m for computational reasons.) I found the best fit to these points to be:

$$E(h) = -13.6 \times h^{0.217} + 31.5 \tag{7.1}$$

where E is elevation angle, in degrees, and h is altitude, in kilometers. This curve is plotted on top of the ray-traced neutral atmosphere delay contour plot in figure 7-1. This curve does not perfectly fit the data points, particularly in the middle altitudes, but is always above the 1 m delay line. To be conservative, add a constant offset of 0.5° to this equation to be certain there is no chance of neutral atmosphere delay.

With this approach, the neutral atmosphere delay correction should be correct to within the errors of the NATO or Altshuler model during launch and reentry. We saw in section 6.4 that this error will be at worst ∼1 m.

7.1.2 Develop New Model

Two choices must be made when developing a neutral atmosphere signal delay model: the form of the model and the maximum altitude for which the model is valid.

Form

Using a mapping function to describe the neutral atmosphere signal delay will be more complex for receivers at altitude than for receivers on the ground. First, the zenith delay does not contain contributions from all of the altitudes that the signal passes through. (In figure 7-2, the signal passes through altitudes that are below the altitude of the receiver.) The mapping function would have to calculate the zenith delay above the altitude of closest approach, or compensate in some other manner. Further, the mapping function will have take into account that the signal passes through the altitudes which are below the receiver twice. (See figure 1-2; in figure 7-2, the signal passes through certain altitudes in the lightly shaded region, below the receiver, twice.) The geometries involved will change with altitude, making a mapping function quite complex. A more explicit fit to the delay for all altitudes

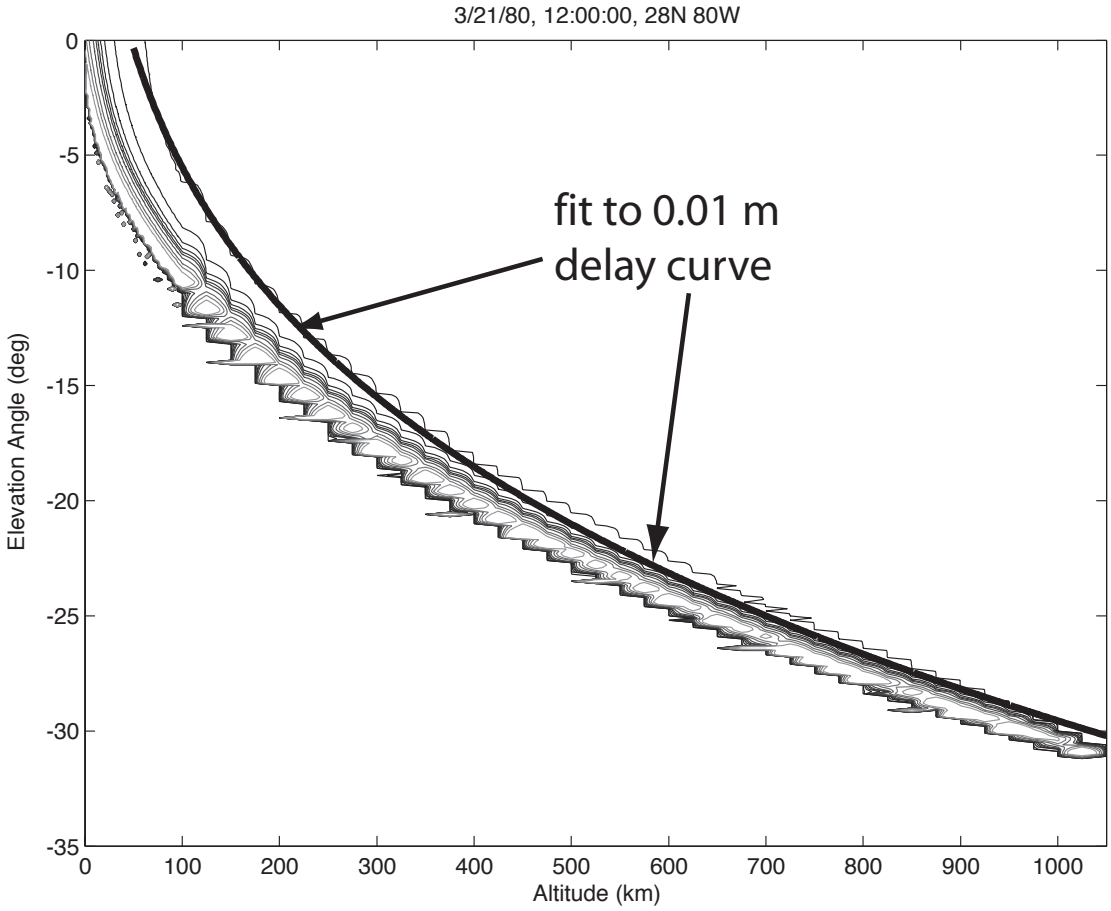

Figure 7-1: Ray-traced neutral atmosphere signal delay. Contour lines at 0.01, 1, 5, 10, 15, 20, 30, 50, 70, 100 m delay. Equation 7.1 is plotted over the points of 0.01 m delay.

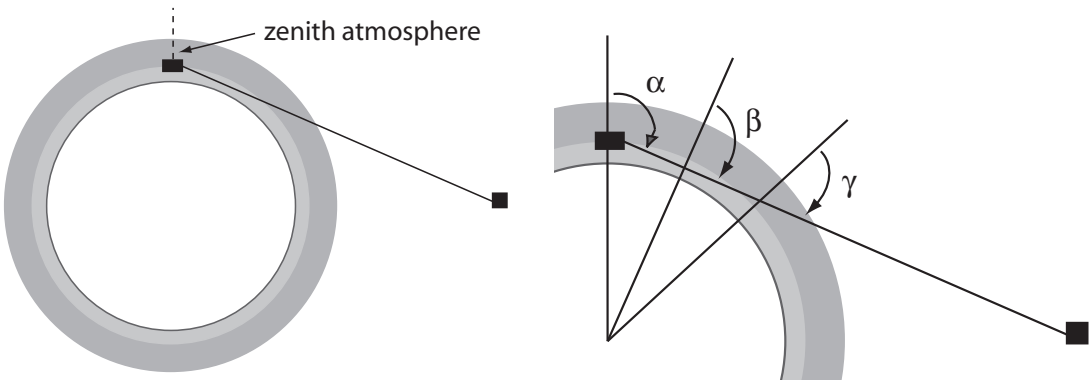

Figure 7-2: Signals to a spacecraft that is within the atmosphere will pass through some layers of the atmosphere once and others twice. The angle at which a signal passes through each layer is different.

and elevation angles (instead of to the zenith delay and a separate mapping function) may be best. A second option would be to place the ray-tracing results in tables and interpolate to find the desired values.

We observed in section 6.1 that neutral atmosphere delays did not vary much as a function of any of the conditions examined, although we expect significant variations due to water vapor in the troposphere. The largest difference in ray-traced neutral atmosphere delays occurred by varying the latitude. The observed differences in delays at three altitudes are outlined in tables 7.1, 7.2 and 7.3. In general, the delay at 28°N was roughly the average delay, and the delays at 75°N and 0°N had the minimum and maximum observed delay at a given elevation angle. (This did not hold true at all altitudes and elevation angles.) Based on the results in these tables, if we construct a model based on the neutral atmosphere delays provided by the control profiles (figure 6-7; neutral atmosphere delay experienced by signals from a satellite due east of a receiver at 28°N 80°W at noon on 3/21/80), the calculated neutral atmosphere delay will be roughly within 7% of the correct ray-traced value. (For signals that pass through the troposphere traveling between a real satellite and a real receiver, there may be a significantly larger difference between the calculated neutral atmosphere delay and the actual delay.) A more detailed Monte Carlo statistical analysis using the perturbation capabilities of the GRAM-99 code and alternate water vapor distributions is warranted.

EA (deg)	28°N (control) delay (m)	75°N delay (m)	Δ(%)	0°N delay (m)	Δ(%)
45	3.51	3.28	−6.6	3.71	5.7
30	4.95	4.63	−6.5	5.25	6.1
15	9.44	8.84	−6.4	9.99	5.8
5	25.29	23.78	−6.0	26.81	6.0
1	60.95	58.63	−3.8	64.75	6.2

Table 7.1: Variations in neutral atmosphere signal delay from that observed at 28°N as a function of latitude for 0 km receiver altitude. EA stands for geometrical elevation angle.

EA (deg)	28°N (control) delay (m)	75°N delay (m)	Δ(%)	0°N delay (m)	Δ(%)
−5	4.03	3.78	−6.2	4.03	0.0
−6.5	47.32	40.86	−13.7	48.65	2.8
−7.5	169.57	166.13	−2.0	172.32	1.6

Table 7.2: Variations in neutral atmosphere signal delay from that observed at 28°N as a function of latitude for 50 km receiver altitude. EA stands for geometrical elevation angle.

EA (deg)	28°N (control) delay (m)	75°N delay (m)	Δ(%)	0°N delay (m)	Δ(%)
−18	7.77	8.28	6.6	7.67	−1.3
−18.5	60.86	62.81	3.2	62.71	3.0
−19	215.6	231.6	7.4	221.3	2.6

Table 7.3: Variations in neutral atmosphere signal delay from that observed at 28°N as a function of latitude for 350 km receiver altitude. EA stands for geometrical elevation angle.

Applicable Region

Since the neutral atmosphere signal delay is so sensitive to the precise elevation angle of the satellite at higher altitudes, and the delay is also sensitive to how the ionosphere refracts the signal before it enters the neutral atmosphere, I recommend simply not viewing satellites within 1° of the minimum viewable elevation angle, and when the receiver is above 200 km altitude (or roughly equivalently, when the satellite is below $-12°$ elevation angle), simply not viewing satellites whose signals will experience any neutral atmosphere delay, if possible. I have chosen 200 km for its utility; using this altitude as a cut-off value should allow the tracking of satellites that set during reentry. Above 200 km, equation 7.1 gives the minimum elevation angle for which the receiver can expect little to no neutral atmosphere delay. If one must view satellites below this minimum elevation angle, I recommend increasing the uncertainty on the range measurement to these satellites since the calculated neutral atmosphere delays may have significant errors. It is likely that one will only have to view satellites below this minimum elevation angle when they are setting; rising satellites will likely be too difficult to acquire.

7.2 Correcting the Ionosphere Delay

Since the ionosphere is dispersive, a dual frequency receiver can correct for nearly all of the ionosphere delay as described in section 3.2. This dual frequency correction may break down near the horizon if the signals for the two frequencies have significantly different paths through the atmosphere, but will be sufficient at most elevation angles. Ray-tracing results show that signals that have significant total delay ($\approx$ 700m) because the receiver is at a high altitude ($\approx$ 1700 km) observing a satellite at a low elevation angle (-38°) will deviate 52 km from the straight line path but will remain within roughly 75 m of each other at all points (figure 7-3). (Since the ray-tracing root-finder does not require that the two signals hit precisely the same point for the satellite, there is some error in this calculation. This error will increase as the signal approaches the satellite.) Figure 7-4 shows both the ray-traced ionosphere signal delay and the ionosphere delay calculated using the dual frequency correction (eqn. 3.9). The difference between these two values is largest when there is significant refraction in the neutral atmosphere. This difference (shown in the bottom plot) is in part due to the ray-tracing implementation, where the neutral atmosphere was treated as a vacuum. The difference is also caused by the breakdown in the dual

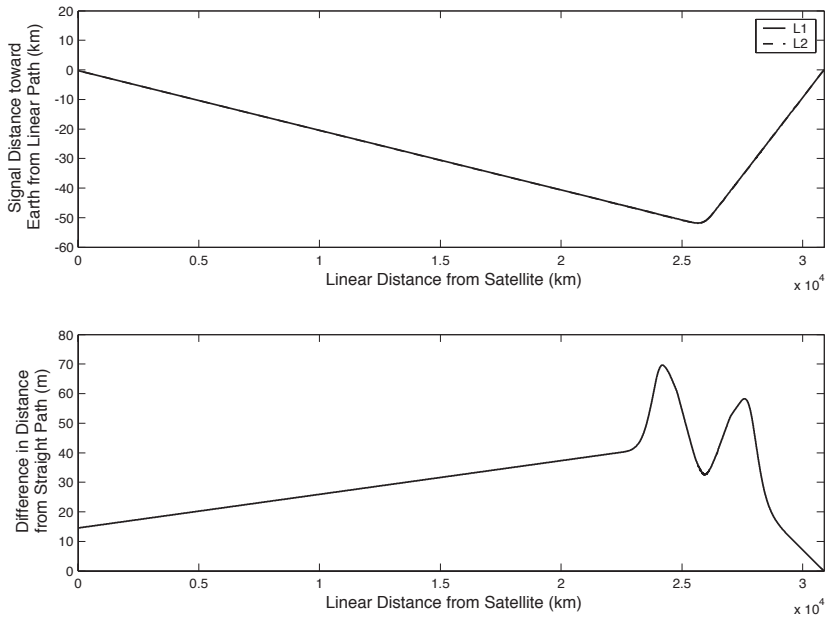

Figure 7-3: The L1 and L2 signals do not follow precisely the same path from the satellite to the receiver due to the dispersive nature of the ionosphere. The ray-tracing root-finder only requires the ray-traced signals to strike within 100 m of the true satellite position, causing the two signals to strike slightly different satellite positions.

frequency correction that occurs when the L1 and L2 paths diverge. The relative magnitudes of these two components is unknown.

As we saw in section 6.2, the ionosphere signal delay experienced by signals from a given satellite can change very rapidly during a space vehicle flight. If a receiver does not have sufficient channels to track all satellites on both the L1 and L2 frequencies the ionosphere correction could introduce a significant error.

I will first discuss why it is impossible accurately to model the ionosphere delay, and then discuss the errors in some ground based receivers' dual frequency correction algorithms and how to fix these errors.

7.2.1 Modeling the Ionosphere Delay

We saw in section 6.1 that the ionosphere signal delay will change significantly with season, solar activity level, time of day, latitude and satellite azimuth. Further, the ionosphere is unpredictable — daily ionosphere delays for a receiver on the ground can vary as much as 25% from the monthly mean. Not only would a model of the ionosphere delay have to be a complex multi-dimensional function of height, elevation

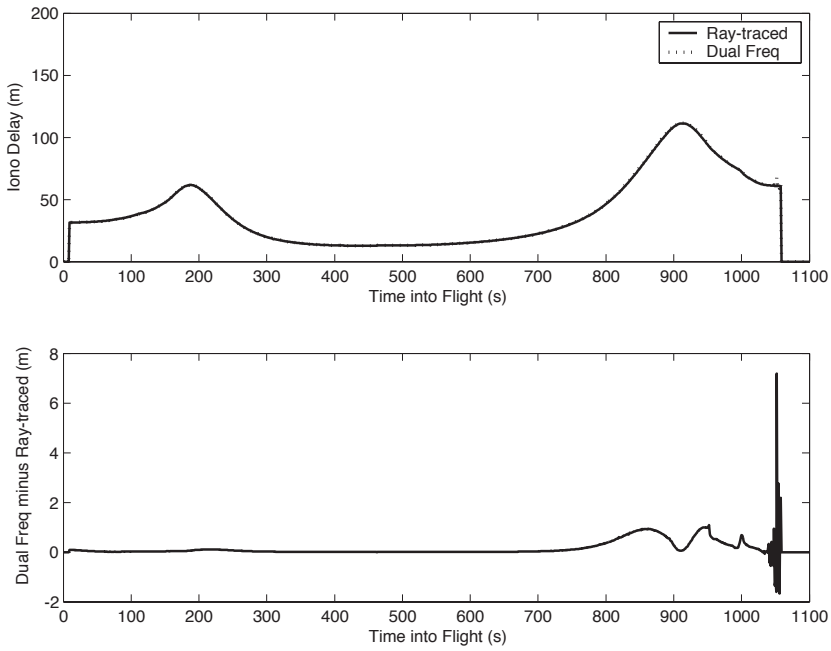

Figure 7-4: A comparison between the dual frequency ionosphere correction and the ray-traced ionosphere delay.

angle and the aforementioned conditions, but it would not be very accurate due to daily variations in electron density.

The satellite azimuth dependence would make the ionosphere delay function very complicated. Unlike the Klobuchar model, we cannot account for the relevant ionosphere contributing the majority of the ionosphere delay (to the east or west, north or south of the receiver) simply by calculating the subionospheric point (the point at which the signal crosses the F2-peak). The Klobuchar model uses the zenith delay at this position with a mapping function to calculate the delay. A receiver aboard a space vehicle can not use this approach. There may be more than one subionospheric point if the satellite is at a negative elevation angle (figure 7-5) and the mapping function will run into the same problems detailed in the neutral atmosphere model section above.

Modeling the ionosphere delay for a user at space vehicle altitudes, taking into account variations in conditions, will be difficult and inaccurate.

7.2.2 Dual Frequency Correction of the Ionosphere Delay

The dual frequency GPS receiver that I examined was designed for a user on or near the surface of the Earth. To conserve channels, it monitored all tracked satellites on

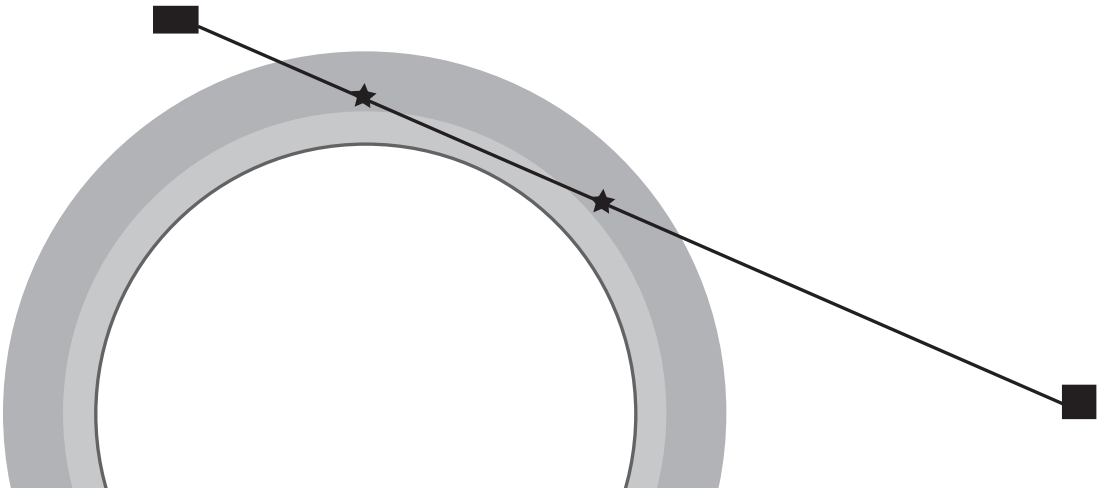

Figure 7-5: Signals traveling to spacecraft potentially pass through two subionospheric points.

their L1 frequency, and dedicated two or three channels to monitor the satellites on their L2 frequency. For a 12 channel receiver, this could mean observing 10 satellites on L1 and using the remaining two channels to cycle through the 10 satellites on L2. If the two channels dedicated to monitoring the L2 frequency spent 5 seconds monitoring each satellite, the receiver would monitor a given satellite for 20 seconds on only L1, followed by 5 seconds spent monitoring the satellite on both L1 and L2. During the 5 seconds of dual frequency monitoring, the receiver estimates an average ionosphere correction, and applies this correction during the next 20 seconds of single frequency monitoring. This type of dual frequency ionosphere correction implementation is called "measure and hold". I have also examined a second dual frequency correction implementation where instead of using the average ionosphere delay during the 20 seconds of single frequency monitoring, I have propagated the slope of the delay calculated during the five seconds of dual frequency monitoring. I call this implementation "measure and propagate".

Both algorithms work well if the ionosphere that a satellite's signals are passing through does not change over the course of 20 seconds. We have seen that this is not the case for a receiver aboard a spacecraft. Figures 7-7, 7-8, 7-9, 7-10, 7-11 and 7-12 show how these two algorithms would perform on six of the satellites from the sample space vehicle flight. In these plots the ionosphere delay calculation during the 5 seconds of dual frequency correction was assumed to be perfect, with no noise on the delay calculation.

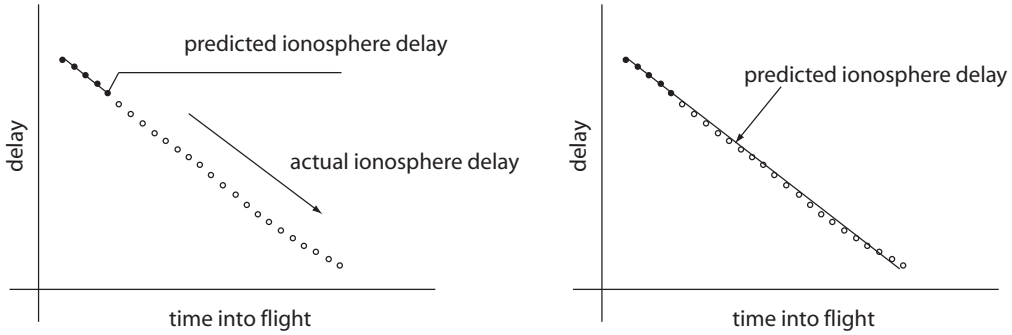

Figure 7-6: LEFT: The measure and hold algorithm makes a dual frequency calculation of the ionosphere signal delay for five seconds, and then uses the average of those measurements as the ionospheric correction over the next 20 seconds. RIGHT: The measure and propagate algorithm maintains the slope of the dual frequency ionosphere correction over the next 20 seconds.

Errors from the measure and hold algorithm are largest when the signal has a large ionosphere delay rate: right after the satellite has risen and the signal is approaching or descending from its peak rising delay, or right before the satellite sets and the signal is approaching or descending from its peak setting delay. During these periods the error in the measure and hold correction can grow to 30 m, but may be quite a bit smaller if the satellite does not change elevation angle quickly. Signals from satellites such as PRN 18 and PRN 9 (in figures 7-11 and 7-12), which are at relatively high elevation angles during the entire flight, only experience ionosphere signal delay during launch and reentry. During these periods, errors from the measure and hold implementation will reach 4–5 m. Satellites that are near the horizon during reentry, such as PRN 10, will have measure and hold errors closer to 10 m.

The measure and propagate algorithm has large errors when the ionosphere delay peaks while the satellite is rising or setting. Since the algorithm cannot predict the change from a positive delay rate to a negative delay rate, it continues to increase the delay, passing the peak and introducing large errors. For the steepest peaks, the measure and propagate algorithm will have errors of over 20 m, but in general, errors remain well below those of the measure and hold algorithm.

Errors from both algorithms can be eliminated by treating satellites that are rising or setting specially. One can either monitor these satellites constantly on two frequencies, or not track these satellites at all. There are two ways to determine when to start tracking on two frequencies or when to stop tracking. First, if the dual frequency correction calculates that there is any ionosphere signal delay, it is

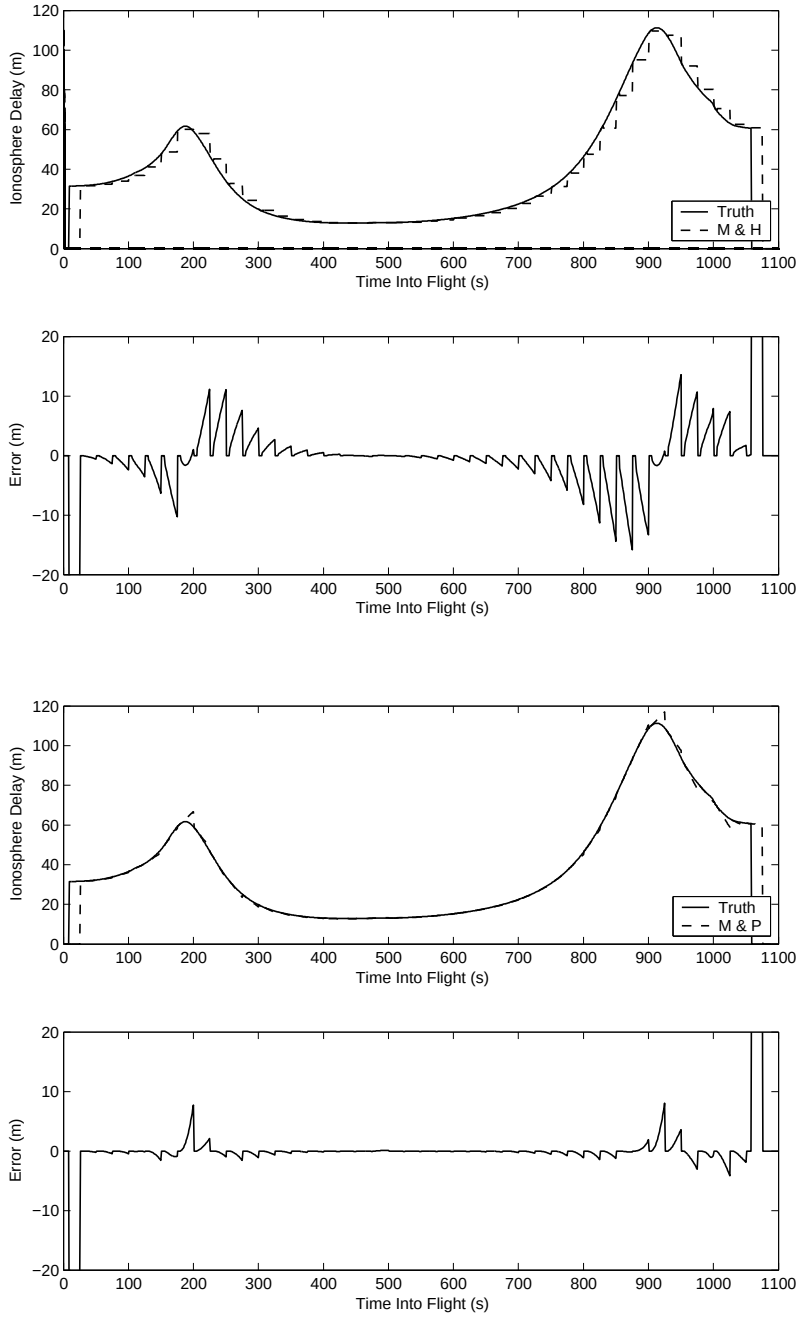

Figure 7-7: TOP: Measure and hold ionosphere delay calculation for PRN 3. SECOND: Error in measure and hold delay calculation. THIRD: Measure and propagate ionosphere delay calculation. BOTTOM: Error in measure and propagate calculation.

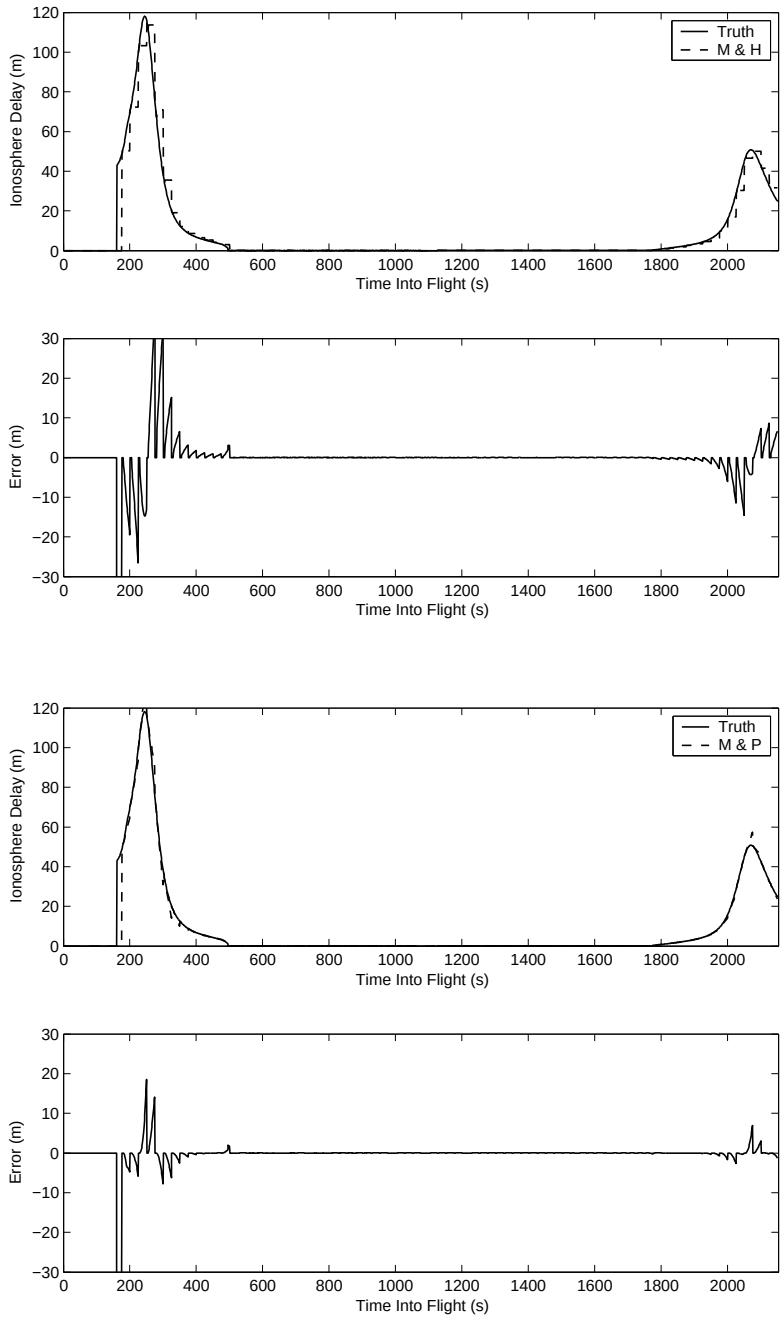

Figure 7-8: TOP: Measure and hold ionosphere delay calculation for PRN 10. SECOND: Error in measure and hold delay calculation. THIRD: Measure and propagate ionosphere delay calculation. BOTTOM: Error in measure and propagate calculation.

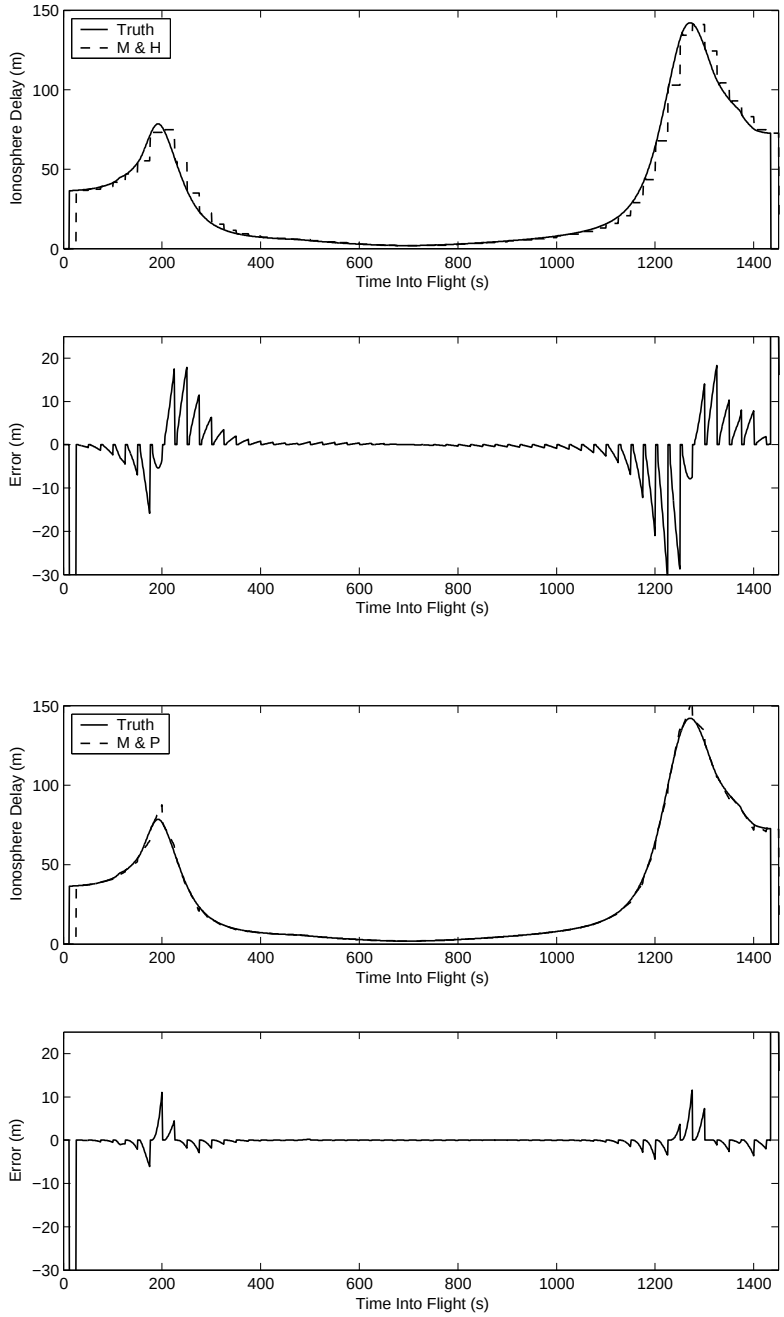

Figure 7-9: TOP: Measure and hold ionosphere delay calculation for PRN 25. SECOND: Error in measure and hold delay calculation. THIRD: Measure and propagate ionosphere delay calculation. BOTTOM: Error in measure and propagate calculation.

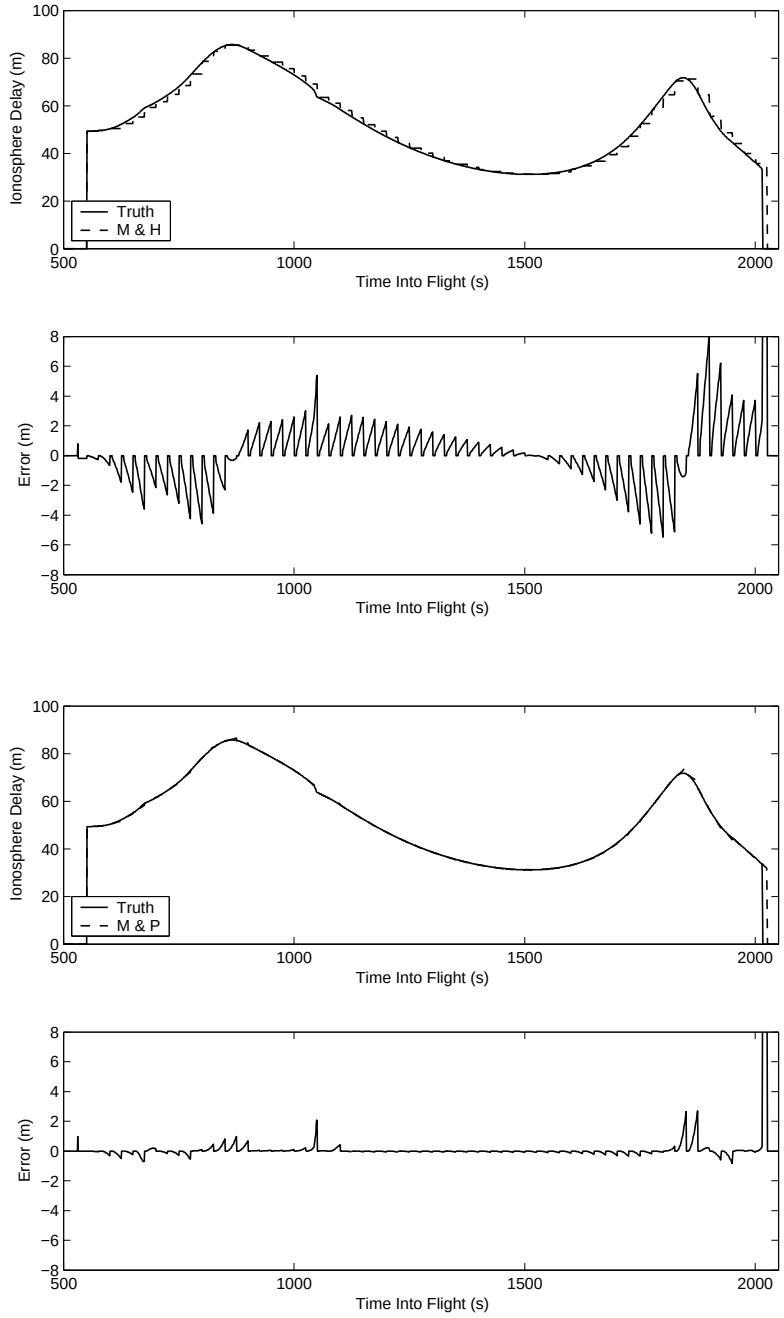

Figure 7-10: TOP: Measure and hold ionosphere delay calculation for PRN 11. SECOND: Error in measure and hold delay calculation. THIRD: Measure and propagate ionosphere delay calculation. BOTTOM: Error in measure and propagate calculation.

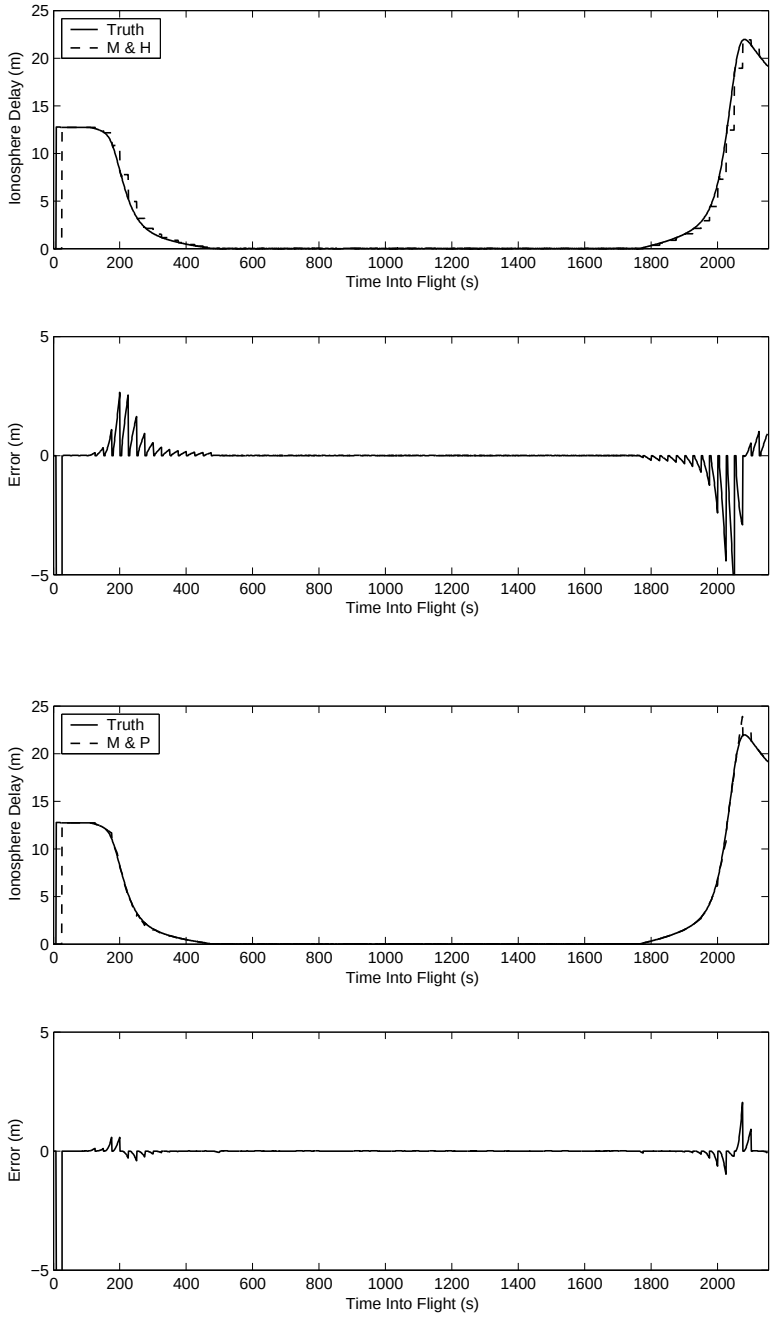

Figure 7-11: TOP: Measure and hold ionosphere delay calculation for PRN 18. SECOND: Error in measure and hold delay calculation. THIRD: Measure and propagate ionosphere delay calculation. BOTTOM: Error in measure and propagate calculation.

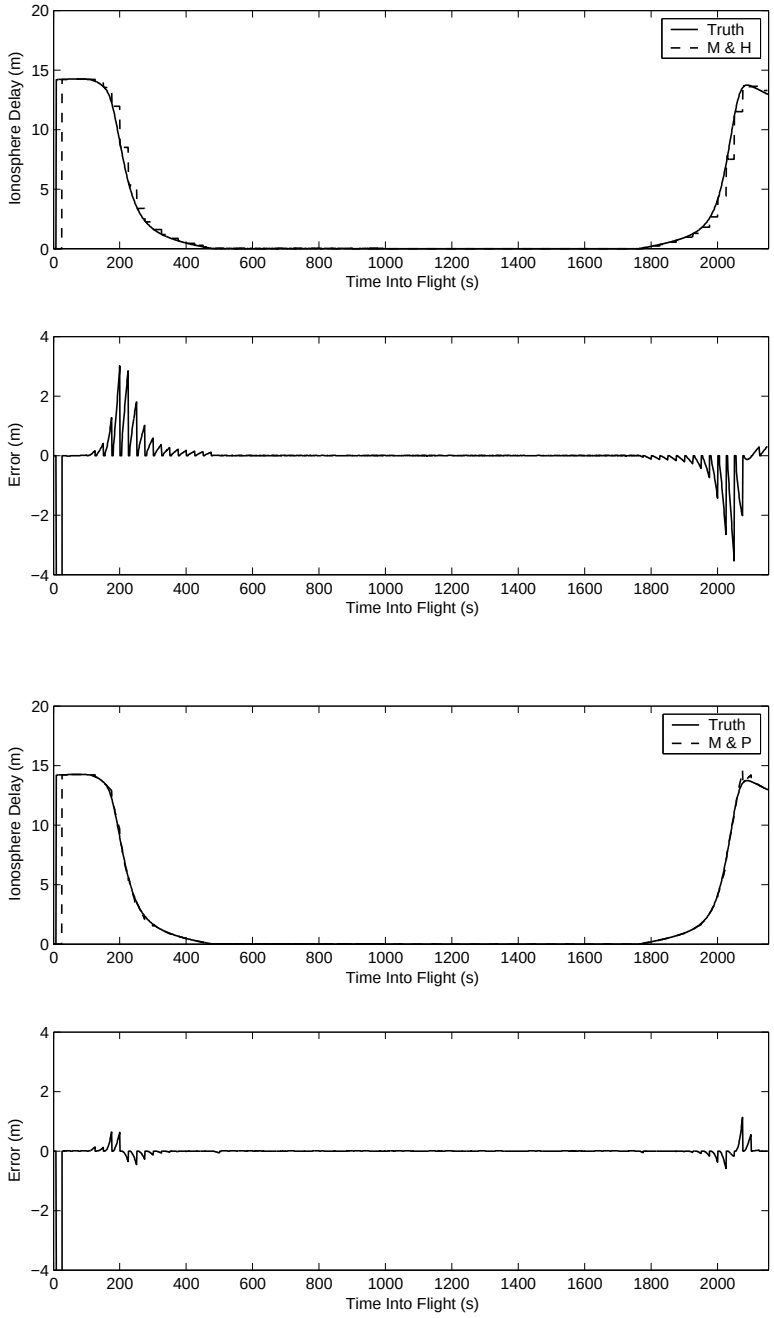

Figure 7-12: TOP: Measure and hold ionosphere delay calculation for PRN 9. SECOND: Error in measure and hold delay calculation. THIRD: Measure and propagate ionosphere delay calculation. BOTTOM: Error in measure and propagate calculation.

reasonable to expect that signals from that satellite are about to experience a change in delay. This applies even to those satellites that are visible at high elevation angles — if signals from those satellites experience ionosphere signal delay, the receiver is traveling through the ionosphere itself and potentially experiencing measure and hold errors of around 3 m (satellites PRN 18 and PRN 9 in figures 7-11 and 7-12). Using the existence of ionosphere delay as a trigger will potentially force monitoring on two frequencies a few minutes before it is really needed. To shorten the time spent dedicating two channels to one satellite (but increase the risk of some measure and hold errors), the receiver could monitor the ionosphere delay rate to each satellite. If the delay rate exceeds a certain value (for example, 0.05 m/s), the receiver will start tracking that satellite on two frequencies. Care would need to be taken to ensure that the selected cut-off values (for either triggering method) are not triggered by noise in the dual frequency ionosphere delay calculation.

If the receiver decides to track a satellite continuously on two frequencies, it must stop tracking another satellite. Instead of tracking ten satellites, it can only track nine since three channels are now dedicated to tracking on L2. The receiver must consider satellite selection criteria such as dilution of precision when deciding whether to track a given satellite on two frequencies (and drop another satellite) or to drop it and start tracking a new satellite instead.

As was mentioned above, the results presented assume that there is no noise in the dual frequency ionosphere delay calculation. Noisy measurements will have significantly more effect on the measure and propagate algorithm than on the measure and hold algorithm, since calculating the average of the measurements is less sensitive to noise than calculating the delay rate. Fortunately, we can calculate the delay rate using phase measurements, which will be less noisy than the calculating the delay rate from code position measurements.

Alternatively, one could eliminate the errors from these two algorithms by solving for the phase cycle integer ambiguity (N) in equation 3.13 during a period of dual frequency correction and use single frequency code and carrier phase range measurements to track the ionosphere signal delay once the L2 signal has been dropped. (See section 3.3 for details.) This method will only work if the L1 signal to a satellite is continuously tracked without cycle slips.

Ideally, one can simply use a receiver that can track all satellites simultaneously on two frequencies.

Testing of the receiver designed for use on or near the surface of the Earth has shown that other aspects of the receiver design need to be modified for use on a space

vehicle. The time scales used in the filter to smooth out the noise in the ionosphere correction are too long for use in a space vehicle receiver. The filter ends up smoothing out the entire peak in delay observed after the satellite rises or before the satellite sets. The receiver also includes reasonableness checks which disregard the large ionosphere delay rates observed by space vehicles as unreasonable.

Shortening the time scales used in the filter and increasing the magnitude of ionosphere delay rate used in a reasonableness check will make the receiver more susceptible to errors from noisy measurements. However, these changes are necessary.

Chapter 8

Summary

8.1 Aim

In this thesis, I sought to provide average neutral atmosphere, ionosphere and total signal delays experienced by GPS signals traveling to a space vehicle during launch, orbit and reentry. With these delays, I sought to determine how well current receiver signal delay corrections performed, particularly during reentry. I found the current receiver delay corrections to be inadequate for use by a space vehicle GPS receiver, and therefore proposed means of correcting these models and algorithms.

8.2 Results

In this thesis, I presented neutral atmosphere, ionosphere and total signal delays calculated by ray-tracing through the Global Reference Atmosphere Model (1999) and the International Reference Ionosphere (2001). I presented delays for a receiver at 0 km to 1700 km altitude as a function of season, solar activity level, time of day, latitude and satellite azimuth.

The delays on signals traveling to space vehicles are potentially much larger than the delays on signals traveling to a receiver on the ground.

A receiver on the surface of the Earth will typically track signals with anywhere from 2–25 m neutral atmosphere delay. I found that at the horizon, signals to a receiver on the ground may experience significant refraction, and up to 130 m neutral atmosphere delay. A receiver on a space vehicle will often be above the neutral atmosphere, and therefore signals will only experience neutral atmosphere delay at a sliver of negative elevation angles right near the horizon. At these elevation angles,

signals will pass through a given altitude of the neutral atmosphere twice — once on the way in and once on the way out. These signals will be refracted significantly, deviating as much as 60 km from the straight path between the receiver and the satellite, and experience hundreds of meters of delay.

Neutral atmosphere signal delays vary as a function of latitude, and to a lesser extent season. These variations are typically within 7% of the mean in the ray-tracing results.

Signals traveling to a receiver on the Earth will typically experience 2–30 m of ionosphere signal delay, whereas those traveling to a receiver at altitude may experience upwards of 150 m of ionosphere delay. The ionosphere both covers a thicker range of altitudes and at a higher altitude than the neutral atmosphere. All signals that pass through the neutral atmosphere will also pass through the ionosphere, but the converse is not true. Therefore, when the receiver is at altitude, signals from a larger percentage of GPS satellites will pass through the ionosphere than will pass through the neutral atmosphere. The ionosphere delay on these signals will be largest when the signals pass tangent to the layer of the ionosphere with the highest electron density, at approximately 350 km altitude.

Ionosphere signal delays vary as a function of season, solar activity level, time of day, latitude and satellite azimuth. Changing each of these conditions varies the electron density distribution in the ionosphere. These variations have two main effects on the ionosphere delay. If the altitude of the peak electron density increases, the elevation angle at which the signal delay peaks also increases. If the overall electron density decreases, the signal delay decreases as well.

GPS signals traveling to a space vehicle will only experience neutral atmosphere delay in four situations: launch, reentry and when a satellite is rising or setting. A current receiver-based neutral atmosphere delay model, the NATO model, will adequately predict neutral atmosphere delays during launch and reentry, with errors less than 1 m. However, the NATO model will not predict delays when a satellite is rising or setting since it does not cover satellites at most negative elevation angles. While a space vehicle is orbiting, it may simply stop tracking these satellites since many others are likely visible. I have provided a formula to determine the appropriate elevation angle at which to stop tracking GPS satellites for this purpose. During reentry, however, a significant portion of the sky may be blocked by the both the vehicle body and the Earth, and the space vehicle may want to continue to track these satellites until they set. In this case, I recommend creating an explicit function for the neutral atmosphere delay of elevation angle and receiver altitude. Ray-tracing

results suggest errors will typically be at most 7%. As the signal passes closer to the surface of the Earth and refraction becomes more significant, the delay will become very sensitive to the precise elevation angle of the satellite, and the errors from this correction will increase.

GPS signals traveling to a space vehicle will likewise experience ionosphere delay during launch and reentry, and when satellite is rising or setting. However, since the ionosphere is significantly thicker than the neutral atmosphere, the length of time for which a given satellite's signals will experience ionosphere delay is much longer than the time for which they will experience neutral atmosphere delay. It is possible for a satellite to remain near the horizon, with its signals experiencing ionosphere delay, for over 20 minutes. The Klobuchar model will not be adequate when the receiver is at altitude. During reentry, the model often performs well with errors on the order of a few meters or less, but it will perform poorly for satellites near the horizon, with errors as high as 30 m. For dual frequency receivers employing a measure and hold algorithm, the dual frequency ionosphere correction can be off by as much as 30 m during flight. During reentry, these errors will decrease, but remain as large as 10 m. Unlike the neutral atmosphere, an ionosphere delay model for receivers at altitude is not a viable option. The errors from daily variations alone are too large. I therefore recommend improving the measure and hold algorithm. I propose implementing one of two improvements. One could trigger continuous dual frequency correction for a satellite when the ionosphere delay rates exceed a threshold value, or more conservatively, trigger dual frequency correction when the ionosphere signal delay is measured to be above a certain level. This approach has two problems: noise in the dual frequency ionosphere delay measurement could falsely trigger continuous dual frequency correction, and there are potentially tens of seconds when the ionosphere delay could be changing before the continuous dual frequency correction is triggered. Alternatively, one could solve for the phase cycle integer ambiguity during a period of dual frequency ionosphere correction and use single frequency code and carrier phase range measurements to track the ionosphere signal delay once the L2 signal has been dropped. This second approach will allow more satellites to be tracked since no satellites will require continuous dual frequency tracking.

8.3 Future Research

Two aspects of this research must still be addressed.

First, equations describing the neutral atmosphere delay as a function of receiver

altitude and elevation angle to satellite must be constructed. A full Monte Carlo analysis of the likely errors of this model using the GRAM-99 perturbation model is advisable.

Second, the suggested modifications to the measure and hold algorithm need to be tested and the errors from the implementation analyzed.

Bibliography

[1] Klobuchar, J. A., "Ionospheric Time-Delay Algorithm for Single-Frequency GPS Users," *IEEE Transactions on Aerospace and Electronic Systems*, Vol. AES-23, No. 3, May 1987, pp. 325–331.

[2] Hopfield, H. S., "Tropospheric Effects on Low-Elevation-Angle Signals: Further Studies," Tech. Rep. APL/JHU SDO 4588, Johns Hopkins University Applied Physics Laboratory, Laurel, MD, Aug 1976.

[3] Hopfield, H. S., "Two-Quartic Tropospheric Refractivity Profile for Correcting Satellite Data," *Journal of Geophysical Research*, Vol. 74, No. 18, Aug 1969, pp. 4487–4499.

[4] Hitney, H. V., Richter, J. H., Pappert, R. A., Anderson, K. D., and Jr., G. B. B., "Tropospheric Radio Propagation Assessment," *Proceedings of the IEEE*, Vol. 73, No. 2, Feb 1985, pp. 265–283.

[5] Hecht, E., *Optics*, Addison-Wesley, New York, 1998.

[6] Seeber, G., *Satellite Geodesy*, Walter de Gruyter, New York, 2nd ed., 2003.

[7] Hofmann-Wellenhof, B., Lichtenegger, H., and Collins, J., *GPS Theory and Practice*, SpringerWein, New York, 5th ed., 2001.

[8] Boudouris, G., "On the Index of Refraction of Air, the Absorption and Dispersion of Centimeter Waves by Gases," *Journal of Research of the National Bureau of Standards – D. Radio Propagation*, Vol. 67D, No. 6, Nov-Dec 1963, pp. 631–684.

[9] Davis, J. L., Herring, T. A., Shapiro, I. I., Rogers, A. E. E., and Elgered, G., "Goedesy by radio interferometry: Effects of atmospheric modeling errors on estimates of baseline length," *Radio Science*, Vol. 20, No. 6, Nov-Dec 1985, pp. 1593–1607.

[10] Smith, E. K. and Weintraub, S., "The Constants in the Equation for Atmospheric Refractive Index at Radio Frequencies," *Proceedings of the I.R.E.*, Vol. 41, 1953, pp. 1035–1037.

[11] Owens, J. C., "Optical Refractive Index of Air: Dependence on Pressure, Temperature and Composition," *Applied Optics*, Vol. 6, No. 1, Jan 1967, pp. 51–59.

[12] Thayer, G. D., "An improved equation for the radio refractive index of air," *Radio Science*, Vol. 9, No. 10, Oct 1974, pp. 803–807.

[13] Spilker, J. J., *Global Positioning System: Theory and Applications, Vol. 1*, Vol. 163 of *Progress in Astronautics and Aeronautics*, chap. 13, AIAA, Washington, D.C., 1996, pp. 517–546.

[14] Hill, R. J., Lawrence, R. S., and Priestly, J. T., "Theoretical and calculational aspects of the radio refractive index of water vapor," *Radio Science*, Vol. 17, No. 5, Sept-Oct 1982, pp. 1251–1257.

[15] Hill, R. J., "Dispersion by Atmospheric Water Vapor at Frequencies Less Than 1 THz," *IEEE Transactions on Antennas and Propagation*, Vol. 36, No. 3, March 1988, pp. 423–430.

[16] Bevis, M., Businger, S., Chiswell, S., Herring, T. A., Anthes, R. A., Rocken, C., and Ware, R. H., "GPS Meteorology: Mapping Zenith Wet Delays onto Precipitable Water," *Journal of Applied Meteorology*, Vol. 33, March 1994, pp. 379–386.

[17] Birnbaum, G. and Chatterjee, S. K., "The Dielectric Constant of Water Vapor in the Microwave Region," *Journal of Applied Physics*, Vol. 23, No. 2, Feb 1952, pp. 220–223.

[18] Mendes, V. B., *Modeling the Neutral-Atmosphere Propagation Delay in Radiometric Space Techniques*, Ph.D. thesis, University of New Brunswick, Fredericton, New Brunswick, Canada, Sept 1998.

[19] Davies, K., *Ionospheric Radio*, Peter Peregrinus Ltd., London, 1990.

[20] Klobuchar, J. A., *Global Positioning System: Theory and Applications, Vol. 1*, Vol. 163 of *Progress in Astronautics and Aeronautics*, chap. 12, AIAA, Washington, D.C., 1996, pp. 485–515.

[21] Wallace, J. M. and Hobbs, P. V., *Atmospheric Science: An Introductory Survey*, Academic Press, New York, 1977.

[22] Hopfield, H. S., "Improvements in the Tropospheric Refraction Correction for Range Measurement," *Philosophical Transactions of the Royal Society of London, Series A, Mathematical and Physical Sciences*, Vol. 294, No. 1410, Jan 1980, pp. 341–352.

[23] Tascione, T. F., *Introduction to the Space Environment*, Orbit Foundation Series, Krieger Publishing Company, 2nd ed., 1994.

[24] Schüler, T., *On Ground Based GPS Tropospheric Delay Estimation*, Ph.D. thesis, Universität der Bundeswehr, Munich, Germany, February 2001.

[25] Saastamoinen, J., *The Use of Artificial Satellites for Geodesy*, No. 15 in Geophysical Monograph, American Geophysical Union, Washington, D.C., 1972, pp. 247–251.

[26] Craig, R. A., *The Upper Atmosphere: Meteorology and Physics*, Academic Press, New York, 1965.

[27] Elgered, G., Rönnäng, B. O., and Askne, J. I. H., "Measurements of atmospheric water vapor with microwave radiometry," *Radio Science*, Vol. 17, No. 5, Sept-Oct 1982, pp. 1258–1264.

[28] Niell, A. E., "Global Mapping Functions for the Atmosphere Delay at Radio Wavelengths," *Journal of Geophysical Research*, Vol. 101, No. B2, February 1996, pp. 3227–3246.

[29] Bilitza, D. K., "International Reference Ionosphere 2000," *Radio Science*, Vol. 36, No. 2, Mar-Apr 2001, pp. 261–275.

[30] Stephanakis, S. J., Balwanz, W. W., Ament, W. S., and Faust, W. R., "The Possibility of Electromagnetic Wave Ducting in the Ionosphere," Tech. Rep. NRL-7171, Naval Research Laboratory, Washington, D.C., 1970.

[31] Mendes, V. B. and Langley, R. B., "An Analysis of High-Accuracy Tropospheric Delay Mapping Functions," *Physics and Chemistry of the Earth, A*, Vol. 25, No. 12, 2000, pp. 809–812.

[32] Altshuler, E. E. and Kalaghan, P. M., "Tropospheric Range Error Corrections for the NAVSTAR System," Air Force Surveys in Geophysics (AFSG) No. 286 AFCRL-TR-74-0198, Air Force Cambridge Research Laboratories, Bedford, MA, 1974.

[33] Altshuler, E. E., "Tropospheric Range-Error Corrections for the Global Positioning System," *IEEE Transactions on Antennas and Propogation*, Vol. 46, No. 5, May 1998, pp. 643–649.

[34] Braasch, M. S., "A Signal Model for GPS," *Navigation: Journal of the Institute of Navigation*, Vol. 34, No. 4, Winter 1990, pp. 363–377.

[35] Cohen, C. E., Pervan, B., and Parkinson, B. W., "Estimation of Absolute Ionospheric Delay Exclusively through Single-Frequency GPS Measurements," *Proceedings of ION GPS-92: Fifth International Technical Meeting of the Satellite Division of The Institute of Navigation*, Albuquerque, New Mexico, Sept. 1992, pp. 325–330.

[36] Trethewey, M. L., Catchpole, I., and Hansla, A., "Single Frequency Ionosphere Determination Using GPS," *Proceedings of ION GPS-93; Sixth International Technical Meeting of the Satellite Division of the Institute of Navigation*, Salt Lake City, UT, Sept. 1993, pp. 1373–1381.

[37] Nisner, P. and Trethewey, M. L., "GPS Ionosphere Determination Using L1 Only," *Proceedings of the National Technical Meeting of the Institute of Navigation*, Santa Monica, CA, Jan. 1996, pp. 625–635.

[38] *U.S. Standard Atmosphere, 1976*, NOAA; U.S. Committee on the Extension to the Standard Atmosphere, Washington, D.C., 1976.

[39] *Guide to Reference and Standard Atmosphere Models*, No. ANSI/AIAA G-003A-1996, AIAA, 1997.

[40] Justus, C. G. and Johnson, D. L., "The NASA/MSFC Global Reference Atmosphere Model – 1999 Version (GRAM-99)," Tech. Rep. NASA/TM-1999-209630, NASA/Marshall Space Flight Center, MSFC, Alabama, May 1999.

[41] Justus, C. G., Jeffries, W. R., Yung, S. P., and Johnson, D. L., "The NASA/MSFC Global Reference Atmosphere Model – 1995 Version (GRAM-95)," Tech. Rep. NASA TM-4715, NASA/Marshall Space Flight Center, MSFC, Alabama, 1995.

[42] *Guide to Reference and Standard Ionosphere Models*, No. ANSI/AIAA G-034-1998, AIAA, 1999.

[43] Rush, C. M., "Ionospheric Radio Propagation Models and Predictions – A Mini-Review," *IEEE Transactions on Antennas and Propagation*, Vol. AP-34, No. 9, Sept. 1986, pp. 1163 –1170.

[44] Llewellyn, S. K. and Bent, R. B., "Documentation and Description of the Bent Ionospheric Model," Tech. Rep. AFCRL-TR-73-0657, SAMSO TR-73-252, Atlantic Sciences Corporation, Indialantic, FL, July 1973.