

RAY-TRACING OPTIONS TO MITIGATE THE NEUTRAL ATMOSPHERE DELAY IN GPS

FELIPE GEREMIA NIEVINSKI

January 2009



**TECHNICAL REPORT
NO. 262**

RAY-TRACING OPTIONS TO MITIGATE THE NEUTRAL ATMOSPHERE DELAY IN GPS

Felipe Geremia Nievinski

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

January 2009

© Felipe G. Nievinski 2009

PREFACE

This technical report is a reproduction of a thesis submitted in partial fulfillment of the requirements for the degree of Master of Science in Engineering in the Department of Geodesy and Geomatics Engineering, January 2009. The research was supervised by Dr. Marcelo Santos, and support was provided by the Canadian International Development Agency, the Geomatics for Informed Decisions Network Centres of Excellence and by the Natural Sciences and Engineering Research Council of Canada.

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

Nievinski, Felipe G. (2009). *Ray-tracing Options to Mitigate the Neutral Atmosphere Delay in GPS*. M.Sc.E. thesis, Department of Geodesy and Geomatics Engineering Technical Report No. 262, University of New Brunswick, Fredericton, New Brunswick, Canada, 232 pp.

Dedication

To Giovana.

Abstract

As the radio signals emanating from GPS satellites propagate through the Earth's electrically neutral (i.e., un-ionized) atmosphere, they suffer refraction. The effect of refraction on GPS timing measurements is a delay compared to what would be measured had the signal propagated in a vacuum. Equivalently, assuming vacuum speed of propagation, refraction makes the apparent distance measured with GPS larger than the geometric distance between receiver and satellite. If not adequately mitigated, that delay corrupts estimates, such as receiver position, obtained from GPS observations. One way of quantifying the neutral atmosphere radio propagation delay is supposing the signal to be a ray, and tracing that ray along its path, from satellite to receiver, through a model for the atmosphere; we call such a procedure ray-tracing, and it constitutes our main interest in this work. Ray-tracing has connections to many different subject areas. Among those, we see the present work falling under the umbrella of geodesy. More specifically, we see it situated along the thread of developments of the so-called mapping functions for radio space geodetic applications.

The main research contribution from this work is the identification, classification, and comparison of alternative models for the ray-path and the atmospheric structure employed in ray-tracing. It is a three-part contribution, parts that we now discuss. First we distinguished among the ray-tracing options known as atmospheric source, atmospheric structure, and ray-path model. Such distinction classifies the myriad of options available in

ray-tracing in separate groups, disentangling aspects that are typically (sometimes arguably conveniently) lumped together. For example, a sentence such as “this ray-tracer assumes spherical symmetry” actually makes separate statements about the assumed ray-path and the atmospheric structure models. Secondly, we identified model alternatives within each of the three options above, namely, spherical concentric, spherical osculating, ellipsoidal, gradient, and 3d atmospheric structures; and zenith, straight-line, bent-2d, and bent-3d ray-path models. Thirdly, we compare experimentally different models. More specifically, we quantified the discrepancy in delay between different models and we also assessed their impact in GPS positioning. In addition to the three-part main contribution above, a secondary contribution is a classification of the delay mitigation techniques available in GPS, developed to support the design of the GPS experiments.

The findings of this work are as follows. (i) Regarding the ray-path, the bent-2d model, albeit not strictly valid in a 3d atmosphere, introduces only negligible errors, compared to the more rigorous bent-3d model (in a 15 km horizontal resolution atmospheric model). Regarding atmospheric structures, we found that (ii) the oblateness of the Earth cannot be neglected when it comes to predicting the neutral atmosphere delay, as demonstrated by the poor results of a spherical concentric atmosphere; (iii) the spherical osculating model is the only one exhibiting azimuthal symmetry; (iv) the oblateness of the Earth is adequately accounted for by a spherical osculating model, as demonstrated by the small discrepancy between a spherical osculating and a more rigorous ellipsoidal model; (v) a gradient atmosphere helps in accounting for the main trend in azimuthal asymmetry exhibited by a 3d atmosphere, but there remains secondary directions of azimuthal asymmetry that only a full 3d atmosphere is able to capture. Furthermore, (vi) we found experimental evidence confirming the theoretical expectation that gradient and especially 3d atmospheric structures offer promising benefits for GPS positioning. Finally, beyond the comparison of atmospheric structures above, an interesting side conclusion regarding atmospheric sources

was that (vii) atmospheric models of higher resolution might offer significant improvements in mapping functions.

Acknowledgments

Besides institutions and individuals with whom I was affiliated during my Master's studies, I acknowledge the support received from the following:

- IGS, the International GNSS Service, for providing raw GPS tracking data as well as products such as satellite orbits and clocks and also station positions and velocities [Dow et al., 2005];
- the Weather Office at Environment Canada, for providing the numerical output from GEM, the Canadian Regional Global Environmental Multiscale Numerical Weather Model [Côté et al., 1998];
- Rodrigo Leandro, for providing and adapting GAPS, a processing software for Precise Point Positioning [Leandro et al., 2007], and also for numerous helpful discussions and suggestions;
- Terje O. Espelid, for his globally adaptive numerical integration routines [Espelid, 2007];
- the Goddard VLBI group, for providing atmospheric pressure loading time series¹ [Petrov and Boy, 2004];
- numerous digital libraries, such as the Smithsonian Astrophysical Observatory/NASA Astrophysics Data System² and the NASA Technical Reports Server³, for digitizing

¹Available at <http://gemini.gsfc.nasa.gov/aplo>

²Available at <http://ads.harvard.edu/>

³Available at <http://ntrs.nasa.gov/>

and providing free access to historical technical and scientific documents.

I acknowledge funding provided by the Canadian International Development Agency, CIDA, and the GEOIDE, Geomatics for Informed Decisions Network of Centres of Excellence. I would like to thank my supervisor, Dr. Marcelo Santos, for giving me the opportunity to pursue graduate studies at the present Department, for providing a nurturing working environment throughout that period, and for always being there when help was needed, gratitude that I extend to his wife, Ms. Denise Ferreira, and their son, Thales. Finally, I would like to dedicate this thesis to Giovana Ferrer Vazatta, who became my wife during the period in which the work reported here was developed: This journey would not have been nearly as fun without you.

“What I cannot create, I do not understand.”
— Richard P. Feynman

Table of Contents

Dedication	ii
Abstract	iii
Acknowledgments	vi
Table of Contents	ix
List of Tables	xii
List of Figures	xiii
Abbreviations	xvii
1 Introduction	1
1.1 Overview	1
1.2 Scope	2
1.3 Contributions from this work	3
1.4 Review of previous related work	5
1.5 Structure of this document	6
2 Background	7
2.1 Index of refraction	8
2.2 Defining delay	9
2.2.1 Integrands and integration paths	9
2.2.2 Path lengths	11
2.2.3 Total, along-path, and geometric delays	13
2.2.4 Geometric delay vs. bending delay	14
2.3 Independent variables and rates of changes	15
2.3.1 On the need for modeling to mitigate the delay	17
2.4 Decomposing the delay: zenith delay and slant factors	17
2.4.1 Scope: slant factors in, zenith delays out	20
2.5 Separating the delay: larger and more stable vs. smaller and more variable	22
2.5.1 Refractivity components	22
2.5.2 Slant delay components	24

2.5.3	Zenith delay and slant factor components	26
2.5.4	Motivation	27
2.6	Developing a mapping function	28
2.6.1	Atmospheric modeling	29
2.6.2	Delay modeling	31
2.6.3	Slant factor modeling	31
2.6.4	Validation and assessment	34
2.7	Atmospheric sources	35
2.7.1	Radiosondes	36
2.7.2	Numerical weather models	37
2.7.3	Climatologies	39
2.8	Our scope in the historical context	40
3	Ray-tracing options	52
3.1	Ray-path models	54
3.1.1	Bent-3d	55
3.1.2	Bent-2d	56
3.1.3	Straight-line	57
3.1.4	Zenith	57
3.2	Atmospheric structures	58
3.2.1	Spherical concentric	59
3.2.2	Spherical osculating	60
3.2.3	Ellipsoidal	60
3.2.4	Gradient	61
3.2.5	Three-dimensional (3d)	65
4	The neutral atmosphere delay in GPS positioning	67
4.1	Vulnerabilities	67
4.2	Mitigation techniques	72
5	Experiments	81
5.1	Delay	82
5.2	GPS positioning: delay mitigation via prediction only	103
5.3	GPS positioning: delay mitigation via estimation	122
6	Conclusions	134
	References	138
	Appendices	153
I	Ray-tracing	155
I.1	Delay formulation	155

I.1.1	Calculating the total delay as apparent ray-path length minus geometric distance	156
I.1.2	Calculating the total delay as along-path delay plus geometric delay	158
I.2	Solver	161
I.3	Bent ray-path	164
I.3.1	Integration limits and orientation	164
I.3.2	Notable points and directions	165
I.3.3	Boundary conditions	171
I.3.4	Ray evolution	173
I.3.4.1	Special cases	174
I.4	Cross-validation	176
I.4.1	Cross-validating ray-path models	177
I.4.2	Cross-validating atmospheric structure models	180
I.5	Further numerical aspects	181
I.5.1	Along-path delay	181
I.5.2	Geometric delay	182
I.5.3	Bent ray-path	185
I.5.3.1	Initial value problem	185
I.5.3.2	Boundary value problem	188
I.5.3.3	Tolerance for boundary conditions	190
I.5.3.4	Gradients and partial and directional derivatives	191
II	Geopotential height	194
II.1	Increments in geopotential height in atmospheric sciences	199
II.2	Increments in geopotential height in geodesy	202
II.3	Conversion to geopotential height	204
III	Pressure interpolation	209
IV	Humidity conversion	212
V	Loxodrome in a gradient atmospheric structure	221
VI	Refractivity coefficients	226
VII	Related work in secondary areas	230
	Vita	

List of Tables

2.1	Four combinations of integrand and integration path.	11
5.1	Epoch and location of first experiment.	84
5.2	Bias in static position estimates using various mapping functions.	128
I.1	Comparisons for cross-validating ray-path models.	178
I.2	Comparisons for cross-validating atmospheric structures.	178
II.1	RMS of discrepancy in geopotential height due to different references. . .	207
IV.1	Conversion <i>to</i> partial pressure of water vapor.	213
IV.2	Conversion <i>from</i> partial pressure of water vapor.	213
VI.1	Refractivity coefficient values.	228
VI.2	Zenith and slant delays based on refractivity coefficients from IUGG. . . .	228
VI.3	Discrepancy (with respect to IUGG) in zenith delay (in m) due to different determinations of the refractivity coefficients.	228
VI.4	Discrepancy (with respect to IUGG) in slant delay (in m) due to different determinations of the refractivity coefficients.	229
VI.5	Discrepancy (with respect to IUGG) in slant factor (multiplied by a nom- inal zenith delay, thus in m) due to different determinations of the refrac- tivity coefficients.	229

List of Figures

1.1	Diagram of options available in ray-tracing.	4
2.1	Diagram of definitions of delays in terms of path lengths.	15
2.2	Diagram of scope concerning prediction and estimation of zenith delays and slant factors.	20
2.3	Diagram of development of mapping functions.	29
2.4	Sample radiosonde.	37
2.5	Sample numerical weather model	38
2.6	Sample climatology.	39
2.7	Diagram of scope of this work.	40
3.1	Diagram of options available in ray-tracing.	53
3.2	Sample spherical concentric iso-indicial (in red).	59
3.3	Sample spherical osculating iso-indicial (in red).	61
3.4	Angles, distances, and vectors involved in a spherical osculating atmosphere.	62
3.5	Sample ellipsoidal iso-indicial (in red).	63
3.6	Sample gradient iso-indicial (in red) in the vicinity of a base point (marked as a cross).	64
3.7	Angles, distances, and vectors involved in a gradient atmosphere.	65
3.8	Sample 3d iso-indicial.	66
4.1	Shift in ranging due to shift in zenith delay and receiver height.	70

4.2	Zenith distance of the same satellite at two different receivers.	76
5.1	Diagram of options available in ray-tracing.	82
5.2	Diagram of sequence of pair-wise comparisons.	83
5.3	Results of first comparison: spherical concentric minus spherical osculating.	86
5.4	Additional results of first comparison: spherical concentric delay (itself).	88
5.5	Additional results of first comparison: spherical concentric delay (self-discrepancy with respect to north azimuth).	89
5.6	Additional results of first comparison: spherical concentric total delay (itself).	90
5.7	Additional results of first comparison: spherical concentric total delay (self-discrepancy with respect to north azimuth).	91
5.8	Horizons of each concentric and osculating spheres	92
5.9	Radial directions of each concentric and osculating spheres.	93
5.10	Results of second comparison: spherical osculating minus ellipsoidal.	94
5.11	Constant radius of osculating sphere vs. azimuth-varying radius of ellipsoid	96
5.12	Results of third comparison: ellipsoidal minus gradient.	97
5.13	Additional results of third comparison: spherical osculating minus gradient.	99
5.14	Results of fourth comparison: gradient minus 3d.	100
5.15	Results of fifth comparison: bent-2d minus bent-3d.	102
5.16	Additional results of fifth comparison: bent-2d minus bent-3d (total delay).	103
5.17	Raw residuals (spherical concentric atmosphere).	106
5.18	Number of observations per bin.	108
5.19	Remaining residuals after discard (spherical concentric atmosphere).	109
5.20	Remaining residuals after discard (spherical osculating atmosphere).	110
5.21	Remaining residuals after discard (3d atmosphere).	111
5.22	Remaining residuals after discard (GMF).	112

5.23	Standard deviation of residuals per bin, as well as their respective 95% confidence interval (shown as error bars).	113
5.24	Mean of residuals per bin, as well as their respective 95% confidence interval (shown as error bars).	115
5.25	Slant delay error due to nominal zenith delay error (1 cm).	117
5.26	Slant delay error due to nominal zenith delay error (−1 cm).	117
5.27	Replicated residuals (see text for details).	119
5.28	Replicated residuals (see text for details).	119
5.29	Estimated and predicted zenith delay.	120
5.30	Slant delay error due to realistic zenith delay error.	120
5.31	Replicated residuals (see text for details).	121
5.32	Replicated residuals (see text for details).	121
5.33	Zenith delay.	127
5.34	Vertical displacement due to non-tidal atmospheric pressure loading, over the week July 22–28, 2007, at station ALGO.	132
I.1	Along-path delay as the area under the curve of refractivity vs. along-path distance.	161
I.2	Diagram of steps in delay solver.	162
I.3	Receiver–satellite geometry.	166
I.4	Notable positions in a bent ray-path.	167
I.5	Notable relative directions in a bent ray-path.	168
I.6	Definition of maximum along-path distance.	169
I.7	Notable tangent directions in a bent ray-path.	170
I.8	Diagram of options available in ray-tracing.	176
I.9	Notable positions in a bent ray-path.	184

II.1	Distortion in a refractivity profile due to errors in the datum and in relative geopotential heights.	196
II.2	Refractivity versus error in relative geopotential height in a vertical profile.	197
II.3	Nominal rate of change of zenith hydrostatic delay with respect to height, along a vertical profile.	198
II.4	Discrepancy in geopotential height due to different gravity formulations. .	205
II.5	Discrepancy in zenith hydrostatic delay due to different gravity formulations.	206
III.1	Consequence of linear versus log-linear interpolation of pressure.	210
III.2	Consequence of linear versus log-linear interpolation of pressure (zoom).	211

List of Symbols, Nomenclature or Abbreviations

α	azimuth	rad
β	virtual temperature lapse-rate	K/m
β'	virtual temperature lapse-rate	K/m
ΔZ	relative geopotential height	m
Δz	relative geometric height	m
ε	elevation angle	rad
ℓ	along-path distance	m
ℓ_{exit}	along-path distance at which atmosphere vanishes	m
∇n	gradient of refractivity (vector)	1/m
\mathbf{A}	design matrix	undefined
\mathbf{C}_x	a priori covariance matrix of the parameters	undefined
\mathbf{J}	Jacobian of transformation between local and global Cartesian coordinates	m/m
φ	geodetic latitude	rad
$\hat{\mathbf{r}}$	estimated residual vector	undefined
$\hat{\mathbf{x}}$	estimated parameters vector	undefined
ℓ	vector of observations	undefined
$\tilde{\mathbf{x}}$	approximate parameters vector	undefined
\mathbf{w}	misclosure vector	undefined
λ	longitude	rad

RH	relative humidity	(percent)
ρ	total mass density	kg/m ³
$\hat{\mathbf{s}}$	tangent direction (unit vector)	unitless
$\hat{\mathbf{s}}_0$	initial ray tangent direction (unit vector)	unitless
$\hat{\mathbf{s}}_{\text{exit}}$	ray tangent direction at the pierce point (unit vector)	unitless
$\hat{\mathbf{s}}_{\text{miss}}$	ray tangent direction at the point of closest approach (unit vector)	unitless
\mathbf{a}	curvature (vector)	1/m
\mathbf{g}	actual gravity (vector)	m/s ²
\mathbf{r}	position (vector)	m
\mathbf{r}'	eccentric position (vector)	m
\mathbf{r}_0	initial position (vector)	m
\mathbf{r}_{rec}	receiver position (vector)	m
\mathbf{r}_{sat}	satellite position (vector)	m
\mathbf{r}_c	position of center of osculating sphere (vector)	m
\mathbf{r}_{exit}	pierce point position (vector)	m
\mathbf{r}_{exit}	position of the point of closest approach between the bent ray-path and the satellite position (vector)	m
c	vacuum speed of light	m/s
D	geometric straight-line length	m
d	total delay	m
d'	straight-line total delay	m
d^z	total zenith delay	m
d_a	along-path delay	m
d_b	bending delay	m
d_g	geometric delay	m
d_h	hydrostatic delay	m

d_h^z	hydrostatic zenith delay	m
d_{nh}	non-hydrostatic delay	m
d_{nh}^z	non-hydrostatic zenith delay	m
G	geometric bent ray-path length	m
g	magnitude of actual gravity	m/s ²
g_c	a nominal, constant, value of gravity	m/s ²
h	ellipsoidal height	m
k	total slant factor	unitless
k_1	refractivity coefficient	K/Pa
k_2	refractivity coefficient	K/Pa
k_2'	derived refractivity coefficient	K/Pa
k_3	refractivity coefficient	K ² /Pa
k_{h+g}	hydrostatic plus geometric slant factor	unitless
k_{nh}	non-hydrostatic slant factor	unitless
L	apparent bent ray-path length	m
L'	apparent straight-line length	m
M_d	molar mass of dry gas	kg/kmol
M_w	molar mass of water vapor	kg/kmol
N	refractivity	unitless
n	index of refraction	unitless
N_d	dry refractivity	unitless
N_h	hydrostatic refractivity	unitless
N_{nh}	non-hydrostatic refractivity	unitless
N_w	wet refractivity	unitless
P	total pressure	Pa
P_d	partial pressure of dry gases	Pa

P_w	partial pressure of water vapor	Pa
q	specific humidity	(unitless)
R	Gaussian radius of curvature of the ellipsoid	m
r	mixing ratio	(unitless)
r	radius or distance to origin	m
R^*	universal gas constant	J/kmol \times K
R_d	dry gas constant	J/kg \times K
R_w	water vapor gas constant	J/kg \times K
T	temperature	K
t	epoch	undefined
T_d	dew point temperature	K
T_v	virtual temperature	K
T_v	virtual temperature	K
v	electromagnetic speed	m/s
W	geopotential	m ² /s ²
X, Y, Z	global Cartesian coordinates	m
x, y, z	local Cartesian coordinates	m
Z	absolute geopotential height	m
Z_{ref}	reference geopotential height	m

Chapter 1

Introduction

1.1 Overview

As the radio signals emanating from GPS¹ satellites propagate through the Earth's electrically neutral (i.e., un-ionized) atmosphere, they suffer *refraction*. Refraction affects the signal velocity in two ways: the signal speed is retarded and the signal direction gets bent. At any point along the signal path the amount of retardation and bending are given by the gradient of the index of refraction — more specifically, by the components of the gradient tangent to the ray and perpendicular to it, respectively.

The effect of refraction on GPS timing measurements is a *delay* compared to what would be measured had the signal propagated in a vacuum. Equivalently, assuming vacuum speed of propagation, the apparent distance measured with GPS is larger than the geometric distance between receiver and satellite. This effect typically reaches 2.3 m for a receiver at sea-level and a satellite in the zenith direction, and is approximately ten times larger for a satellite near the horizon. If not adequately mitigated, that delay corrupts estimates, such as receiver position, obtained from GPS observations.

¹For an introduction to GPS, we recommend Kleusberg and Teunissen [1998]; Leick [2003]; Blewitt [2007].

One way of quantifying the neutral atmosphere radio propagation delay is supposing the signal to be a ray, and tracing that ray along its path, from satellite to receiver, through a model for the atmosphere; we call such a procedure *ray-tracing*. Ray-tracing constitutes our main interest in this work; more specifically, the present work was developed based on the following proposition:

One can find significant discrepancies in ray-traced neutral atmosphere delays, due to reasonable variations in the ray-tracing procedure.²

1.2 Scope

Before we proceed, we would like to set the scope of the phenomenon of interest under this work. In the present section we set the scope broadly; then, in Chapter 2, after introducing the necessary background and vocabulary, we refine that scope.

Even though we mention only GPS and speak of GPS satellites and receivers, this work applies equally well to other satellite navigation constellations (e.g., Galileo, GLONASS) and to other space geodetic techniques. The applicability is greatest for techniques employing similar microwave radiation or electromagnetic signal as GPS, such as in VLBI and InSAR; this work is also applicable, albeit to a smaller extent, to techniques employing radiations other than microwave, such as optical in satellite and lunar laser ranging.

We are interested in the refraction due to the Earth's electrically neutral atmosphere, as opposed to the ionosphere, the ionized portion of the atmosphere. The fact that the ionosphere refracts radio signals of different frequencies to different degrees (i.e., it is dispersive) allows one to use simultaneous measurements at two or more frequencies to eliminate most of this effect. Indeed, that was the very reason for the introduction of a

²Of course, there is always the actual, true value of the delay, as experienced by a GPS receiver, but that is usually unknown; therefore the proposition compares only ray-traced delays among themselves, *not* with respect to the actual, true delay.

second frequency in the design of GPS. The neutral atmosphere, on the other hand, is non-dispersive for radio waves employed in GPS, preventing us from mitigating its effects in GPS in the same way as for the ionosphere.

In this work we disregard the use of atmospheric models in geodesy for purposes other than delay mitigation, such as deformations of the crust due to atmospheric pressure loading and variations in the Earth rotation (as quantified by, e.g., length-of-day, polar motion) due to atmospheric mass redistribution and exchanges with the solid Earth [Salstein et al., 1993].

In this work we also neglect the effect of the atmosphere on the signal amplitude (i.e., attenuation) or polarization. That is because in radiometric space geodetic techniques one rarely employs measurements other than the signal's phase itself (or measurements closely related to it).

Finally, in this work we adopt the framework of classical geometric optics, which allows us to speak about the electromagnetic signal as a ray. It implies that we assume the wavelength of the radiation under study to be negligibly small compared to the extent of the perturbations in the medium of propagation. In other words, we assume that the atmosphere is nearly uniform at spatial distances comparable to that wavelength, roughly speaking. Consequently, under that theory we are unable to account for effects such as diffraction and scattering.

1.3 Contributions from this work

The main research contribution from this work is the identification, classification, and comparison of alternative models for the ray-path and the atmospheric structure employed in ray-tracing (below we briefly introduce those concepts and indicate the sections where a more detailed explanation can be found). Now we discuss the three parts of that main

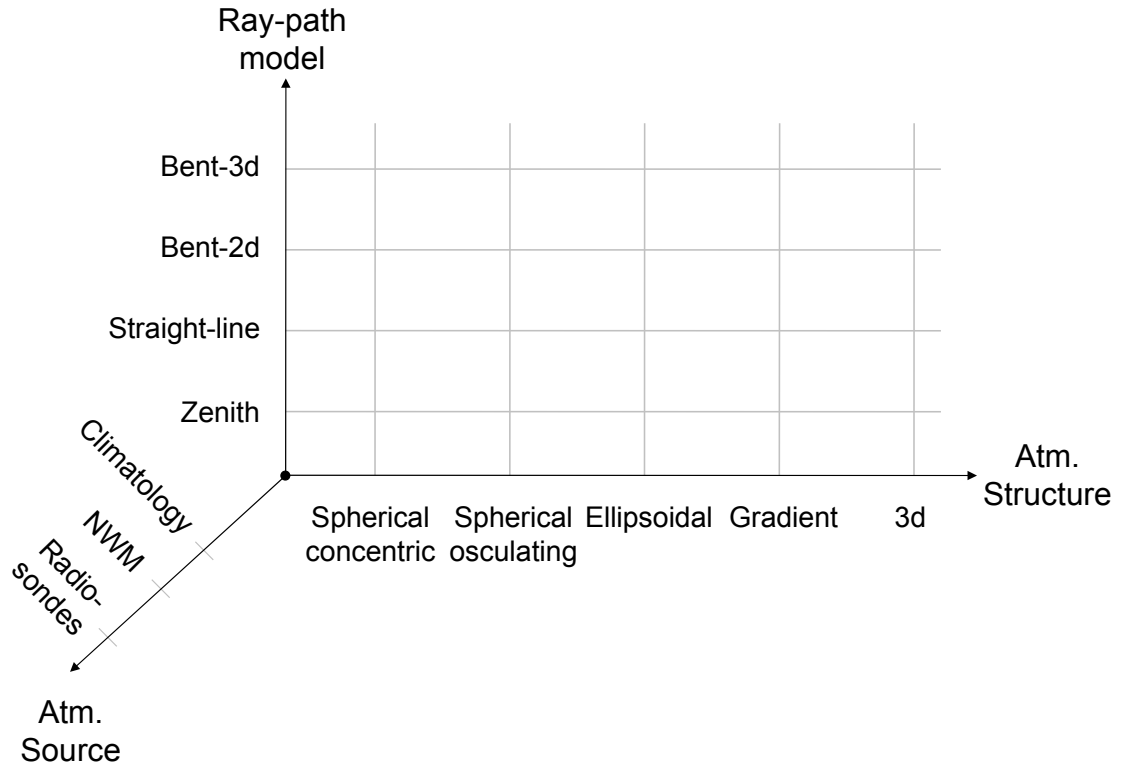


Figure 1.1: Diagram of options available in ray-tracing.

contribution.

The first part is the distinction among the ray-tracing options known as atmospheric source, atmospheric structure, and ray-path model, each represented as an axis in Figure 1.1. It classifies the options available in ray-tracing in separate groups, disentangling aspects that are typically (sometimes arguably conveniently) lumped together (section 2.6). For example, a sentence such as “this raytracer assumes spherical symmetry” actually makes separate statements about the assumed ray-path and the atmospheric structure models. Luckily, the aspects of three groups — ray-path, atmospheric structure, and atmospheric data source — are mutually orthogonal. That orthogonality, shown in Figure 1.1, implies that alternative models compete only within the same group and not with models from a different group.

The second part is the identification of model alternatives within each group (sec-

tion 2.7 and Chapter 3) or along each of the axes in Figure 1.1. The taxonomy above helps one to appreciate the range of options available. In fact, the ordering of the models in a crescendo within each group — from simpler to more complicated or realistic — was especially fruitful in model identification. For example, the ellipsoidal atmospheric structure model was only identified as a result of that particular ordering.

The third part is the experimental comparison of different models (Chapter 5). We quantified the discrepancy in delay between different models (section 5.1) and we also assessed their impact in GPS positioning (sections 5.2 and 5.3).

Besides the main contribution above, a secondary contribution is a classification for the delay mitigation techniques available in GPS (section 4.2) We developed it to support the design of the above-mentioned GPS experiments: that classification helped us in finding prospect applications, in which there is still room for improvement in positioning due to a possibly better delay model. The reader might find that classification helpful in designing how best to deal with the neutral atmosphere delay in his or her own GPS experiments.

Finally — if not original research itself, still useful for *doing* research —, we provide a detailed description of the ray-tracing technique (Appendix I), entailing its numerical and practical aspects. We hope that it might facilitate the introduction of the technique to new researchers interested in studying trans-atmospheric signal propagation — it certainly would have been of help for ourselves when we first started working on this topic.

1.4 Review of previous related work

Ray-tracing has connections to many different subject areas. Among those, we see the present work falling under the umbrella of *geodesy*. More specifically, we see it situated along the thread of developments in mapping functions for radio space geodetic applications. In Appendix VII we quickly skim through the secondary connections. In section 2.8

we delve into that thread, reviewing its historical development; we decided to postpone that review to the end of the background chapter (2) because we need the terminology and definitions only introduced throughout that chapter.

1.5 Structure of this document

We start with Chapter 2, in which we *(i)* introduce the definitions and terminology necessary to discuss the subject in a precise manner and *(ii)* further restrict the scope of this work, which we broadly set in section 1.2. Chapter 3 is the main focus of this work, in which we describe the options available for the ray-path and the atmospheric structure employed in ray-tracing. In Chapter 4 we make a pause in the discussion of the delay itself and discuss its effects and mitigation in GPS positioning. Finally, in Chapter 5 we provide experimental evidence to compare and assess the performance of the different models discussed in Chapter 3. In Chapter 6 we summarize the work developed throughout this document. In parallel to the body of this document, the appendices provide essential details for the reader intending to replicate or build upon the results presented in this work.

Chapter 2

Background

The objective of the present chapter is twofold: (i) to introduce the definitions and terminology necessary to discuss the subject in a precise manner and (ii) to further restrict the scope of this work, something that we already started in section 1.2.

The remainder of this chapter is organized as follows. In section 2.1 we briefly show how to quantify the electromagnetic speed of propagation in the neutral atmosphere. In section 2.2 we define and contrast the various types of electromagnetic propagation delay, resulting from the presence of the atmosphere, as opposed to vacuum. In section 2.3 we introduce the independent variables of the delay and we scan through the complex variation exhibited by the delay over those variables. The aim with section 2.3 is to motivate the need for modeling the delay, if we wish to mitigate its effects on GPS measurements. The manner in which the delay is modeled is essentially to decompose and separate it into different components, based on different rates of change for each component; we explore two modeling techniques in sections 2.4 and 2.5. Among the various delay components defined below we choose to focus on the so-called slant factors, whose models are known as mapping functions. At section 2.6 we delve into the *process* of developing such mapping functions, at which point we further restrict our scope to only one of the three main

tasks involved in that development. The section that follows, 2.7, in principle could be included in Chapter 3, because it is one of the options available in ray-tracing; nevertheless we chose to include it in the present background chapter because we do not see it as part of our contributions. Finally, in section 2.8 we summarize the scope refined throughout the present chapter, and try to justify it in terms of the historical evolution of the process of developing mapping functions.

2.1 Index of refraction

The *index of refraction* n of a medium is defined as:¹

$$n \equiv \frac{c}{v_{\text{phase}}}, \quad (2.1)$$

where c is the speed of light in vacuum and v_{phase} is called the phase velocity (or, more precisely, phase speed) in that medium. There is one formulation of n for each media and type of radiation traveling through it. We call non-dispersive a medium in which the phase velocity is the same for all frequencies in a given band. That is the case (to a very good approximation) of the Earth's neutral atmosphere in the radio band used in GPS. In such a case, the phase velocity equals the group velocity, so we may drop the subscript in v_{phase} , obtaining:

$$v = c/n.$$

It is useful to define also the *refractivity* N :

$$N = 10^6(n - 1), \quad (2.2)$$

which describes the deviation of refraction in the medium, compared to vacuum. The 10^6

¹For a derivation from first principles, we recommend Davis [1986, Sections 1.2.ii–1.2.iv, p. 24–33].

factor is used because the deviation in the Earth’s atmosphere is at most just a few parts per thousand. Both N and n are unitless quantities. N has typical values of 260 on the geoid and decreases with increasing height mainly because of the rapid decrease in density.

For the Earth’s neutral atmosphere and microwave radiation, N is formulated as:

$$N = k_1 \frac{P_d}{T} + k_2 \frac{P_w}{T} + k_3 \frac{P_w}{T^2}, \quad (2.3)$$

where P_d and P_w are, respectively, the partial pressures due to dry gases and water vapor (both in Pa), and T is temperature (in K). The constants k_1 , k_2 , k_3 are coefficients determined empirically in laboratory. In Appendix VI we discuss en passant the discrepancy in delay among different determinations of those coefficients.

In section 2.5.1 we compare two different ways in which total refractivity can be separated into components, and in section 2.5.4 we provide motivation for those separations in terms of their benefits for modeling the corresponding delay components.

2.2 Defining delay

Whereas the delay itself is only defined in section 2.2.3, its definition is streamlined by the prior definition of path lengths (section 2.2.2), which, in their turn, follow naturally once we clarify the consequences of the presence and absence of the atmosphere on the ray propagation (section 2.2.1).

2.2.1 Integrands and integration paths

The delay is defined below in terms of line integrals, running from the satellite to the receiver. Those line integrals have the same independent variable, the along-path distance, denoted ℓ , and the same element of infinitesimal arc-length, $d\ell$. Yet those line integrals

differ in two parts: in their integrand and in their integration path. We employ those two mathematical parts to model the consequences of the presence and absence of the atmosphere on the ray propagation.

In the absence of the atmosphere — i.e., in a vacuum — the integration path is formulated as a straight-line joining receiver and satellite,² and the integrand is unity, corresponding to a ray speed constant and equal to the vacuum speed of light:

$$\int_{\text{straight line}} 1 \, d\ell. \quad (2.4)$$

The presence of the atmosphere — as opposed to a vacuum — brings two changes to the ray velocity: the ray speed is retarded and the ray direction gets bent. Speed retardation is modeled by the integrand $n(\ell)$, where the index of refraction n is allowed to vary as function of the along-path distance ℓ . Direction bending is accounted for by the integration path, which now changes from the straight-line to a model for the actual, bent ray-path:

$$\int_{\text{bent ray-path}} n(\ell) \, d\ell. \quad (2.5)$$

So we notice that the presence and the absence of the atmosphere are modeled in two parts: in the integration path (bent and straight-line) and in the integrand ($n(\ell)$ and unity). Those options make up four different combinations. We have seen two combinations in the two equations above; the remaining two combinations we discover in the next section.³

²See discussion in Appendix I.1.1.

³The numerical evaluation of such integrals is discussed in Appendix I.

2.2.2 Path lengths

The two integration paths defined above (bent and straight-line), along with the two integrands ($n(\ell)$ and unity), make up four different combinations, summarized in Table 2.1. Each combination corresponds to a different path length, denoted by an upper-case letter in Table 2.1, that now we define and discuss.

Table 2.1: Four combinations of integrand and integration path.

		Integration path	
		Atmospheric	Vaccum
Integrand	Atmospheric	L	L'
	Vaccum	G	D

Below, the adjectives “apparent” and “geometric” act as reminders for the type of integrand employed: $n(\ell)$ and unity, respectively. “Apparent” corresponds to the length as it would be experienced by the ray, considering its variable speed along the path (either bent or straight-line). “Geometric” corresponds to the length as measured by a ruler, roughly speaking.

The first combination is called *apparent bent ray-path length* (or simply *apparent ray-path length*):⁴

$$L \equiv \int_{\substack{\text{bent} \\ \text{ray-path}}} n(\ell) \, d\ell, \quad (2.6)$$

Equation (2.6) is a model for what the ray experiences in reality, under the presence of the atmosphere.

The second combination is the *geometric straight-line length* (or simply *geometric distance*):

$$D \equiv \int_{\substack{\text{straight} \\ \text{line}}} 1 \, d\ell. \quad (2.7)$$

⁴Also known as *optical* or *radio path length*, depending on the type of radiation employed.

Equation (2.7) is a model for what the ray would experience ideally, under the absence of the atmosphere.

The first and second combination above (eqs. (2.6) and (2.7)) are extreme cases, in the sense that they correspond to the complete presence or absence of the atmosphere. The two remaining combinations, presented below, are physically impossible. They correspond to fictitious intermediary states of semi-presence and semi-absence of the atmosphere, as you can notice inspecting the mixture of types (Table 2.1) in their integrands and integration paths. However physically implausible those combinations are, they prove useful in defining and understanding the delay, as done in the section that follows.

The third combination is the *geometric bent ray-path length* (or simply *geometric ray-path length*):

$$G \equiv \int_{\text{bent ray-path}} 1 \, d\ell. \quad (2.8)$$

Please note that $G \geq D$, because the straight-line is geometricly the shortest path.

Finally, the fourth combination is the *apparent straight-line length*:

$$L' \equiv \int_{\text{straight line}} n(\ell) \, d\ell. \quad (2.9)$$

A remarkable fact about refraction is that, in spite of the bent ray-path being geometrically longer than the straight-line ($G \geq D$), the bent ray-path is apparently shorter than the straight-line ($L \leq L'$), considering the actual speed of propagation in that medium. Moreover, L is not only smaller but it is also *least* among all possible paths between receiver and satellite. In other words, in a space metricized by the index of refraction n , the actual, bent, ray-path represents the geodesic between two points — analogously to the straight-line being the geodesic in a space of unity metric. That is Fermat's least time

principle.

2.2.3 Total, along-path, and geometric delays

Having defined the path lengths above, now it is straightforward to define the delays. The *total delay*⁵ is:

$$d \equiv L - D = \int_{\text{bent ray-path}} n(\ell) \, d\ell - \int_{\text{straight line}} 1 \, d\ell, \quad (2.10)$$

expressed in units of length, not of time. In words, the total delay is the discrepancy between the apparent ray-path length and the geometric distance. In yet other words, the total delay equals what happens in reality (L) minus what would happen ideally (D). Please note that $d \geq 0$.

We can rewrite the total delay d as:

$$d = d_a + d_g, \quad (2.11)$$

where d_a is the newly defined *along-path delay*:

$$d_a \equiv L - G = \int_{\text{bent ray-path}} n(\ell) \, d\ell - \int_{\text{bent ray-path}} 1 \, d\ell = 10^{-6} \int_{\text{bent ray-path}} N(\ell) \, d\ell, \quad (2.12)$$

which is evaluated along the actual, bent, ray-path, and d_g is the *geometric delay*:

$$d_g \equiv G - D = \int_{\text{bent ray-path}} 1 \, d\ell - \int_{\text{straight line}} 1 \, d\ell, \quad (2.13)$$

which can be stated in words as the discrepancy between the geometric length of the actual,

⁵This is only the total *neutral atmosphere* delay, only one among numerous other delays making up the total delay experienced by a GPS signal, such as ionospheric delays, hardware delays, etc; see, e.g., Leick [2003].

bent, ray-path and the geometric length of the straight-line joining receiver and satellite.

Re-phrasing the statement about refraction in terms of delays instead of path lengths, it is remarkable that the total delay d is least along the actual, bent, ray-path. On the contrary, if we force the ray to follow a straight-line, we will decrease the geometric delay to zero, but only at the expense of increasing the along-path delay by a greater extent, attaining a net increase in total delay as a result.

2.2.4 Geometric delay vs. bending delay

The along-path delay d_a and the geometric delay d_g are not the separate results of each speed retardation and direction bending. Especially, direction bending has an effect not only on d_g but also on d_a , because d_a (eq. (2.12)) is evaluated along the actual, bent, ray-path.

It seems legitimate, therefore, to define a “*bending delay*” d_b , distinct from the geometric delay d_g , encapsulating the net effect of bending (both the increase in geometric delay and the decrease in along-path delay):

$$d_b \equiv L - L' = \int_{\text{bent ray-path}} n(\ell) d\ell - \int_{\text{straight line}} n(\ell) d\ell. \quad (2.14)$$

Consequently, the total delay d is *not* the sum of the along-path delay d_a and the above-defined bending delay d_b , i.e.:

$$d = d_a + d_g \neq d_a + d_b. \quad (2.15)$$

The distinction between the geometric delay d_g and the bending delay d_b is not just a curious theoretical aspect; it is useful in practice, too. Say that we were given a “*straight-*

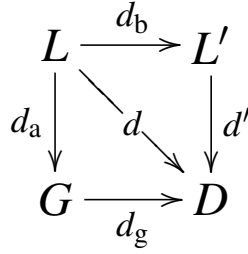


Figure 2.1: Diagram of definitions of delays in terms of path lengths.

line total delay” d' that was computed neglecting ray bending:

$$d' \equiv L' - D = \int_{\text{straight line}} n(\ell) d\ell - \int_{\text{straight line}} 1 d\ell = 10^{-6} \int_{\text{straight line}} N(\ell) d\ell, \quad (2.16)$$

and we wish to correct it to obtain the actual total delay d . The correction to be added, then, is the negative bending delay ($d_b \leq 0$ because $L \leq L'$), *not* the positive geometric delay:

$$d = d' + d_b. \quad (2.17)$$

The corrective bending delay having negative sign reminds us one last time that bending is beneficial, meaning that without bending the total delay would be even greater.

Figure 2.1 summarizes the definitions of delays in terms of path lengths, as an extension to Table 2.1.

2.3 Independent variables and rates of changes

The neutral atmosphere delay, in its most general form, is a function of date and time, receiver position, and satellite direction. That translates into six variables: epoch t , latitude

φ , longitude λ , height h , elevation angle ε^6 , and azimuth α :

$$d = f(t, \varphi, \lambda, h, \varepsilon, \alpha). \quad (2.18)$$

Just like the delay itself, its rates of change are also a function of those same independent variables. In other words, the partial derivatives of the delay depend upon the point at which they are evaluated along *all* independent variables $(t, \varphi, \lambda, h, \varepsilon, \alpha)$, regardless of the one variable v with respect to which the partial is being taken:

$$\frac{\partial d}{\partial v} = f'(t, \varphi, \lambda, h, \varepsilon, \alpha),$$

where v can be substituted by any of $t, \varphi, \lambda, h, \varepsilon, \alpha$.

For example, the delay changes much more rapidly — over any variable v — for a receiver near the ground than aloft, i.e:

$$\text{abs} \left(\frac{\partial d}{\partial v} \Big|_{h=0} \right) \gg \text{abs} \left(\frac{\partial d}{\partial v} \Big|_{h=1 \text{ km}} \right);$$

(which is simply another way of saying that the atmosphere is smoother aloft). As another example, the delay at a lower elevation angle (say, $\varepsilon = 10^\circ$) varies over any variable (not only over ε itself) more rapidly than at zenith. That is because the portion of the ray-path within the atmosphere is greater at a lower elevation angle, therefore putting that delay subject to the influence of a greater portion of the atmosphere.⁷ As one last example, the delay varies more rapidly in a humid region than in an arid region, because the presence of greater humidity in the air, itself highly variable in both time and space, leads to greater variability in the delay. The complexity of the variation of the delay involved in such a

⁶Elevation angle is sometimes replaced by its complement, the zenith angle (sometimes called zenith distance) $z = 90^\circ - \varepsilon$.

⁷The same fact also makes the delay greater at a lower elevation angle, but our point here is variability, not magnitude.

six-dimensional space $(t, \varphi, \lambda, h, \varepsilon, \alpha)$ must start to be apparent.

2.3.1 On the need for modeling to mitigate the delay

If the ultimate goal is to *mitigate* the effect of the delay in GPS positioning, that can be achieved basically in two ways: through prediction or estimation. Here we distinguish between the two on the basis of whether GPS observations are employed to obtain the delay or parameters modeling it. When GPS observations are employed — especially when *redundant* observations are involved in a least-squares sense — we speak of *estimation*. In contrast, when the delay is obtained using only observations and models *external* to GPS, such as pressure measurements, then we speak of *prediction*.⁸

The purpose of modeling is to make the delay easier to predict and to estimate. The motivation is, on one hand, the sheer complexity of the variation exhibited by the delay over its independent variables and, on the other hand, our limited theoretical and observational knowledge about that variation. The technique for modeling is basically to decompose, separate, and approximate the underlying variation of the delay with simpler yet satisfactory models. Below, in sections 2.4 and 2.5, we discuss those techniques.

2.4 Decomposing the delay: zenith delay and slant factors

Elevation angle ε is by far the independent variable over which the delay changes the most. Consequently, it is standard practice to isolate that variation, decomposing the delay as:

$$d = d^z \times k, \tag{2.19}$$

⁸A more elaborate classification is introduced in section 4.2.

where d^z , the so-called *zenith delay*, is defined as:

$$d^z \equiv d(t, \varphi, \lambda, h, \varepsilon = 90^\circ, \alpha), \quad (2.20)$$

for any azimuth α . From eqs. (2.19) and (2.20), the so-called *slant factor*, denoted k , follows as:

$$k \equiv d/d^z. \quad (2.21)$$

Likewise, sometimes we might call d *slant delay*, to emphasize the distinction with the zenith delay, d^z . We therefore postulate d^z the same for all satellites observed at the same epoch (t) from the same receiver (φ, λ, h):

$$d^z = f(t, \varphi, \lambda, h); \quad (2.22)$$

in this way, the variation of the delay with respect to the satellite direction (ε, α) stays confined to the slant factor k :

$$k = f(t, \varphi, \lambda, h, \varepsilon, \alpha). \quad (2.23)$$

We decided to keep the azimuth α in the definition of slant factor; the variation of k over α is not nearly as significant as that over ε , therefore it is not uncommon to see α neglected entirely.

Sometimes we also speak of a *mapping function*, which we define as a model for the variation of the slant factor values with respect to its independent variables. In other words, sometimes we wish to distinguish the slant factor *model* or mapping function (denoted k) from a particular slant factor *value* (denoted k), the latter resulting from the evaluation of

the former at a particular epoch, position, and direction:

$$k = k(t, \varphi, \lambda, h, \varepsilon, \alpha). \quad (2.24)$$

However subtle might seem the distinction between k and k , it will prove necessary to discuss the subject in a precise manner.

The decomposition in eq. (2.19) brings enormous advantages. For prediction (see section 2.3.1) it offers zenith delays that are simpler than slant delays because they depend upon fewer independent variables. Moreover, slant factors can be predicted with less uncertainty than either slant or zenith delays separately. That is a consequence of the fact that d and d^z share the same epoch and receiver position; thus a major part of their individual uncertainty gets canceled out by their significant correlation, upon propagation to k .⁹

For estimation, a predicted slant factor — that can be assumed known beforehand, somehow — turns the zenith delay into a parameter that is *redundant* over the GPS ob-

⁹A few examples. At zenith, the slant delay happens to equal the zenith delay, $d = d^z$, thus the slant factor equals unity, $k = d^z/d^z = 1$, a perfectly known constant whose uncertainty is therefore null:

$$\sigma_k(\varepsilon = 90^\circ) = 0.$$

As the ratio of uncertainty to expected value of the zenith delay approaches zero, $\sigma_{d^z}/\mu_{d^z} \rightarrow 0$, the zenith delay approaches a constant, and the slant factor approaches a normal distribution. The general case requires a little care as, e.g., the ratio of two independent standard normally distributed variates follows a Cauchy distribution, whose variance is not even defined. Under certain conditions, though, it can be shown [Hayya et al., 1975] that the ratio $w = y/x$ of two normally distributed random variables $x \sim N(\mu_x, \sigma_x^2)$, $y \sim N(\mu_y, \sigma_y^2)$, x, y not necessarily independent, is approximately normally distributed:

$$w \sim N(\mu_w = w\mu_x - \mu_y, \sigma_w^2 = \sigma_y^2 - 2w\sigma_x\sigma_y\rho_{xy} + w^2\sigma_x^2),$$

where ρ_{xy} is the correlation coefficient between x and y . The above approximation for the uncertainty σ_w of the ratio w is valid over a wider range of conditions than the typical approximation below, based on a first-order Taylor expansion of the non-linear w , in terms of x and y :

$$\begin{aligned} \sigma_w^2 &= \left(\frac{\partial w}{\partial y}\right)^2 \sigma_y^2 + \left(\frac{\partial w}{\partial x}\right)^2 \sigma_x^2 + 2\frac{\partial w}{\partial y}\frac{\partial w}{\partial x}\sigma_y\sigma_x\rho_{xy}, \\ \left(\frac{\sigma_w}{w}\right)^2 &= \left(\frac{\sigma_y}{y}\right)^2 + \left(\frac{\sigma_x}{x}\right)^2 - 2\frac{\sigma_y}{y}\frac{\sigma_x}{x}\rho_{xy}. \end{aligned}$$

See Hayya et al. [1975] for conditions of applicability of each of the two approximations above.

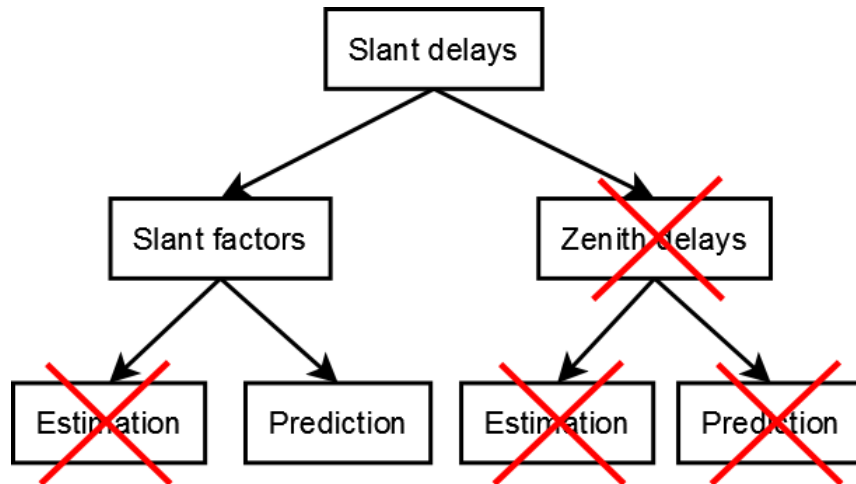


Figure 2.2: Diagram of scope concerning prediction and estimation of zenith delays and slant factors.

servations collected from all satellites visible from the same receiver, thus allowing its estimation. And an estimated zenith delay will always be more accurate than its predicted counterpart.¹⁰ Finally, in estimating a zenith delay, one accounts for the bulk of the slant delay, and such a de-trending of the observation residuals is a pre-requisite for any attempt to estimate slant delays from GPS observations through the stochastic exploitation of their correlations.

2.4.1 Scope: slant factors in, zenith delays out

Our interest in this work concerns slant factors; more specifically, their prediction via ray-tracing (not as much in the attempts to estimate them or, equivalently, to estimate residual slant delays). To maintain a reasonable focus it seems therefore required that we put outside the scope a large body of research dealing with zenith delays. Before moving on, though, let us briefly sketch what was left out in that regard, following Figure 2.2.

The estimation of slant factors from GPS observations is a very hard problem. The

¹⁰As a caveat, we have to admit that zenith delay prediction models are still a valuable product for GPS applications in which the zenith delay estimation is not possible or desired [Schüler, 2006]; see further discussion in section 4.2.

challenge is, on one hand, the separation of other co-existing and poorly accounted for error sources, such as multipath, and on the other hand, to overcome the poor redundancy, poor sky coverage, and poor precision (due to, e.g., antenna gain, atmospheric attenuation) of GPS observations at lower elevation angles. Of course, it is not an entirely unattainable enterprise, it is just hard — after all, we do expect the existence of some correlation across nearby satellite directions that one could take advantage of.

The problem of predicting zenith delays is solved assuming the availability of a reliable atmospheric model, describing the variation of the underlying atmospheric parameters upon which the delay depends [Krueger et al., 2004; Gutman et al., 2003a; Collins, 1999]; see section 2.8 for further discussion. The problem of estimating zenith delays can be attacked splitting it in two, each considering the variation with respect to only epoch or receiver position.

The position part can be ignored in the case of a single stationary receiver, i.e., with $\varphi, \lambda, h = \text{const.}$; in the case of a network there is some redundancy due to correlations across nearby receivers [Stoew and Elgered, 2005; Eresmaa and Jarniven, 2005] that could be exploited to increase the degree of freedom in the GPS estimation; the case of a single moving receiver requires extra care, especially for airborne receivers, as the delay varies very rapidly with height; nevertheless one can still find nearly constant parameters describing a height-varying zenith delay [Colombo, 2006].

The epoch part can be neglected provided the tracking session is short enough. Therefore the simplest model for estimating d^Z corresponds to a short session of a stationary receiver: $d^Z = \text{const.}$ More generally, the variation with respect to epoch can be modeled in either functional or stochastic estimation models. Functionally, one would estimate coefficients of constant, linear, or higher-order polynomial pieces of arbitrary duration (say, 30 min, 1 h). Stochastically, one would leave the zenith delay free to vary from epoch to epoch under a specific variance. Either approach can be tuned in by setting the duration

of the pieces or the variance of the process, respectively [van der Wal, 1995].

2.5 Separating the delay: larger and more stable vs. smaller and more variable components

In the present section we introduce a separation reflecting the contribution of different gas constituents of the Earth's atmosphere. In sections 2.5.1, 2.5.2 and 2.5.3 we introduce that separation for refractivity, slant delays, and slant factors, respectively; then, in section 2.5.4, we provide motivation for that separation.

2.5.1 Refractivity components

Total refractivity eq. (2.3) can be separated as:

$$N = N_d + N_w. \quad (2.25)$$

The first term:

$$N_d \equiv k_1 \frac{P_d}{T}, \quad (2.26)$$

is called *dry refractivity* and the second one:

$$N_w \equiv k_2 \frac{P_w}{T} + k_3 \frac{P_w}{T^2}, \quad (2.27)$$

is called *wet refractivity*, named after the partial pressures employed in each (P_d, P_w).

There exists a different separation for the same total refractivity:

$$N = N_h + N_{nh}, \quad (2.28)$$

where:

$$N_h \equiv k_1 R_d \rho, \quad (2.29)$$

and:

$$N_{nh} \equiv k'_2 \frac{P_w}{T} + k_3 \frac{P_w}{T^2}. \quad (2.30)$$

In eqs. (2.29) and (2.30), ρ is the *total* mass density,¹¹ and $k'_2 \equiv k_2 - k_1 R_d / R_w$ is a derived coefficient.¹² Following Mendes [1999], we call eq. (2.29) *hydrostatic* and eq. (2.30) *non-hydrostatic* refractivity.

Both are a bit of misnomers, not to be taken literally. In 1985, Davis et al. suggested the name “hydrostatic”, fittingly arguing that it would be inaccurate to name “dry” a term that depends on the total density of the mixed humid air. But they were suggesting a name for the zenith delay component, not the corresponding refractivity component. More specifically, they were referring to a model for that zenith delay component, model which makes use of the assumption of hydrostatic equilibrium — hence the name.¹³ Eventually, somewhere in the literature, the name was carried back from the zenith delay model to the corresponding refractivity component, which is not unreasonable. The misnaming becomes clear, though, when we realize that one can validly obtain a “zenith hydrostatic delay” integrating numerically a profile of “hydrostatic refractivity” without ever buying

¹¹ ρ can be expressed in at least two equivalent ways; the first way is in terms of the individual densities:

$$\rho = \rho_d + \rho_w = \frac{P_d}{R_d T} + \frac{P_w}{R_w T},$$

and the second one is in terms of virtual temperature, $T_v \equiv T / (1 - (P_w/P)(1 - M_w/M_d))$ (where $P = P_d + P_w$ is *total* pressure):

$$\rho = \frac{P}{R_d T_v}.$$

¹²For R_d and R_w , please see Appendix IV, p. 214.

¹³That particular zenith hydrostatic delay model, due to Saastamoinen [1972], assumes hydrostatic equilibrium to be able to relate the vertical integral of total density (the only variable in N_h) to the easy-to-measure pressure exerted by the overhead column of mixed humid air. That model was revised in 1986 by Davis and in 2007 by Bosser et al..

into the assumption of hydrostatic equilibrium!¹⁴ Davis [1986] continued to call “wet” the complement to hydrostatic. Even though that is accurate (as eq. (2.30) depends on the partial pressure of water vapor only) it is also ambiguous with respect to the other wet, that complements the dry. Therefore, for the sake of clarity, in 1999, Mendes fittingly suggested the name “non-hydrostatic” instead. But, again, “non-hydrostatic” has nothing to do with a possible deviation from the assumed condition of hydrostatic equilibrium; it is just the complement to hydrostatic in making up total refractivity or the along-path delay.

2.5.2 Slant delay components

Recall that the total (slant) delay, as given by eq. (2.11), is divided in the along-path delay d_a and the geometric delay d_g :

$$d = d_a + d_g.$$

Following the separation introduced for refractivity (section 2.5.1), we can introduce a corresponding separation in the along-path delay:

$$d_a = d_h + d_{nh}, \quad (2.31)$$

where:

$$d_h = 10^{-6} \int_{\text{bent ray-path}} N_h(\ell) d\ell \quad (2.32)$$

and:

$$d_{nh} = 10^{-6} \int_{\text{bent ray-path}} N_{nh}(\ell) d\ell. \quad (2.33)$$

¹⁴We have to admit that such a profile is seldom available. In other words, the assumption of hydrostatic equilibrium is pervasive. Most state-of-art atmospheric models have that assumption built-in. Even the height values provided by radiosondes are not measured, rather derived under that same assumption, given the measurements of pressure, temperature, and humidity. Exceptions are, e.g., radiosondes equipped with *both* GPS and pressure sensors, and non-hydrostatic atmospheric models.

Likewise, we call d_h and d_{nh} *hydrostatic* and *non-hydrostatic delay*, respectively. A similar separation in delay also exists for *dry/wet*, of course. Care should be exercised when comparing results from different authors, to assure that they all follow the same separation, $d_h + d_{nh}$ or $d_d + d_w$.

Please notice that the separation in refractivity stays confined to the integrand and does *not* carry over to the integration path (borrowing from the terminology discussed in section 2.2.1). In other words, the actual, bent, ray-path is unique and defined by the field of *total* index of refraction. In yet other words, the ray-path along which the integrals eqs. (2.32) and (2.33) are evaluated is one and the same. The alternative — two different ray-paths, each defined by the hydrostatic and non-hydrostatic indices of refraction — besides being physically wrong, would bring the highly undesirable consequence of the hydrostatic and non-hydrostatic delays not summing up to the along-path delay.

The fact that we do not separate the contribution of hydrostatic and non-hydrostatic components to the definition of the ray-path implies that we are unable to separate the geometric delay:

$$d_g = \int_{\substack{\text{bent} \\ \text{ray-path}}} 1 \, d\ell - \int_{\substack{\text{straight} \\ \text{line}}} 1 \, d\ell,$$

into its hydrostatic and non-hydrostatic components. Therefore we express the total slant delay no further detailed than:

$$d = d_h + d_{nh} + d_g. \tag{2.34}$$

Please see caveat below (p. 27) regarding practice in mapping functions, though.

2.5.3 Zenith delay and slant factor components

The extension to zenith delays is trivial, as the zenith delay is a type of slant delay (whose direction happens to be that of zenith):

$$d_h^Z \equiv d_h(t, \varphi, \lambda, h, \varepsilon = 90^\circ, \alpha),$$

$$d_{nh}^Z \equiv d_{nh}(t, \varphi, \lambda, h, \varepsilon = 90^\circ, \alpha),$$

for any azimuth α , following the definition of zenith delay (eq. (2.20)). The extension to slant factors would be straightforward too (e.g., $k_{nh} \equiv d_{nh}/d_{nh}^Z$), were it not for the fact that the geometric delay d_g is zero at zenith,¹⁵ what prevents us from defining a slant factor for geometric delay (k_g).

Yet we wish to maintain consistency with the definition of total delay: the sum of delay components should always make up the same total delay. One first possible approach is to define an additive geometric delay, i.e.:

$$d = k_h d_h^Z + k_{nh} d_{nh}^Z + d_g; \quad (2.35)$$

a second approach is to define a compound factor, k_{h+g} , accounting for both d_h and d_g :

$$k_{h+g} \equiv (d_h + d_g) / (d_h^Z + d_g^Z) = (d_h + d_g) / d_h^Z, \quad (2.36)$$

$$k_{nh} \equiv d_{nh} / d_{nh}^Z, \quad (2.37)$$

(where d_g^Z happens to be zero), so that, from eqs. (2.11), (2.19) and (2.31):

$$d = k_{h+g} (d_h^Z + d_g^Z) + k_{nh} d_{nh}^Z = k_{h+g} d_h^Z + k_{nh} d_{nh}^Z; \quad (2.38)$$

¹⁵The zenith geometric delay is exactly zero in the most common model for the ray-path (bent-2d); for the more general bent-3d ray-path model, albeit it is not exactly zero, it is so to very good approximation; see Appendix I.3.4.1 for details.

a third approach is to define a similar compound factor, k_{nh+g} , involving instead the non-hydrostatic delay, and leave the hydrostatic delay alone in k_h ; the resulting expressions are analogous to eq. (2.36).

The second approach (k_{h+g}, k_{nh}) is favoured in the recent literature [Niell, 1996; Boehm et al., 2006b]. The reasons are as follows. The second approach is preferred over the third one (that of k_h, k_{nh+g}) because the slant non-hydrostatic delay, free of geometric delay (but still under the influence of ray bending, as per section 2.2.4), is more similar to the delay determined by water vapor radiometers [Davis, 1986]. The second approach is preferred over the first one (that of k_h, k_{nh} and an additive d_g) because it requires one fewer quantity to be provided for end-users (i.e., two quantities in the second approach instead of three quantities in the first approach).

To unclutter the notation, it is common practice to call the combination of hydrostatic and geometric delays simply “hydrostatic delay”, in which case we would have:

$$d = k_h d_h^Z + k_{nh} d_{nh}^Z, \quad (2.39)$$

where the assignment $d_h \leftarrow (d_h + d_g)$ is taking place implicitly — without this caveat, eq. (2.39) would be slightly misleading.

We could define similar zenith delay and slant factor components following the dry/wet separation.

2.5.4 Motivation

Dry gases constitute the major part of the atmosphere; furthermore, they are homogeneously mixed. Water vapor, in contrast, is not only a minor constituent, but is also highly variable both in space and in time, tending to exist in lumps in the atmosphere [AGU, 1995]. Therefore the delay component due to dry gases is larger and more stable than the

smaller yet far more variable component due to water vapor.

The dry/wet separation represents those different characteristics directly. The hydrostatic/non-hydrostatic separation, in contrast, combines the whole contribution from dry gases with a small part of the contribution from water vapor to provide a new component dependent on the *total* mass density of the mixed humid air. That new component (hydrostatic) has a tremendous advantage, namely, the fact that its corresponding zenith delay can be reliably predicted in terms of the pressure exerted by the overhead column of air, dismissing (under the assumption of hydrostatic equilibrium) the need for knowing the actual atmospheric vertical distribution. That advantage in terms of prediction, in its turn, leads to an advantage in terms of estimation: one needs then to estimate only the non-hydrostatic delay — otherwise, one would have a little trouble trying to estimate both simultaneously, as they are very highly correlated. Yet another advantage is that the non-hydrostatic component, contrary to the wet one, corresponds to the delay determined by water vapor radiometers [Davis et al., 1985]. We speculate, though, that it should have a minor disadvantage, namely, that the hydrostatic component is not as stable as the dry one; in other words, we speculate that the water vapor included in the hydrostatic delay should manifest itself as increased variability, as compared to the purely dry delay.

2.6 Developing a mapping function

In section 2.4 we have seen that what makes possible the estimation of zenith delays from GPS observations is the assumption that the mapping function is somehow known beforehand.¹⁶ But how do we fulfill that assumption? In other words, how do we develop a mapping function? That is the subject of the present section.

¹⁶Whereas the development of mapping functions in earlier contributions was generally part of an effort aiming at predicting total slant delays, most later contributions in the literature focused on offering mapping functions to aid in the estimation of zenith delays.

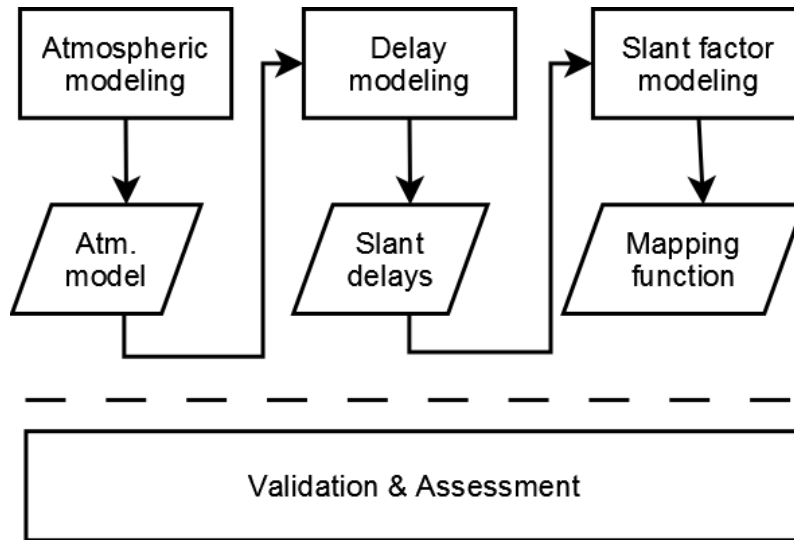


Figure 2.3: Diagram of development of mapping functions.

Figure 2.3 depicts the tasks and intermediary products (as boxes and parallelograms, respectively) involved in the process of developing a mapping function. In the sub-sections that follow, we discuss each task individually. We also take on the opportunity to refine the scope of this work, putting certain tasks or parts of tasks outside the scope, which can only be done after we at least define all the tasks.

Before we proceed, though, let us make the caveat that we describe the process of developing mapping functions as carried out in this work. But the way described is not the only one possible. In section 2.8 we discuss how that process evolved historically, and why we believe that the version of that process adopted here represents the current state-of-art.

2.6.1 Atmospheric modeling

On one hand, the delay depends on the state of the atmosphere along the entire ray-path, from satellite to receiver. On the other hand, it is impractical to measure that state directly. Therefore, the development of a mapping function requires the creation or adoption of

an atmospheric model (leftmost box in Figure 2.3). Pragmatically, an atmospheric model can be interpreted as a function, having as argument or input an epoch t and a position φ, λ, h (corresponding to a point along the ray-path), and returning as output a description of the atmospheric state at that point. Among the numerous quantitative descriptions of the atmospheric state, we are interested specifically in knowing pressure P , temperature T , and humidity (expressed in any of numerous ways; see Appendix IV) or, equivalently, refractivity N or index of refraction n .

We distinguish between two aspects making up a more comprehensive atmospheric model: structure and source. Atmospheric structure is a label for the arrangement of the iso-indicials¹⁷ in a given atmospheric model. For example, if the index of refraction is constant along spherical shells, we speak of a spherical atmospheric structure. Atmospheric source, on the other hand, denotes the origin of the data making up the atmospheric model (such as radiosondes) or the purpose in generating that atmospheric model (such as weather or climate modeling).

Atmospheric sources and structures are further discussed in sections 2.7 and 3.2, respectively. In the rest of this work, beyond the present background chapter, we put atmospheric source outside the scope; atmospheric structure remains inside. The reasons for putting source outside are discussed in section 2.8 — in a nutshell, we see it as an atmospheric modeler’s job, not a geodesist’s.

In the subsequent tasks (sections 2.6.2 and 2.6.3), all part of the process of developing a mapping function, we take for granted the availability of an atmospheric model; i.e., we assume that it is possible to obtain the atmospheric parameters at any epoch and position needed.

¹⁷An iso-indicial is a surface of constant refractivity or index of refraction.

2.6.2 Delay modeling

The objective of the present task, delay modeling, is to provide values of slant delays that would be observed at a particular epoch, receiver position, and satellite direction, given an atmospheric model resulting from the previous task, atmospheric modeling (section 2.6.1). There are a few key ingredients in the present task. One of the ingredients is a refractivity formulation; as seen in section 2.1, it relates a particular atmospheric state to the electromagnetic speed of propagation. Another ingredient is a ray-path model; it dictates the path, from satellite to receiver, of a ray arriving at a particular epoch, position, and direction. As we shall see later, in Chapter 3, there is a whole range of ray-path models, from simpler to more complicated and realistic.

In the rest of this work, beyond the present background chapter, we keep ray-path models inside the scope; the refractivity formulation is put outside, because even though it might still require further research (Appendix VI), it can be pursued independently from the aspects elected to form the scope of this work.

2.6.3 Slant factor modeling

The previous task, delay modeling (section 2.6.2), results in individual slant delays at several particular epochs, positions, and directions. The corresponding slant factors can be obtained as per their definition, eq. (2.19). The objective of the present task, slant factor modeling, hence, is to provide end-users with a convenient, usually compact, representation of slant factors.

The present task presupposes that it is impractical or undesirable to have users interacting directly with the delay model and obtaining slant factors as per their definition, eq. (2.19). Otherwise, there would be no need for an explicit mapping function — in fact, such an approach was put into practice by several authors, as will be discussed in

section 2.8.

In the rest of this work, beyond the present background chapter, we adopt the approach of an implicit mapping function, putting slant factor modeling (but not slant factors, of course) entirely outside the scope.

Current practice

Before moving on, let us briefly sketch how slant factor modeling is performed currently. Please keep in mind the independent variables of slant factors: $t, \varphi, \lambda, h, \varepsilon, \alpha$. Hydrostatic (including geometric, see p. 27) and non-hydrostatic components are treated separately.

The variation with respect to satellite direction (elevation angle ε and azimuth α) is accounted for first, individually at each different epoch and receiver position. The variation with respect to elevation angle is modeled with Marini's [1972] continued fraction expansion in $1/\sin \varepsilon$, normalized to yield unity at zenith, as given by Herring [1992]:

$$k(\varepsilon) = \frac{1 + \frac{a}{1 + \frac{b}{1 + \frac{c}{\dots}}}}{\sin \varepsilon + \frac{a}{\sin \varepsilon + \frac{b}{\sin \varepsilon + \frac{c}{\dots}}}}. \quad (2.40)$$

The result is made of coefficients a, b, c, \dots for each epoch and receiver position.¹⁸ The variation with respect to azimuth is, most of the time, neglected, and sometimes accounted for with a single main direction of asymmetry:

$$k(\varepsilon, \alpha) = k_0(\varepsilon) + \delta k(\varepsilon, \alpha), \quad (2.41)$$

where k_0 and δk accounts for the azimuthally symmetric and asymmetric parts, respectively; the former is as in eq. (2.40); the latter is [Davis et al., 1993]:

$$\delta k(\varepsilon, \alpha) = k_0(\varepsilon) \cot \varepsilon (Z_N \cos \alpha + Z_E \sin \alpha), \quad (2.42)$$

where $Z_N \equiv \Xi_N/d^z$ and $Z_E \equiv \Xi_E/d^z$ are, respectively, north and east coefficients describing the direction and magnitude of asymmetry exhibited by slant factors, defined in terms of the asymmetry exhibited by slant delays, Ξ_N and Ξ_E , normalized by the zenith delay d^z .

The variation with respect to receiver position (latitude φ , longitude λ , height h) is accounted for secondly, with geographical maps describing the variation of some of Marini's elevation coefficients and Davis et al.'s azimuth coefficients vs. latitude and longitude, at a specific height (usually constant zero or the variable ground's height). The variation with respect to height is accounted for as a rate of change of the coefficients, valid only near the specific height at which the maps above are given. (Some coefficients are assumed con-

¹⁸Yan and Ping [1995] offer an alternative expression for the elevation-angle dependence, also in the form of a continued fraction, related to the complementary error function. Does Yan and Ping's [1995] fit experimental ray-tracing results better (i.e., with fewer terms in the expansion) than Marini's [1972], *given the same atmospheric model*? Mendes [1999] did include in his extensive comparisons a mapping function based on Yan and Ping's [1995], but it is not clear how much of its performance is due to the mathematical expression per se or the atmospheric model upon which the mapping function was fit. Yan et al. [2002, Fig. 2–3] presents results showing general better convergency in very low elevation angles, with worse convergency for a few geographical locations. Definitive conclusion would require a larger dataset and also homogeneity in the ray-tracing results upon which one is to fit Marini's [1972] and Yan and Ping's [1995] expressions. Bear in mind, though, that “if the observation elevation angles are constrained above 5°, the choice of the continued fraction of the mapping function is not a critical problem” [Yan et al., 2002, p. 573].

stant.) Those maps are represented either analytically, as an expansion in surface spherical harmonics [Boehm et al., 2006a], or numerically, discretized in grids (Kouba [2008] or Niell [1996] for higher and lower resolution grids, respectively).

The variation with respect to epoch (t) is accounted for last, as a time series for each grid node making up the map discretization or for each coefficient in the map's expansion in spherical harmonics. The time series, in their turn, can be represented numerically, as a discretization in snapshots at specific epochs (say, four times daily — Boehm et al. [2006b], Kouba [2008]), or analytically, fitting an empirical model (say, seasonal — Niell [1996], Boehm et al. [2006a]).

At last, let us list a few open problems regarding the modeling of the variation of slant factors with respect to its independent variables:

- variation with respect to azimuth: beyond a single main direction of azimuthal asymmetry (e.g., secondary directions [Seko et al., 2004]; arbitrary spherical harmonics [Böhm and Schuh, 2001]);
- variation with respect to elevation angle: negative elevation angles;
- variation with respect to any independent variable at positions far from the ground (e.g., airplanes, radiosondes, dropsondes, the Space Shuttle) — see Thessin [2005] and Yan et al. [2002, Fig. 1].

2.6.4 Validation and assessment

The present task, validation and assessment, runs in parallel to all previous tasks (sections 2.6.1 to 2.6.3) in the process of developing a mapping function. Here, in section 2.6.4, we distinguish between the tasks of *validation* and *assessment*. The former is a measure of *internal* consistency; it seeks to guarantee that an implementation conforms to its specification, i.e., to its intended design. The latter compares the performance

of a given model to that of competing models; the crux in such comparisons is to find a reliable *benchmark*, preferably external to the models being compared.

Both tasks apply to each individual model: atmospheric (source and structure), delay (refractivity and ray-path), and slant factor. In other words, validation and assessment can be done not only with the final product but also with the intermediary ones. Let us give a few examples illustrating the description above.

Validation usually does its job through “sanity checks”, i.e., checks on whether a given quantity indeed behaves as expected. Trivial examples are if pressure is always positive and decreases with height; if total refractivity is the same regardless of how it is separated ($N_d + N_w$ or $N_h + N_{nh}$); if the total delay along the actual ray-path is smaller than that along a forced straight-line ($d < d'$).

Assessment can be done directly, in terms of the modeled quantity itself. An example is comparing modeled pressure, temperature, and humidity to their observed counterparts (provided by, e.g., radiosondes). It can also be done indirectly, in terms of the impact on the performance of a particular use of that model. An example is assessing the atmospheric model by comparing the resulting non-hydrostatic slant delay predictions to those measured by a water-vapor radiometer. The ultimate impact assessment involves employing a given model in GPS positioning and checking either or both observation residuals and estimated parameters (e.g., estimated receiver position).

We validate the models developed in this work in Appendix I.4. We also assess their performance in GPS positioning, in sections 5.2 and 5.3.

2.7 Atmospheric sources

In section 2.6.1 we introduced the distinction between two aspects making up a more comprehensive atmospheric model: structure and source. Then we put atmospheric source out-

side the scope, and kept atmospheric structure inside. In the present section, atmospheric sources, we give a minimal background on that subject, before moving on; atmospheric structures are further discussed in section 3.2.

2.7.1 Radiosondes

The source constituted of radiosondes [Dabberdt et al., 2002] is the closest to reality, as they are (mostly — see next paragraph) actual observations. Usually the sensors are released from the ground hanging off a balloon, but sometimes they are dropped during special flights hanging off a special parachute, in which case they receive the special name *dropsondes*. Their measurements have large vertical resolution (Figure 2.4) but have very small horizontal and time resolution: they are usually released from approximately 800 sites worldwide, twice daily.

It is interesting to notice that radiosondes, as used for predicting slant delays or slant factors, are not purely observations — they do require some modeling. First, height is not measured but rather derived, under the assumption of hydrostatic equilibrium, from the actual measurements of pressure, temperature, and humidity. Second, it requires interpolation among the vertically scattered measurements, for which one has to postulate a (linear, quadratic, log-linear, hydrostatic, etc.) model. Third, it requires interpolation for horizontal positions other than the launching sites' [Ghoddousi-Fard and Dare, 2007; Ifadis and Savvaidis, 2001], and even more dangerous, extrapolation beyond the coverage area of the ensemble of launching sites. Fourth, it requires interpolation and/or extrapolation at epochs other than the release epoch (which, by the way, is just a nominal value — the balloon takes up to a few hours to finish its course). Fifth, it requires a supplementary atmospheric model after the balloon bursts at its maximum height.

Radiosondes remain as the most important observation type in atmospheric modeling, since they provide vertical profiles throughout the troposphere; in contrast, vertically

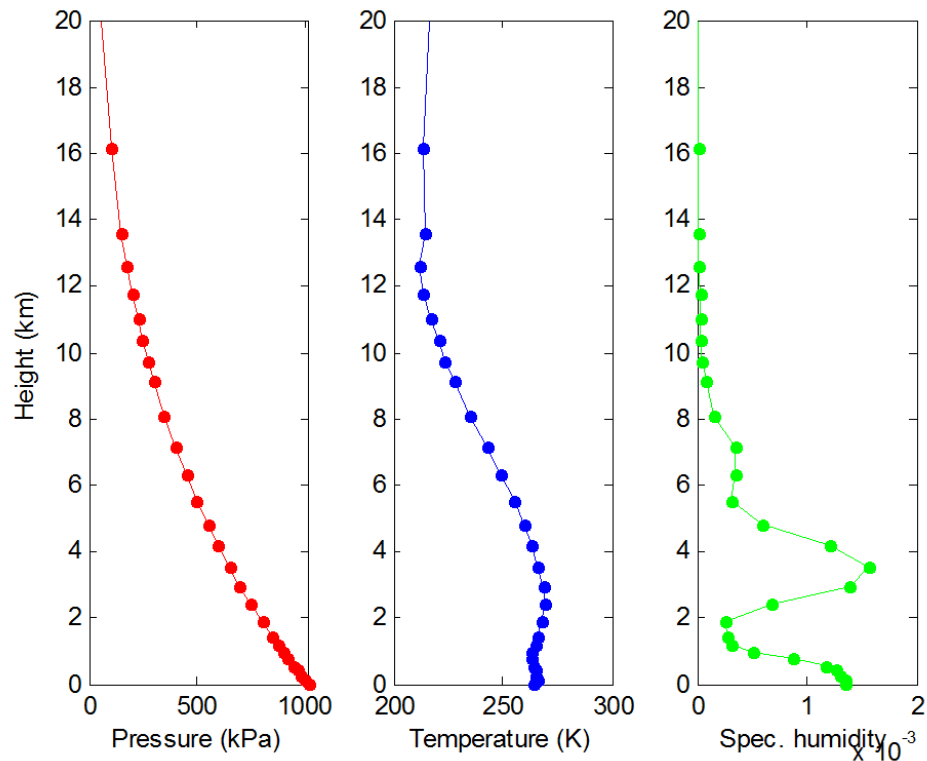


Figure 2.4: Sample radiosonde: vertical profile collected at a single site and epoch.

integrated observations, such as GPS-estimated zenith delay or satellite images, cannot be inverted meaningfully without careful constraints (as in any tomography problem, in general).

2.7.2 Numerical weather models

Another source is that of numerical weather models (NWM, hereafter) or, more precisely, their output [Lorenz, 2002; Buizza, 2002; Golding, 2002]; see Figure 2.5. When that model output refers to a future epoch, one speaks of prediction; otherwise, analysis; which is the reason why we prefer to speak in general of a numerical weather *model*, regardless of whether it refers to a future or past epoch. Its output is a representation of the atmospheric fluid at specific instantaneous epochs. The representation includes a 3d field snapshot for

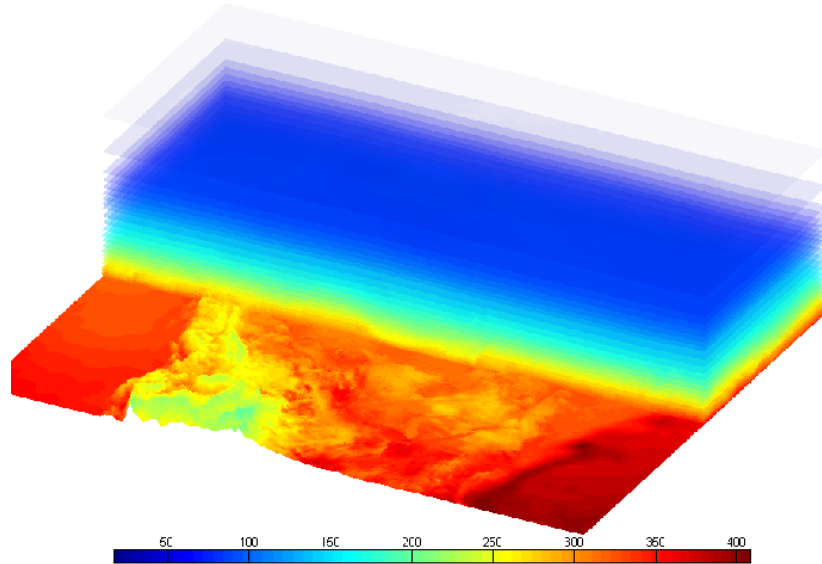


Figure 2.5: Sample numerical weather model: refractivity field (unitless) over North America on Aug 16, 2004 at 22:45 UTC, as given by the Canadian Regional Model [Côté et al., 1998] (vertical scale exaggerated 100x).

each atmospheric parameter — pressure, temperature, humidity, amongst many others. It usually takes the form of either an expansion in (volumetric) spherical harmonics or a discretization on a grid. Its main advantage over radiosondes is its much larger horizontal and time resolution. On the other side of the same coin there is the disadvantage of its large size (in bytes) requiring above-average digital equipment for storage and computations.¹⁹

¹⁹For use in ray-tracing, a NWM usually needs to be supplemented with a climatology, because a NWM does not extend vertically as high as necessary.

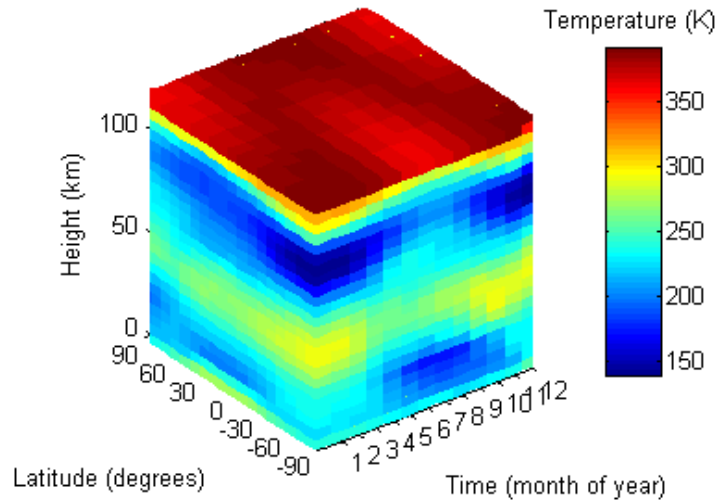


Figure 2.6: Sample climatology: temperature, as given by the CIRA86 model [Chandra et al., 1990; Fleming et al., 1988].

2.7.3 Climatologies

Yet another atmospheric source is that of climate models [Hartmann, 2002]; see Figure 2.6. As the saying goes, climate is what we expect and weather is what we get.²⁰ Consequently, a climatology will never be as good a source as a NWM, concerning the actual atmospheric conditions affecting GPS observations. Compared to radiosondes, climatologies are less accurate, but radiosondes have a rather sparse area of coverage. The amount of discrepancy expected and the accuracy requirements for the intended GPS application will dictate which atmospheric source is better suited. Yet one particular climatology might be better than others. (For a good discussion, please see Mendes [1999, p. 48], Johnson et al. [2002], and Thessin [2005].)

²⁰The definition of climate can be worded in a few slightly different ways: climate comprises the slowly varying components of the weather [Glickman, 2000]; climate, more rigorously, is the statistical description of the weather, in terms of its expected value (e.g., average) and variability (e.g., standard deviation, empirical probability distribution functions, etc.) [IPCC, 2007, Annex I]. In any case, a critical concept in any of those definitions is the *averaging period*, as the so-defined climate will be different for different averaging periods. The classical value is 30 years, as defined by the World Meteorological Organization [IPCC, 2007, Annex I].

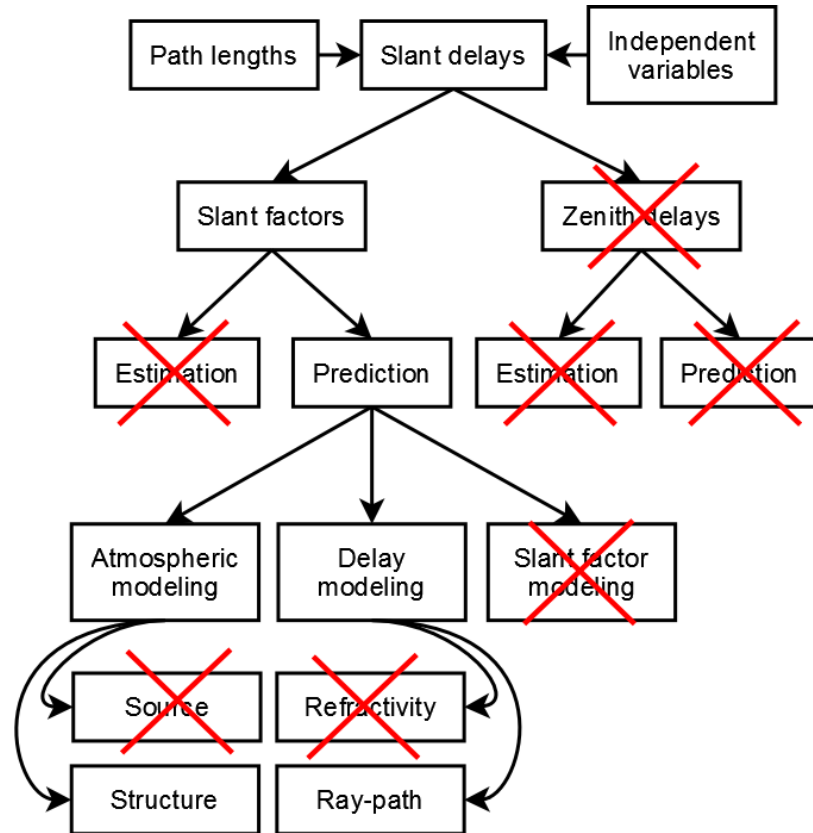


Figure 2.7: Diagram of scope of this work.

2.8 Our scope in the historical context

The objective of the present section is to justify and motivate the scope for this work; the scope is summarized in the next paragraph; the justification and motivation is presented in the rest of the present section.

The scope of this work was broadly set in section 1.2: it is the study of the neutral atmosphere delay in space geodetic observations, under the theory of classical geometric optics. That scope was further refined throughout the background chapter, after we introduced the necessary concepts and terminology. The refined scope is summarized in Figure 2.7, which we now briefly review. From top to bottom, we see how slant delays come about from the definition of delay (section 2.2), on one hand, and from its indepen-

dent variables and corresponding rates of change (section 2.3), on the other hand. Then we split slant delays into slant factors and zenith delays (section 2.4). At that point we refined the scope (section 2.4.1). Zenith delays were entirely put outside, both their estimation and their prediction. As for slant factors, their estimation (or, equivalently, the estimation of slant delays) from GPS observations was put outside the scope, while their prediction (based on models and observations external to GPS) was kept inside. Then we proposed a process for developing mapping functions (section 2.6), involving three modeling tasks — atmospheric, delay, and slant factor — each of which is made of more elementary models; for example, the atmospheric model is made of a source and a structure; a delay model is made of a refractivity formulation and a ray-path model. The introduction of that process of developing mapping functions allowed us to further refine our scope: we elected to focus, beyond the present background chapter, on atmospheric structure and ray-path model, effectively putting outside the scope atmospheric source, refractivity formulation, and all of slant factor modeling.

The decision about the scope taken at section 2.4.1 needs not be justified, as it is based on a decomposition (eq. (2.19)) that is standard and commonplace. It is the decision concerning the process of developing mapping functions, taken at sections 2.6.1 to 2.6.3, that we see the need to justify and motivate now. That need exists because, even though that process is not innovative (in the sense that other researchers have pursued it before), we understand that it is not consensual either (in the sense that it is still common to see researchers pursuing alternative approaches).

The justification and motivation for the scope is presented below in the form of a list of individual remarks. Our intention with such a form was to facilitate for the reader to agree or disagree with only parts of our whole reasoning or argument. At best, we try to convince the reader that the version of the process of developing mapping functions adopted in this work represents the current state-of-art; at least, we provide some food-

for-thought concerning the historical evolution of that process.²¹

(1) We can choose the primary modeling realm: The variation of slant factors with respect to their independent variables — epoch, receiver position, and satellite direction — can be accounted for either directly in the slant factor model itself (section 2.6.3) or indirectly in the underlying atmospheric model (section 2.6.1).

In the past, in the absence of a better atmospheric model, one would obtain samples of slant factors at scattered epochs, positions, and directions, and then try to fit an explicit model to them. Consequently we say that the slant factor model was the primary modeling realm. The availability of a satisfactory atmospheric model already implicitly accounting for that variation causes a shift in the primary modeling realm, from the slant factor model to the atmospheric model.

Examples of the latter practice are all the contributions based on numerical weather models, e.g., Boehm et al. [2006b]; an example of the previous practice is Mendes [1999].

(2) The atmosphere is the best primary modeling realm: Given the choice in remark (1), it is better to choose to account for the variation of slant factors in terms of the underlying atmospheric state, than directly in terms of the slant factor itself. (Even though we admit that it is more complicated.)

The first reason is that the atmospheric state is a more fundamental quantity than the highly processed slant factor. Therefore there are certain behaviours exhibited by slant factors that can only be explained in terms of the underlying atmospheric state (e.g., passing of weather fronts, occurrence of droughts or hurricanes, etc.).

The second reason is that there are many more research areas other than geodesy and astrometry interested in general atmospheric quantities such as pressure, temperature, and

²¹Herring [1992]; Niell [1996]; Sovers et al. [1998]; Mendes [1999]; Yan et al. [2002] present good reviews of past contributions to the development of mapping functions, up to their publication dates.

humidity than in the neutral atmosphere delay. Consequently there are more resources invested in understanding the variation of general atmospheric quantities than in understanding the delay. It seems sensible, therefore, to leverage on those resources already invested instead of trying to model delay or slant factors from scratch.

(3) Atmospheric modeling is better done by experts: Atmospheric modeling is better done by expert atmospheric scientists than by, say, geodesists.

At hindsight, the sheer complexity of modern atmospheric models warrants that it would be futile to expect non-experts to be able to improve upon them. In other words, we advocate that a mapping function developer adopt an atmospheric model designed, implemented, validated, and operated by expert atmospheric scientists, instead of naïvely trying to develop a possibly better atmospheric model of their own. Otherwise, one must admit that he or she would be entering in the business of atmospheric modeling, and therefore should judge his or her atmospheric models by the stricter standards of the expert atmospheric science. Also the painstaking job of running and quality-controlling such systems on a daily basis has costs that any single user community, such as the geodetic one, probably could not afford alone.

In summary: we see such a division of labor as liberating geodesists and astrometrists from the burden of atmospheric modeling. Our rationale is that, to build a good atmospheric model, is not an easy task!

(4) NWM are the state-of-art in atmospheric modeling: The shift in the modeling realm discussed in remark (2) was made possible by the availability of models providing a continuous, higher resolution representation of the atmospheric state.

A continuous model is able to readily provide pressure, temperature, and humidity at any reasonable epoch and position, even unsampled ones. That is in contrast with *partitioned* atmospheric models, valid here and there at scattered locations. Examples of

continuous models are NWM and climatologies; radiosonde-based atmospheric models usually start as partitioned models, upon which one can build a continuous model, after postulating an interpolating procedure [Leandro et al., 2006b].

By resolution we mean not simply the spacing between nodes making up the model discretization, but the scale of atmospheric features that it accounts for, either explicitly through the model dynamics or implicitly through the model physics [COMET-UCAR, 1999]. Typical temporal resolutions are a few minutes internally (for numerical integration of the atmospheric equations) and a few hours externally (for model output). Typical spatial resolution is a few tens of kilometers along horizontal coordinates and variable along the vertical (from tens of metres near the ground to kilometres far aloft).

The improvement in resolution is rather astonishing when we compare NWM to climatologies previously used for mapping functions. Temporal and spatial resolution were one month and a thousand kilometres, respectively. Furthermore, time was modeled modulo one year (i.e., only intra-annual variations considered, inter-annual variations neglected); spatial variation was modeled only along latitude (variations with longitude were neglected).

In summary: the current state-of-art in atmospheric modeling is embodied in the so-called numerical weather models; we refer to Thorpe [2005] for an excellent overview.

(5) How NWM came to be used as an atmospheric source for mapping functions:

Zenith delays, being simpler than slant delays, were naturally tried before [Bevis et al., 1996; Vedel et al., 2001; Jensen, 2005; Cove, 2005]. One could obtain zenith delays from scratch, integrating a vertical profile of refractivity, or more easily and less error-prone, plugging in surface proxies (such as pressure or integrated water vapor — see remarks (9) and (11)), into an appropriate zenith delay model (e.g., Saastamoinen's [1972]).

From zenith to slant delays there is a huge jump in complexity in the way that NWM

can be employed, at which point the remarkable versatility of NWM is revealed. One simpler way is to extract upper-air proxies (see remarks (9) and (11)) from the NWM [Niell, 2003]. Another way is to extract a single vertical profile from the NWM and build a spherical atmosphere around it, just like radiosonde profiles have been used previously [Boehm et al., 2006b; Rocken et al., 2001]. Yet another way is to extract two profiles, each describing mean and gradient conditions, corresponding to zero-th and first order terms of a Taylor expansion around a base point [Hulley and Pavlis, 2007; Boehm and Schuh, 2007; MacMillan and Ma, 1998]. A more elaborate way would be fitting a higher-order polynomial over a desired vicinity and using that polynomial as a surrogate for the full 3d NWM [Chen and Herring, 1997]. The ultimate way is to use the full 3d NWM in its plenitude, without simplifications [Hobiger et al., 2008a; Nievinski et al., 2005; Pany et al., 2001].

It seems fitting to conclude that any attempt to directly predict slant delays necessarily implies a level of distrust towards the state-of-art mapping functions — otherwise one could predict instead only simpler zenith delays and then couple those zenith delays with a state-of-art mapping function. In other words, we should recognize that, when predicting slant delays directly, one is entering in the business of mapping functions.

(6) Future prospects for the usage of NWM in mapping functions: It is sensible to say that NWM are here to stay. They have recently made their way into the IERS Conventions [McCarthy and Petit, 2004] (more specifically, in its updated version, currently under development²²), therefore having their usage endorsed in the geodetic/astrometric community. The trend into the foreseeable future is for NWM to improve continuously, pushed by advances in computer processing, by increased data availability, and by the societal needs for better weather and climate forecasts. That will continue to happen even

²²<http://tai.bipm.org/iers/convupdt/convupdt.html>

without additional effort from the geodetic community.

At last, fortunately the future advances in NWM will be useful not only for geodetic datasets collected in the future, but also for those collected in the past. That is possible thanks to the so-called re-analysis projects, that run current state-of-art NWM retroactively with historical data, going back the last few decades [Uppala et al., 2005; Kalnay et al., 1996].

(7) Slant factor models: from primary modeling realm to surrogate optional representation: The shift in the primary modeling realm (discussed in remark (1)), from the slant factor model to the atmospheric model, implies a diminishment in the importance of the former.

In an extreme case, the slant factor model becomes absent entirely; the end-user interacts directly with the delay model and obtains slant factors as per its very definition (eq. (2.21)), in terms of slant and zenith delays. In a mild case, the slant factor model is retained, but its role is relegated to that of a surrogate for the variation of slant factors already implicitly modeled by the atmospheric model.

One might argue that a slant factor model is no longer a driving force in the development of mapping functions, because even if it is wasteful in terms of computer storage and processing, it would not affect the accuracy of the results. Nevertheless, that task is still relevant, because even though computing is relatively cheap nowadays, there will always be constrained computing environments (such as on-board miniaturized receivers) in need of more convenient representations (i.e., less demanding in terms of computing resources). This concern applies especially to real-time applications [Collins, 1999; Krueger et al., 2004], like those covered by GNSS navigation augmentation systems, such as the American WAAS or the European EGOS.

Examples of the extreme case are provided by Rocken et al. [2001]; Pany et al. [2001];

Hobiger et al. [2008b];²³ examples of the mild case are given Boehm et al. [2006a,b].

(8) Slant factor modeling as a data compression problem: Remark (7) implies that the task of slant factor modeling is simplified from what it used to be in the past — it is no longer atmospheric modeling in disguise.

We find it fruitful to think of slant factor modeling as a data compression problem [Salomon, 2007]. I.e., one wishes to find a suitable basis upon which to expand the slant factor values, so that the one can still obtain a satisfactory approximation after truncating the expansion at a few leading terms. Surely knowledge about the atmospheric models provides insights about good bases, e.g., Marini’s [1972] expansion in $1/\sin(\varepsilon)$; see also Yan and Ping [1995].

(9) Proxy vs. intrinsic variables: In the light of limited observational or theoretical prior knowledge about the atmosphere, one might be unable to accurately account for the variation of slant factors with respect to their independent variables (epoch, receiver position, satellite direction) — let’s call those *intrinsic variables*. A solution often sought in the past would be to delegate that modeling to what we call *proxy variables*. Those proxies provide the link to the actual conditions at time and location of usage. The solution involving proxy variables requires the augmentation of the model parametrization:

$$k(t, \varphi, \lambda, h, \varepsilon, \alpha) = k(t, \varphi, \lambda, h, \varepsilon, \alpha, P^*, T^*, \dots); \quad (2.43)$$

²³We must mention that all of those aim at providing slant delays, not slant factors; we cite them here nevertheless because one can always obtain slant factors from slant delays.

where each of the proxies P^*, T^*, \dots (say, surface pressure, surface temperature, temperature lapse rate, etc.) depends, in their turn, on time and location:

$$P^* = f(t, \varphi, \lambda, h)$$

$$T^* = f(t, \varphi, \lambda, h)$$

⋮

(10) Ready-to-use vs. provisional vs. on-demand slant factor models: If a given model requires as input only its intrinsic independent variables (as defined in remark (9)), we call it a *ready-to-use model*.²⁴ In contrast, if the model requires as input additional proxy variables, we call it a *provisional model*, as those additional variables provide a sort of calibration to the model prior to its use. Finally, if the model is developed automatically from scratch (i.e., from slant and zenith delays, and those delays are obtained by ray-tracing in an atmospheric model) every time a user requests slant factor values, then we call it an *on-demand model*.²⁵

Examples of provisional models are Ifadis [1986] and most of those studied in Mendes [1999]; of ready-to-use, Niell [1996] and Boehm et al. [2006a]; and of on-demand, Pany et al. [2001]; Hobiger et al. [2008b]; Rocken et al. [2001]; Iwabuchi et al. [2003]. It seems that the turning point between provisional and ready-to-use models was the work of Niell [1996]; between ready-to-use and on-demand, we could cite Rocken et al. [2001].

(11) Surface vs. upper-air proxies: To understand the history of mapping functions, it helps to distinguish between surface-only and upper-air proxy variables. In general, provisional models based on surface proxy variables have limited success, because surface conditions are not necessarily representative of the conditions aloft. For example, the

²⁴Our “ready-to-use” seems equivalent to “empirical” in Boehm et al. [2006a].

²⁵Our “on-demand” seems equivalent to “dynamic” in Iwabuchi et al. [2003].

amplitude of diurnal temperature changes is much larger near the ground than aloft, where, in contrast, changes happen mostly seasonally. As Niell [1996] pointed out, we need to distinguish between warm winter days and cold summer days. In spite of that non-representativeness, the recourse to surface proxies is perfectly understandable: such values are relatively common and easy to obtain. In contrast, provisional models based on upper-air proxy variables have great potential, not always fulfilled, though. The reason for that we discuss in the next remark, (12).

(12) Values for proxy variables: measured in situ vs. provided by auxiliary models:

The obvious way to obtain values for the proxy variables (remark (9)) is to measure them in situ (e.g., with a radiosonde for upper-air proxies). Unfortunately, measurement devices are not always available, especially for upper-air variables. The consequence of that unavailability is that users of provisional mapping functions (see remark (10)) would employ nominal values for those proxy variables, achieving thus only sub-optimal results.

In the lack of in situ observations, a better alternative for obtaining values for proxy variables is to provide an auxiliary model. In that way, one would couple the provisional model requiring the proxy variables with an auxiliary model for those proxies.²⁶ Coupled models are especially fruitful because they foster collaboration among researchers — in other words, it needs not to be the same researcher developing both the provisional and auxiliary models. On the down side, with the information needed to build an auxiliary model for upper-air proxies, one could probably achieve better results dropping the provisional model altogether and building a stand-alone ready-to-use model directly.

Examples of auxiliary models are the numerous ones developed by Mendes [1999] for previously developed provisional models.

²⁶The auxiliary model, in its turn, could be provisional itself (if it related upper-air proxies to more commonly available surface measurements) or it could be ready-to-use (if it depended only on intrinsic variables). The latter case is more useful, because the resulting coupled model becomes ready-to-use as a whole.

(13) Hypothetical vs. statistical models: How does one establish the relationship between proxy variables and the resulting delay or slant factors?

In early contributions (e.g., Saastamoinen [1972]), the relationship would be established through an analytical *hypothetical* model (which is not to say that it would be unreasonable). It would be based on the expected mean behaviour of the atmosphere, such as the prevalent stratification in layers of constant or linear (in height) temperature, and hydrostatic equilibrium for pressure.

Another way is through statistical methods (e.g., Mendes [1999]), in which one builds a large set of results covering a variety of conditions, and then seeks empirical relationships. Ideally, the dataset would cover uniformly the range of independent variables intended for use (e.g., from pole to pole, if it is intended as a global model).

(14) Analytical vs. numerical methods: The development of a mapping function can be carried out, up to varying degrees, in analytical or numerical form. We recognize a historical evolution from purely analytical to mixed analytical/numerical methods. That shift occurred more markedly in the realm of atmospheric models, which we discuss in the next paragraph. In the realm of slant factor models, a similar shift is underway (Kouba [2008] is a prime example), but analytical models are still preferred and we expect them to remain preferred in niches such as computing-constrained environments (see remark (7)).

First let us recall that numerical solutions are usually computationally more expensive than equivalent analytical ones. Second, cheap, powerful computing resources is a modern luxury that was simply not available decades ago. Therefore it is easy to understand why convenient closed-form analytical solutions were preferred in early efforts: one would build the atmospheric model aiming at a fast and easy way to compute the delay later. There is a price to pay for that convenience, though: the underlying models (atmosphere, delay, slant factor — see section 2.6) end up entangled in a single composite model that is

hard to modify. Consequently, in early contributions (e.g., Saastamoinen [1972]) it might not be clear-cut where each task in process of developing a mapping function (section 2.6) ends and another one begins.

As modern computing matured and alleviated the burden of more demanding numerical solutions, analytical models would be replaced with profiles, tables, and grids discretizing the atmospheric state. It allows the original underlying models (atmospheric, delay, slant factor) to be kept intact and separate. That is advantageous for a number of reasons: in validation, when something goes wrong it is easier to identify the culprit; in assessment, it is easier to isolate the individual contributions of each of the underlying models and improve only one at a time. Nonetheless, the existence of both analytical and numerical solutions allows their cross-validation, as they should agree within a certain tolerance.

In this context, we call the computation of slant factor values by a (predominantly) numerical procedure by the name *ray-tracing*.

Chapter 3

Ray-tracing options

In Chapter 2 (p. 51) we have defined ray-tracing as the computation of slant factors by a (predominantly) numerical procedure. Whereas the essential how-to knowledge is presented in detail in Appendix I, here we would like to bring awareness to the options available in ray-tracing.

The options form three groups, each depicted as one axis in Figure 3.1. The axes are mutually orthogonal, meaning that options compete only within the same group and not among different groups. Along each axis, the options are ordered from the simplest, closer to the origin, to more rigorous, farther away from the origin. The simpler options approximate the more rigorous ones sufficiently well over common and useful special cases, as we shall see later.

The main feature of that diagram is the recognition of the orthogonality between ray-path and atmospheric models. It disentangles two aspects that are typically lumped together. For example, a remark such as “*this raytracer assumes spherical symmetry*” encompasses distinct concepts: it makes separate statements about each ray-path and atmospheric models. Certainly, there are some combinations that provide a better match between the two, but we propose that such a choice of options be made and communicated

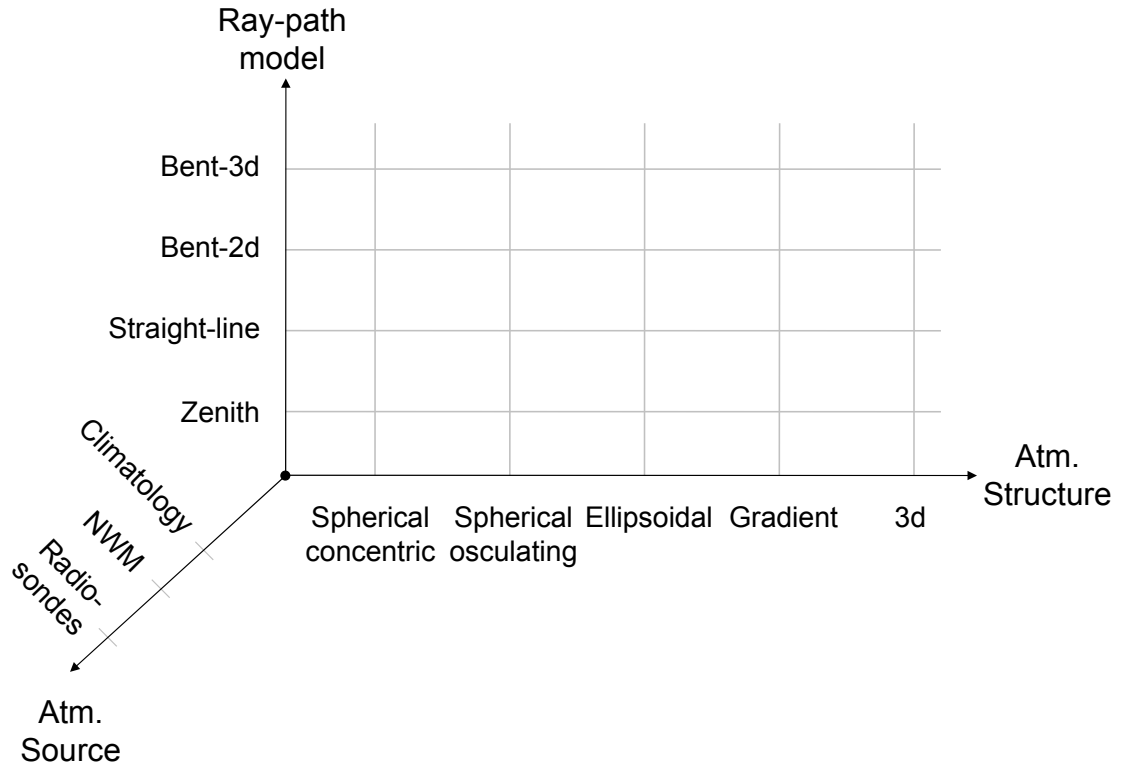


Figure 3.1: Diagram of options available in ray-tracing.

more explicitly. Furthermore, the possibility of mixing and matching different options is a good opportunity to validate more complicated models comparing them to simpler ones at the cases for which we expect no discrepancy; we explore that in Appendix I.4.

A secondary feature of Figure 3.1 is the orthogonality in the aspects of source and structure, making up a more comprehensive atmospheric model. Recalling briefly from section 2.6.1, atmospheric structure is a label for the arrangement of the iso-indicial surfaces, while source denotes the origin or purpose in generating an atmospheric model. It is interesting that we can take any source and arbitrarily postulate a structure for it.

In the next sections of the present chapter we detail the options in regard to ray-path and atmospheric structure; atmospheric source was already discussed in section 2.7.

Hereafter, we make use of the following concepts: geodetic coordinates (latitude φ ,

longitude λ , ellipsoidal height h), global Cartesian coordinates (roughly speaking¹, having origin at the center of mass of the Earth, X axis pointing towards the intersection of the Greenwich meridian with the equator, Z axis pointing towards the north pole, and Y axis in such a way that it completes a right-handed system), and local Cartesian coordinates (x axis pointing towards the east direction, y axis pointing towards the north direction, and z axis pointing upwards, along the ellipsoidal normal). The function denoted as $\overset{XYZ}{\underset{\phi\lambda h}{\left(\right)}}$ converts from geodetic to global Cartesian coordinates. Position vectors such as \mathbf{r} are assumed expressed in global Cartesian coordinates, i.e., $\mathbf{r} = X\hat{\mathbf{i}} + Y\hat{\mathbf{j}} + Z\hat{\mathbf{k}}$; when we need position vectors in local Cartesian coordinates we will indicate that as $\mathbf{r}^{xyz} = x\hat{\mathbf{i}} + y\hat{\mathbf{j}} + z\hat{\mathbf{k}}$. Elevation angle ε is always reckoned from the ellipsoidal horizon or, equivalently, zenith angles are always reckoned from the ellipsoidal normal.

3.1 Ray-path models

In this section we discuss different model for the ray-path, more specifically its shape, scale, position, and orientation.

First of all we must recognize that when we speak about the electromagnetic signal as a ray we have implicitly adopted the framework of classical geometric optics. It implies that we assume the wavelength of the radiation under study to be negligibly small compared to the extent of the perturbations in the medium of propagation. In other words, we assume that the atmosphere is nearly uniform at spatial distances comparable to that wavelength, roughly speaking. Consequently, under that theory we are unable to account for effects such as diffraction and scattering.

The rest of this section is organized as follows. The ray-path, in its general form, is a 3d curve, as described in section 3.1.1. In propagation media with special structure

¹The precise definition is laid down in McCarthy and Petit [2004].

(section 3.2), though, it assumes simpler shapes. For example, in a spherical atmosphere (sections 3.2.1 and 3.2.2), the ray-path degenerates into a plane curve, i.e., it is contained on a plane, as shown in section 3.1.2. As another example, if the ray's tangent direction always coincides with the gradient of refraction in a medium, then the ray degenerates further, into a straight-line, as described in section 3.1.4; the straight-line case, albeit not always rigorously valid, is a good approximation for most of the sky, as we discuss in section 3.1.3.

3.1.1 Bent-3d

The most general ray-path model is defined by the fundamental equation of classical geometric optics, the *Eikonal equation* [Born and Wolf, 1999]:

$$\frac{d}{d\ell} \left(n \frac{d\mathbf{r}}{d\ell} \right) = \nabla n. \quad (3.1)$$

This differential equation describes the *change* in the ray position vector \mathbf{r} along the ray-path. We see that it depends critically on the propagation medium: n is the (scalar) field of index of refraction and ∇n is its gradient (vector) field.² The actual position $\mathbf{r}(\ell)$ at any given distance ℓ along the ray-path depends on both the propagation medium and a set of conditions. A common choice of initial conditions are an initial position and an initial direction, in which case the final position is determined from eq. (3.1). A common choice of boundary conditions are initial and final positions, in which case the initial direction is determined from eq. (3.1). The Eikonal is discussed in detail in Appendix I.

²By field we mean that n and ∇n associate a value (scalar and vector, respectively) to any position in space, not only to positions along the ray-path.

3.1.2 Bent-2d

In an atmosphere with spherical structure the gradient of refractivity ∇n always points to the center of the sphere:

$$\hat{\nabla}n = \frac{\nabla n}{|\nabla n|} = -\hat{r}' = -\frac{\mathbf{r}'}{|\mathbf{r}'|}, \quad (3.2)$$

where $\mathbf{r}' = \mathbf{r} + \mathbf{r}_c$ is the position vector with respect to the center of the (possibly osculating) sphere (section 3.2). The consequence of a null horizontal component in the gradient of refractivity is the absence of any out-of-plane bending in the ray-path. Another way of stating the same is saying that in an atmosphere exhibiting spherical structure the ray-path is rigorously a plane curve (hence the suffix “2d” in the name).

This model is embodied in the widely used Bouguer’s formula [Born and Wolf, 1999, p. 131]:

$$nr' \sin z = \text{const.}, \quad (3.3)$$

where $z = 90^\circ - \varepsilon$ is the zenith angle and r' is the distance to the center of the (possibly osculating) sphere. Following Young [2006, p. 99–100] we can recast eq. (3.3) in the form of a differential equation:

$$dr' = -\tan(z) dn/n. \quad (3.4)$$

Equation (3.4) can be interpreted as a generalization of Snell’s law:

$$n \sin \theta = (n + dn) \sin(\theta + d\theta), \quad (3.5)$$

where θ is the ray’s angle of incidence with respect to the normal at the interface of two media.

In the actual Earth’s atmosphere, the horizontal component, though not exactly null, is always much smaller than its vertical counterpart, making the bent-2d a satisfactory approximation most of the time. In section 5.1 we describe experiments quantifying that

discrepancy.

3.1.3 Straight-line

This model neglects any bending in the ray-path. It is defined:

$$\mathbf{r} = \mathbf{r}_0 + \ell \hat{\mathbf{s}}_0, \quad (3.6)$$

where $\mathbf{r}_0 \stackrel{XYZ}{\leftarrow}_{\varphi\lambda h} (\varphi_0, \lambda_0, h_0)$ is the receiver position vector; $\hat{\mathbf{s}}_0$ is the direction (a unit vector) from receiver to satellite; ℓ is the usual along-path distance, except that now we postulate it to be equal to zero at the receiver and increasingly positive towards the satellite (equivalently, we could define the ray-path from the satellite to the receiver, instead).

The fact that the along-path distance incidentally equals the straight-line distance brings a tremendous advantage over preceding models (bent-3d, bent-2d), namely, that the entire (postulated straight-line) ray-path is known in advance. That is in contrast with models that allow for bending, in which the ray-path needs to be discovered step-by-step or iteratively. Although the straight-line may seem a crude approximation, it is satisfactory for a large portion of the sky, from zenith down to 30° in elevation angle.³ It was very common in early contributions [Saastamoinen, 1972; Hopfield, 1969].

3.1.4 Zenith

This is a special case of both straight-line and bent-2d models, when the ray direction happens to coincide with the local vertical direction:

$$\mathbf{r} \stackrel{XYZ}{\leftarrow}_{\varphi\lambda h} (\varphi_0, \lambda_0, h_0 + \ell). \quad (3.7)$$

³We exploit that fact in Appendix I.5.3.1 to develop a fast solver for bent ray-paths based on the perturbation of a straight-line ray-path.

It is rigorously valid in an atmosphere with spherical osculating or ellipsoidal structure, and an excellent approximation in any structure. It is a very convenient form when one is interested in zenith delays only, because it allows one to simplify the path integrals (section 2.2) to ordinary integrals having limits expressed in terms of height.

3.2 Atmospheric structures

In the present section we discuss the following atmospheric structures (from simplest to more elaborate): spherical concentric (section 3.2.1), spherical osculating (section 3.2.2), ellipsoidal (section 3.2.3), gradient (section 3.2.4), and 3d (section 3.2.5). Just like for the ray-path models, under atmospheric structure we have a general case (3d) and the remaining ones are simplifications of that general case, warranted by common and useful special cases.

The first four structures are profile-based, meaning that all the data they need is stored in the form of a single vertical profile, assumed the same for any horizontal position. In other words, vertical variation is modeled and horizontal variation is neglected. That is a good approximation to the Earth’s atmosphere, in which, e.g., pressure decreases exponentially with increasing height. The differences among those four profile-based atmospheric structures lie in their subtle differences in the definition of horizontal coordinates, or, equivalently, of vertical coordinates. We make those definitions more clear and precise in the remainder of the present section.

Throughout this section, we will be using the form:

$$v = f(\dots),$$

where v stands for any of pressure, temperature, or humidity, and “...” will be replaced

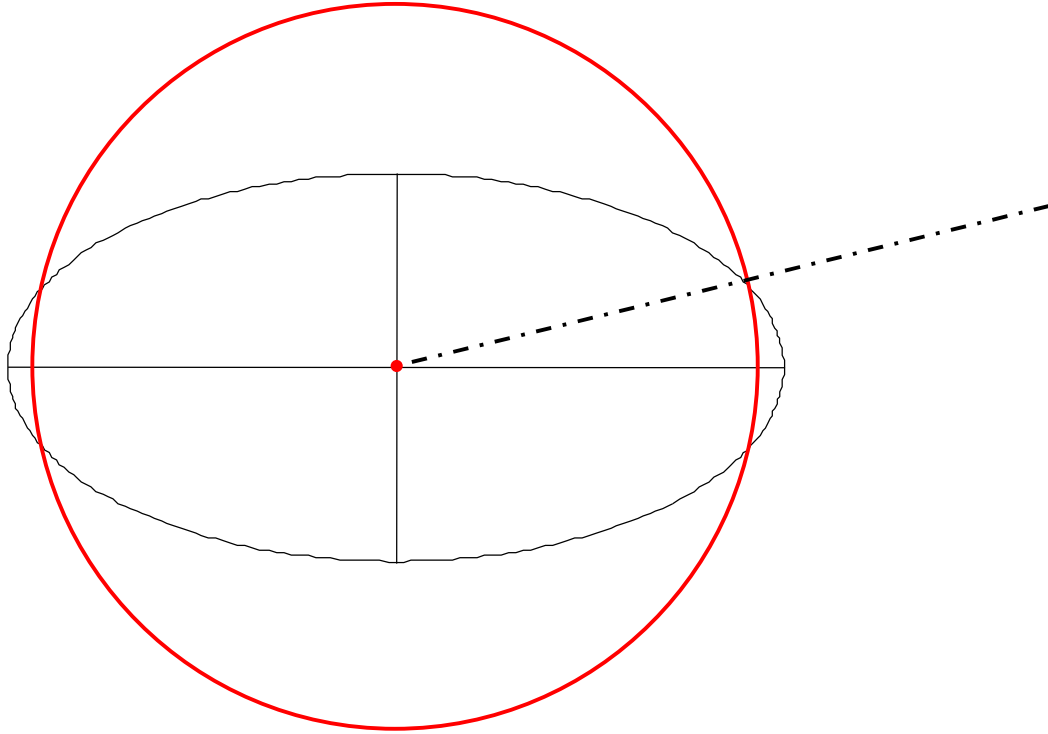


Figure 3.2: Sample spherical concentric iso-indicial (in red).

by the variables upon which the atmospheric parameters depend in each structure.⁴

3.2.1 Spherical concentric

In this model, illustrated in Figure 3.2, the vertical direction is defined coinciding with the radial direction of a sphere centered at the Earth's center of mass (or, more precisely, centered at the origin of the global Cartesian coordinate system (X, Y, Z) , [McCarthy and Petit, 2004]):

$$v = f(r), \quad (3.8)$$

where $r = |\mathbf{r}|$.

⁴“Iso-indicial”, used below, is a surface of constant refractivity or index of refraction.

3.2.2 Spherical osculating

This model takes the vertical direction as the radial direction of a sphere that osculates the ellipsoid at a particular base point. In such a case:

$$v = f(r'), \quad (3.9)$$

where $r' = |\mathbf{r}'|$ is the length of an eccentric position vector \mathbf{r}' :

$$\mathbf{r}' \equiv \mathbf{r} - \mathbf{r}_c,$$

The center of the osculating sphere, \mathbf{r}_c , is defined as:⁵

$$\mathbf{r}_c \stackrel{XYZ}{\leftarrow}_{\varphi\lambda h} (\varphi_0, \lambda_0, h = -R);$$

the latitude φ_0 and longitude λ_0 specify the base point at which the sphere's radial direction coincides with the ellipsoid's normal direction; $R = \sqrt{MN}$ is the Gaussian radius of curvature;⁶ M, N are the radii of curvature along the meridian and prime vertical, respectively; the negative sign in $h = -R$ implies that the center of the osculating sphere is located downward with respect to the ellipsoid's surface (Figures 3.3 and 3.4).

3.2.3 Ellipsoidal

This last profile-based atmospheric structure (Figure 3.5) takes the vertical direction as the ellipsoidal normal:

$$v = f(h), \quad (3.10)$$

⁵Please notice that $\mathbf{r}_c \stackrel{XYZ}{\leftarrow}_{\varphi\lambda h} (\varphi_0, \lambda_0, -R) \neq \mathbf{r}_0 \stackrel{XYZ}{\leftarrow}_{\varphi\lambda h} (\varphi_0, \lambda_0, h_0)$.

⁶“An intrinsic property of the surface independent of the coordinate system used to describe that surface” [Weisstein, 2008].

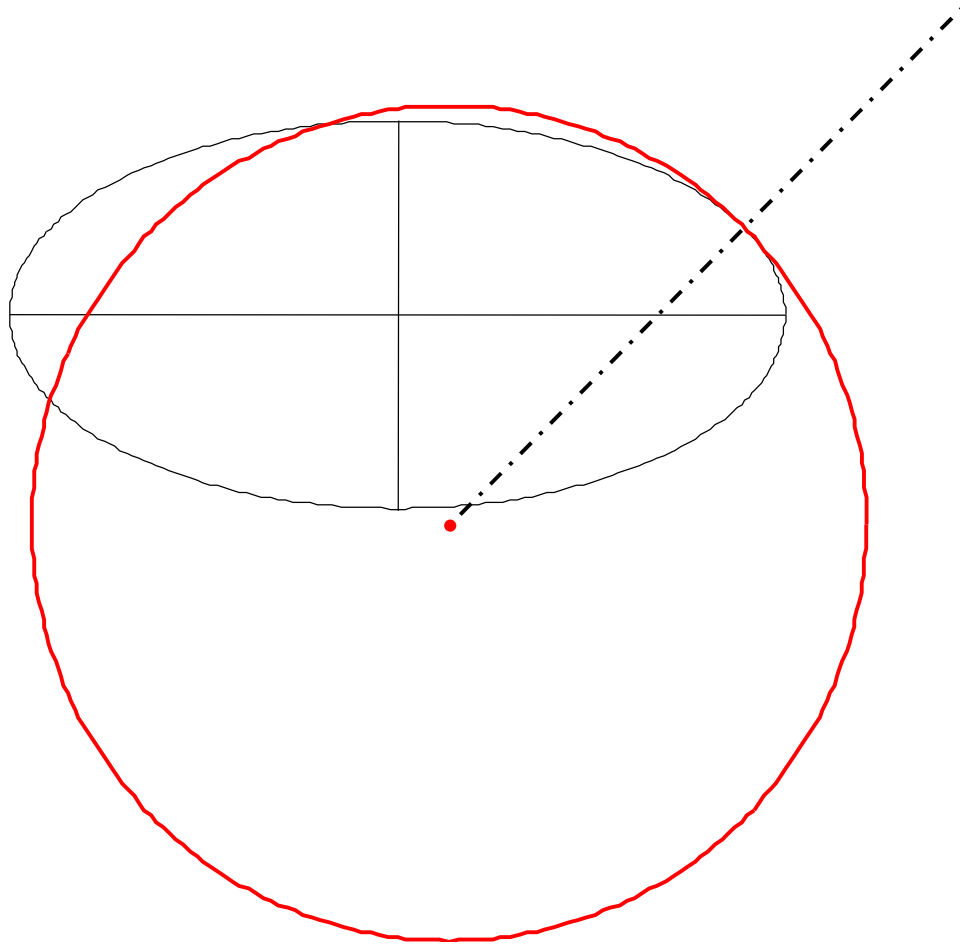


Figure 3.3: Sample spherical osculating iso-indicial (in red).

which obviously yields a simple expression in terms of ellipsoidal height h .

3.2.4 Gradient

The next approximation in terms of atmospheric structure may be described as a double-profile based model: we take the single profile exactly as defined for the ellipsoidal atmosphere (denoted here v_0) and augment it with a second profile ($\nabla_{\text{H}}v_0$). Both profiles, v_0 and $\nabla_{\text{H}}v_0$, refer to a base location, which is specified by φ_0, λ_0 (hence the subscript 0). The scalar profile v_0 describes the mean conditions, whereas the vector profile $\nabla_{\text{H}}v_0$ describes the horizontal gradient, both at the base location. We define the gradient atmo-

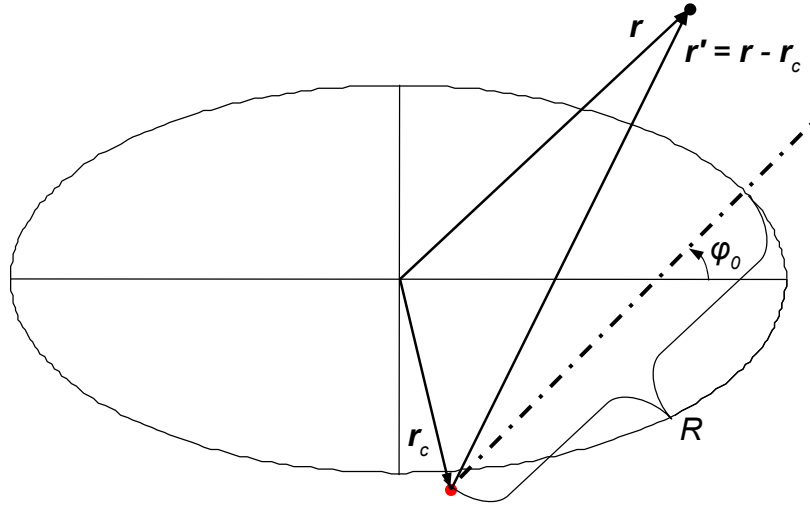


Figure 3.4: Angles, distances, and vectors involved in a spherical osculating atmosphere.

sphere in such a way that there are possibly different values of mean v_0 and horizontal gradient $\nabla_{\text{H}}v_0$ for each different height h , but at any particular height h , v_0 and $\nabla_{\text{H}}v_0$ are the same for any horizontal position φ, λ , i.e.:

$$v_0 = f(h),$$

$$\nabla_{\text{H}}v_0 = f(h).$$

The resulting integrated v still depends on horizontal coordinates, though:

$$v = f(\varphi, \lambda, h) = f_0(h) + f_{\nabla_{\text{H}}}(\varphi, \lambda, h). \quad (3.11)$$

The functions $f_0(h)$ and $f_{\nabla_{\text{H}}}(\varphi, \lambda, h)$ represent, respectively, the contribution from mean conditions and the horizontal gradient of the atmospheric parameter v . The former is exactly the same as the function defined for a single profile in an ellipsoidal atmosphere

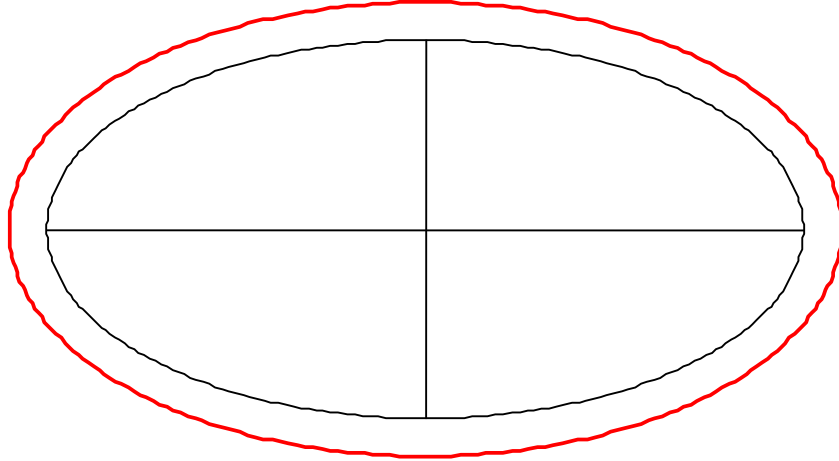


Figure 3.5: Sample ellipsoidal iso-indicial (in red).

(section 3.2.3). The latter is:

$$f_{\nabla_H}(\varphi, \lambda, h) \equiv \int_{\mathbf{r}_{0,h}}^{\mathbf{r}} \nabla_{H^v_0}(\mathbf{r}) \cdot d\mathbf{r}. \quad (3.12)$$

where $\mathbf{r} \stackrel{XYZ}{\leftarrow}_{\varphi\lambda h}(\varphi, \lambda, h)$ is the point of interest, at which the atmospheric parameter v is sought; $\mathbf{r}_{0,h} \stackrel{XYZ}{\leftarrow}_{\varphi\lambda h}(\varphi_0, \lambda_0, h)$ is a point lying along the normal passing through the base location, corresponding to the point of interest: i.e., it has the same vertical coordinate h as the point of interest \mathbf{r} but horizontal coordinates equal to the base location's (φ_0, λ_0) ⁷; “ \cdot ” denotes dot product.

The integral above can be performed along any path. There is one particular path, though, that simplifies the integral to a closed-form expression: it is the path of constant

⁷Contrast:

$$\begin{aligned} \mathbf{r} &\stackrel{XYZ}{\leftarrow}_{\varphi\lambda h}(\varphi, \lambda, h), \\ \mathbf{r}_0 &\stackrel{XYZ}{\leftarrow}_{\varphi\lambda h}(\varphi_0, \lambda_0, h_0), \\ \mathbf{r}_{0,h} &\stackrel{XYZ}{\leftarrow}_{\varphi\lambda h}(\varphi_0, \lambda_0, h); \end{aligned}$$

see also Figure 3.7 on p. 65.

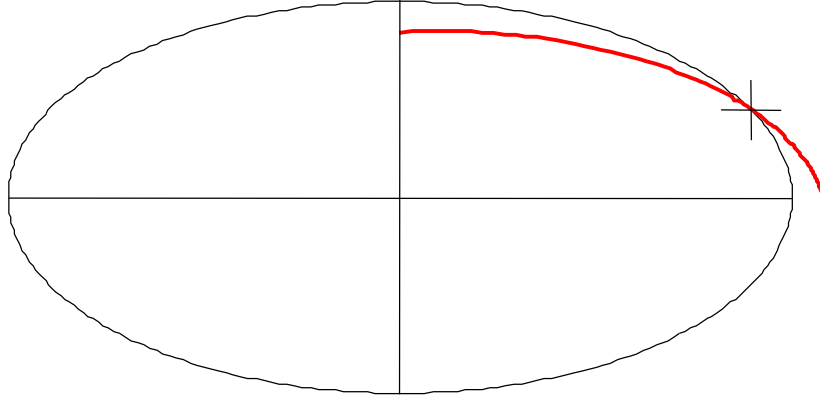


Figure 3.6: Sample gradient iso-indicial (in red) in the vicinity of a base point (marked as a cross).

azimuth, called a *loxodrome*. The resulting closed-form expression is:

$$f_{\nabla_H}(\varphi, \lambda, h) = l \left(\sin a \left. \frac{\partial v_0}{\partial x} \right|_h + \cos a \left. \frac{\partial v_0}{\partial y} \right|_h \right), \quad (3.13)$$

where $\partial v_0/\partial x$ and $\partial v_0/\partial y$ are directional derivatives of v along, respectively, east and north directions (thus making up the horizontal gradient), defined at the base location (φ_0, λ_0) and evaluated at a particular h ; l is the loxodrome length (which is not the same as the along-path distance ℓ nor the straight-line distance), and a is the loxodrome azimuth (which is not the same as the azimuth of the initial ray direction, α). For details, please see Appendix V.⁸ Figures 3.6 and 3.7 illustrate the gradient atmospheric structure.

⁸This is similar to the work presented by Gardner [1976], in the sense that $v_0, \nabla_H v_0$ are postulated constant with respect to horizontal coordinates but can change with height. The main difference is that, whereas Gardner employs a gradient for total refractivity, here we employ instead a gradient for each atmospheric parameter. Also Gardner builds his gradient atmosphere upon a spherical concentric atmosphere, whereas here we build it upon an ellipsoidal atmosphere. Finally, Gardner simplifies the equivalent of integral in eq. (3.12) evaluating it along a geodesic (a spherical great circle in his spherical atmosphere), whereas we evaluate it along a loxodrome, the only path in which the horizontal gradient is rigorously constant.

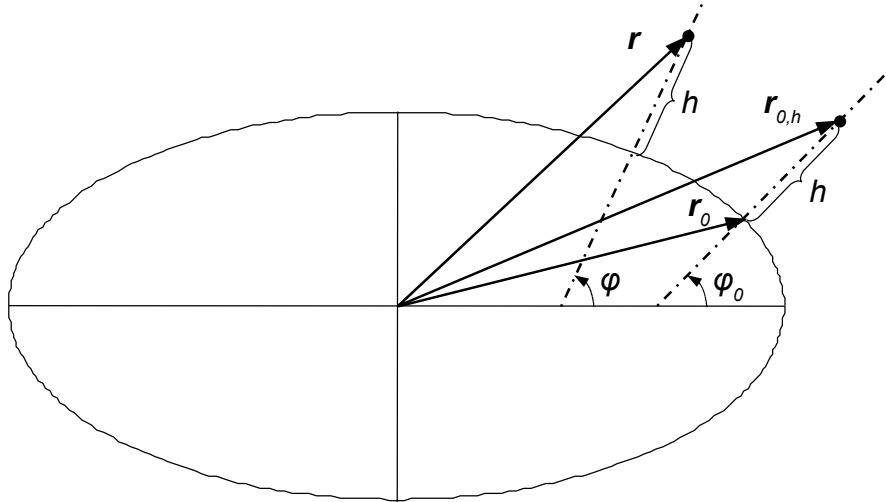


Figure 3.7: Angles, distances, and vectors involved in a gradient atmosphere.

3.2.5 Three-dimensional (3d)

This is the most general case: one makes no assumption about the variation of the atmospheric parameters, neither in the vertical nor horizontal directions. In other words, one takes as-is whatever the atmospheric source is able to represent (Figure 3.8).

With this, we close the description of the main options available in ray-tracing: ray-path models and atmospheric structures. In Chapter 5, the most important combinations of ray-path and atmospheric structure are compared in terms of slant delay and assessed in terms of their impact in GPS positioning. Just before that, in Chapter 4, we take a detour to recapitulate the role of the neutral atmosphere delay in GPS positioning.

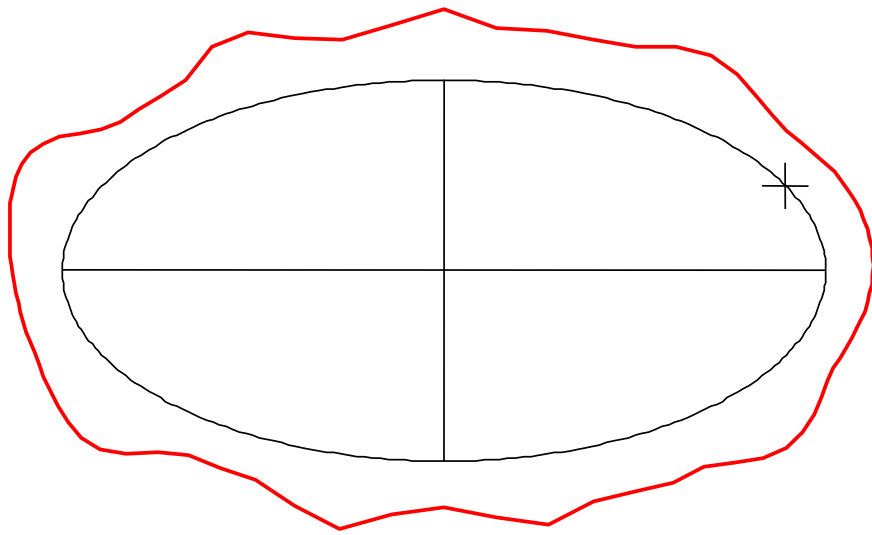


Figure 3.8: Sample 3d iso-indicial.

Chapter 4

The neutral atmosphere delay in GPS positioning

The present chapter exists to support the design of some of the experiments reported in Chapter 5. Here we are interested in selecting prospective GPS applications that could benefit from a possibly better slant factor model, to be assessed in an experiment. The rationale is that, for a GPS application in which the delay is already satisfactorily mitigated, even a much better slant factor model would give only marginally or negligibly better results.

4.1 Vulnerabilities

GPS positioning applications have three general vulnerabilities, that are aggravated by the presence of the neutral atmosphere delay.

(a) **Linear dependency:** In the estimation of unknown parameters, common terms of different error sources cannot be separated if they are linearly dependent. One example is the case of receiver clock error and carrier-phase ambiguity when observing a single

satellite (which is resolved observing several satellites). Another example is given by Leick [2003, p. 277]:

Whereas the propagation media effects on satellite signals, indeed, show a variation with azimuth and elevation angle, there is a station average that can be considered common to all signals.

So the slant delay $d_i^j(t)$ from a receiver i to a satellite j at an epoch t can be split up in a common term, $\bar{d}_i(t_1 < t < t_2)$, valid for all satellites over the period $t_1 < t < t_2$, and a variation over it, $\delta d_i^j(t)$:

$$d_i^j(t) = \bar{d}_i(t_1 < t < t_2) + \delta d_i^j(t). \quad (4.1)$$

If $t_1 = t_2$, then the common term $\bar{d}_i(t_1 < t < t_2)$ cannot be estimated separately but rather is absorbed by the estimate of the receiver clock error. Leick further concludes that modeling of troposphere is useful only if it reduces the variability with respect to the common term.

The common term $\bar{d}_i(t_1 < t < t_2)$ is not to be confused with the zenith delay d^z . The two are obviously related, because the same slant delay $d_i^j(t)$ can be expressed as:

$$d_i^j(t) = k_i^j(t) \times d_i^z(t),$$

where k is the slant factor. From the two expressions above for the same slant delay $d_i^j(t)$ we see that, for a single epoch $t_1 = t_2 = t$, the common term \bar{d}_i equals the zenith delay d^z times an average slant factor \bar{k}_i :

$$\bar{d}_i = \bar{k}_i \times d_i^z,$$

where \bar{k}_i is:

$$\bar{k}_i = \sum_{j=1}^{N_i} k_i^j / N_i,$$

and N_i is the number of visible satellites at the receiver i . If $t_2 > t_1$, then the common term $\bar{d}_i(t_1 < t < t_2)$ corresponds to an average slant delay over all satellites j in the period

$t_1 < t < t_2$:

$$\bar{d}_i(t_1 < t < t_2) = \frac{\int_{t_1}^{t_2} k_i^j(t) \times d_i^z(t) dt}{t_2 - t_1}.$$

In fact, it is exactly this averaging over time that helps to decorrelate receiver clock error (modeled as white noise) and zenith NAD (modeled as a polynomial piece over time or as a stochastic process auto-correlated in time); see more discussion about it on p. 73. Which brings us to the next vulnerability.

(b) Nearly linear dependency or strong correlation: Certain parameters are not completely linearly dependent yet are nearly so; in other words, they are strongly correlated. In the absence of additional observations or constraints, we are prevented from separating those parameters clearly, leading to either a bias in one by the unaccounted for effects of the other or poor resolution (e.g., large standard deviations) when we try to estimate both simultaneously.

That strong correlation is usually the case between zenith delay and receiver height. Treuhaft [1992] illustrated the correlation between a change in relative height (Δh) and a change in zenith delay (d^z) due to change in the (VLBI equivalent of a) GPS ranging measurement (ρ) at a particular elevation angle (ϵ):

$$\partial \rho(\epsilon) \approx \sin \epsilon \partial \Delta h,$$

$$\partial \rho(\epsilon) \approx \frac{1}{\sin \epsilon} \partial d^z.$$

Figure 4.1 demonstrates that the signatures of a shift in zenith delay and in receiver height (in the presence of a typical clock error) can be very similar, especially at higher elevation angles. In other words, it is the ranging measurements from satellites at lower elevation

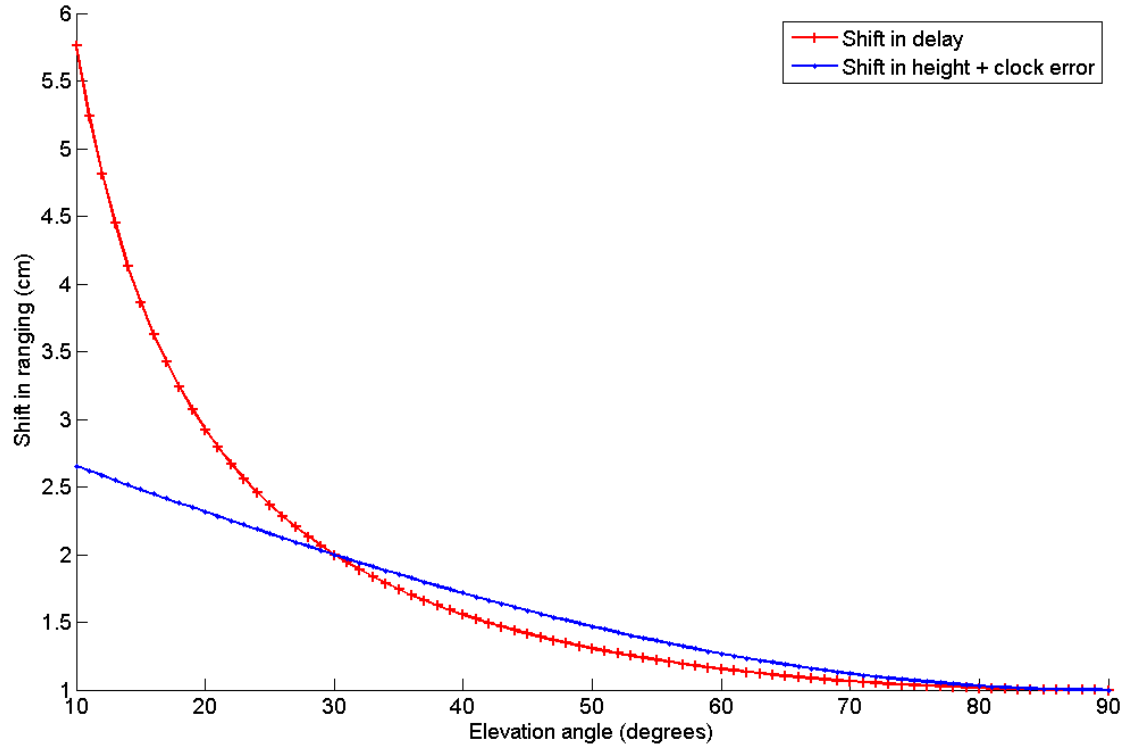


Figure 4.1: Shift in ranging, for different elevation angles, due to (i) a 1 cm shift in zenith delay ($1 \text{ cm}/\sin \epsilon$) and (ii) a -2 cm shift in receiver antenna height, combined with an usual 3 cm clock error ($-2 \text{ cm} \sin \epsilon + 3 \text{ cm}$).

angles that help decorrelate d^z and Δh . At such lower elevation angles ($\epsilon \approx 0$):

$$\left. \frac{\partial \rho}{\partial \Delta h} \right|_{\epsilon \approx 0} \approx 0, \quad \left. \frac{\partial \rho}{\partial d^z} \right|_{\epsilon \approx 0} \approx \infty.$$

Another, even more dramatic case of the vulnerabilities inherent in strong correlations in GPS is that between receiver height and receiver clock error. For the GPS practitioner accustomed to expect worse results for vertical positioning compared to horizontal, it might come as a surprise the fact that, were it not for the receiver clock error, height would be even better determined than horizontal position. Misra [1996, p. 60,62] states:

The problem . . . is not simply that the estimation of the clock bias “uses up” a measurement, but that it alters the basic structure of the position estimation problem in a negative way.

... the price for estimation [or elimination] of the clock bias is being paid for almost entirely by loss of accuracy in the vertical dimension.

Kuang et al. [1996, p. 42] reaches similar conclusions, for both absolute and relative positioning:

Estimating the receiver clock error or eliminating the bias by differencing the measurements [across satellites] ... degrade the information in the average direction of the measurements, which would be the best observed component of the position if the receiver clock error were known.

The connection between heights and the average direction between receiver and satellites is explained next.¹

(c) Geometric weakness of height: While generally we are able to observe GPS satellites at all azimuth quadrants, we can obviously observe satellites only in the less than half a hemisphere above the ground [Spilker, 1996, p. 545]. That makes the vertical direction to coincide approximately with the average direction between receiver and satellites, exactly the direction which will be most deteriorated by receiver clock error, as we saw previously.² A secondary deterioration comes from the fact that, while a uniform sampling across azimuths allows one to effectively average out many systematic effects on estimated horizontal coordinates, the same benefit is unavailable for heights.

¹This analysis considers only observation geometry and neglects all other error sources, most dangerously atmospheric delays, antenna gain, and multipath. But even in the presence of those error sources, the absence of receiver clock error would still dramatically improve vertical positioning, though not necessarily making it better than horizontal positioning.

²On the other side of the same coin, were we able to keep the receiver clock error under tighter control, this same negative vulnerability would become a positive strength [Kuang et al., 1996, p. 42]:

Should good a priori information about the receiver clock bias be available, point positioning solutions, especially the height component, may be significantly improved. The height component of baselines should also be improved ... For the global network, this implies that ... quantities dependent upon the orientation of the coordinate polyhedron such as the geocenter and Earth orientation parameters may be improved.

4.2 Mitigation techniques

In the present section we propose a classification of all the techniques available for the mitigation of the neutral atmosphere delay in GPS positioning. First we enumerate the techniques on p. 72. The names given for each technique were chosen as mnemonics, not to be taken literally; for example, when we say “eliminate” we mean that such a technique eliminates some, hopefully most — certainly never all — of the delay. Then we enumerate the requirements of each technique — requirements in the sense that, if not met, the corresponding mitigation technique becomes less effective, not necessarily completely ineffective. Again, the descriptions in the list are intentionally brief; please see text for detailed explanations. Finally, we discuss the pros and cons of each mitigation technique vis-à-vis the intended GPS application. The ultimate objective is to select a GPS application that could possibly benefit from a better delay model, to be later evaluated in experiments.

- (0) *no* mitigation technique;
- (1) *avoiding* the delay by discarding lower-elevation observations;
- (2) *eliminating* the delay by across-receiver differencing;
- (3) *estimating* the delay from the GPS observations;
- (4) *predicting* the delay with ready-to-use models;
- (5) *determining* the delay with in situ external observations;
- (6) *interpolating* the delay from estimates at nearby stations.

Requirements for each mitigation technique:

- (0) none:
 - (0.a) estimation of receiver clock error;
 - (0.b) no interest in the parameters of receiver clock error or zenith delay, *or*
 - (0.c) no constraint on zenith delay parameter, making it auto-correlated.

- (1) avoiding the delay by discarding lower-elevation observations:
 - (1.a) abundant simultaneously visible satellites;
 - (1.b) no estimation of both zenith delay and receiver height.
- (2) eliminating the delay by across-receiver differencing:
 - (2.a) relative positioning;
 - (2.b) small baseline length;
 - (2.c) small height offset;
 - (2.d) homogeneous weather, with no local anomalies.
- (3) estimating the delay from the GPS observations:
 - (3.a) carrier-phase observations;
 - (3.b) initialization time (first 1 h);
 - (3.c) no estimation of actual vertical movement correlated with zenith delay.
- (4) predicting the delay with ready-to-use models:
 - (4.a) dm-level accuracy suffices.
- (5) determining the delay with in situ external observations:
 - (5.a) funds to purchase and maintain additional equipment.
- (6) interpolating the delay from estimates at nearby stations:
 - (6.a) receiver within network extents;
 - (6.b) sufficiently dense network of stations.

(0) None: Even when we are not employing any mitigation technique, we might benefit from the absorption of an average delay by the receiver clock error estimate (see vulnerability (a) on p. 67). Leick [2003] states that the part of the delay in common among all satellites does not affect the position, but merely changes the estimated clock parameter. Spilker [1996, p. 544] makes a similar statement: “Clearly, a constant bias for four pseudoranges to four different satellites would not affect position error at all but would cause a user clock time bias error.”

At first sight, it might seem that the leaking of tropospheric delay into the receiver clock error should be a nuisance for GPS applications interested in either of those two parameters — applications such as time transfer/frequency exchange and meteorology/climate monitoring. That would be true only if we were estimating tropospheric delay independently every epoch. In such a case, the zenith delay parameter would vary from epoch to epoch as much as the receiver clock error varies — i.e., as a white noise random process.

At closer look, though, we realize the importance of a functional or stochastic constraint on the zenith delay, that introduces some auto-correlation for that parameter. Such a constraint makes the average delay $\bar{d}_i(t_1 < t < t_2)$ encompass more than one epoch, effectively decorrelating it from the numerous independent receiver clock errors at each epoch. As a consequence, both estimated parameters (receiver clock error and zenith delay) become less biased and thus still useful for time/frequency and weather/climate applications of GPS.

The present technique applies specially for point positioning. In contrast, the present technique does not apply for relative positioning. That is because in relative positioning usually one does *not* estimate receiver clock error: such an error is canceled out by differencing observations across satellites simultaneously in view. (Across-satellite differencing is distinct from the across-receiver differencing employed in technique (2).)

(1) Avoiding the delay by discarding lower-elevation observations: In this technique the delay is still present in the observations; we just avoid most of it by discarding the observations most affected by it (i.e., the ones at lower elevation angles). A common variation of this technique is to keep in the estimation procedure the observations at lower elevation angles yet significantly downweight them.

The present technique (1) is not desirable in environments where sky visibility is poor

thus DOP is large, such as in urban canyons. In other words, when the number of visible satellites is close to the minimum (four), discarding observations might be an unfeasible luxury. That situation should be ameliorated in the future with the greater availability of satellites from, e.g., the European Galileo system or the revitalization of the Russian GLONASS system.

If technique (1) is used alone, vulnerability (b) tells us that height will be biased, due to the absorption of unaccounted for delay. Yet the height standard deviation will be unaffected, as the discarded observations at lower elevation angles would contribute very little to the determination of height alone ($\left. \frac{\partial p}{\partial h} \right|_{\epsilon \approx 0} \approx 0$). (That would be different were technique (1) to be used in conjunction with technique (3).)

One must also realize that (1) is employed to mitigate not only the delay but also other error sources that predominate at lower elevation angles, mainly multipath and poor antenna gain off-zenith.

(2) Eliminating the delay by across-receiver differencing: The delay affecting two receivers is the more similar the closer the receivers are. Provided that the baseline joining them is not excessively long (say, less than 30 km) *and* that their height offset is small (say, less than 100 m) — as the delay varies very rapidly with respect to height —, differencing observations across the two receivers will eliminate most of the delay. That technique is obviously unavailable in point positioning.

Technique (2) is so powerful to mitigate the delay that relative positioning applications can be designed having baselines short enough so that the delay is satisfactorily mitigated. Indeed, the classification of short and medium-distance baselines can be based partially on the residual delay itself [Bock, 1998, p. 483]:

One can define, however, a lower limit [for medium-distance baselines] as the shortest distance at which residual ionospheric refraction, tropospheric refraction, and/or orbital errors between sites are greater than total high-frequency site and receiver-specific errors.

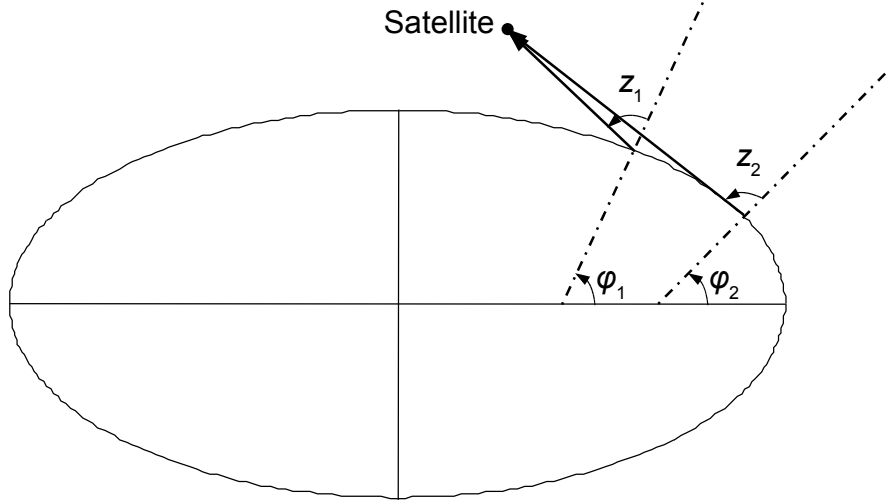


Figure 4.2: Zenith distance of the same satellite at two different receivers.

Conversely, the more homogeneous is the weather between the two ends of a baseline, the longer one can extend that baseline. In contrast, local weather anomalies, such as passing fronts, will make the present technique less effective.

Technique (1) is beneficial to (2), in the sense that the latter is less effective the lower is the satellite in common view. That fact can be understood recalling that the elevation angle (ϵ_1, ϵ_2) of the same satellite at two different receivers (1,2) is different ($\epsilon_1 \neq \epsilon_2$), because the ellipsoidal normals from which (the complement to) the elevation angles are reckoned is different (see Figure 4.2). In fact, the elevation angles at each receiver are the more different the longer is the baseline between the receivers. At higher elevation angles ($\epsilon \approx 90^\circ$), such a discrepancy in elevation angle ($\epsilon_1 - \epsilon_2$) yields a small discrepancy in delay ($d(\epsilon_1) - d(\epsilon_2) \approx 0$), because the delay varies very little with respect to elevation angle near zenith ($\frac{\partial d}{\partial \epsilon} |_{\epsilon \approx 90^\circ} \approx 0$). But at lower elevations ($\epsilon \ll 90^\circ$) the delay varies much more rapidly with respect to elevation angle ($\frac{\partial d}{\partial \epsilon} |_{\epsilon \ll 90^\circ} \gg 0$), so even a small discrepancy in elevation angle ($\epsilon_1 - \epsilon_2 \approx 0$) yields a much greater discrepancy in the delay ($d(\epsilon_1) - d(\epsilon_2) \gg 0$). Precisely for that reason, when differencing observations across receivers, the elimination of the delay affecting a given satellite's observations is less effective the lower

is that satellite at any receiver.

(3) Estimating the delay from GPS observations: The fact that the GPS observations are affected by the delay allows one to invert those very observations to estimate parameters modeling the delay. One follows the decomposition introduced in section 2.4: the slant factor model is assumed known (or more, precisely, predicted from external models and observations) and the zenith delay is estimated, as it becomes a parameter redundant over the observations to all visible satellites.

The present technique (3) relies on the availability of carrier-phase observations, as pseudorange observations are not precise enough. Which implies in the second requirement for technique (3), namely, the need for some time to initialize carrier-phase ambiguities after the receiver is turned on.

In contrast with technique (1), the present technique (3) yields less biased heights, as it prevents the otherwise unaccounted for delay from leaking into height. But the height standard deviation will be worse in (3) than in (1), because heights and zenith delay are highly correlated. Furthermore, technique (1) is detrimental to technique (3) if both are used in conjunction, because one way of decorrelating height and zenith delay thus improving their individual standard deviations would be lowering the cut-off elevation angle, exactly the opposite of what (1) does.

Finally, care should be exercised in applications where (i) the receiver height is estimated independently every epoch and (ii) the receiver is subject to vertical displacements varying similarly to the zenith delay itself, because that could lead to some aliasing of the vertical displacements onto zenith delay and vice-versa [Dach and Dietrich, 2000; Vey et al., 2002]. Examples of applications are given in Dodson et al. [2001]; Urschl et al. [2005].

(4) Predicting the delay with ready-to-use models: Please recall from section 2.8 (p. 48) that ready-to-use models require as input only the intrinsic independent variables upon which the delay depends: epoch, receiver position, and satellite direction. A prime example are the UNB models [Leandro et al., 2006a].

In the present technique (4), the delay effects on GPS estimates are due to the discrepancy between predicted and actual delay. It is important to recognize that (4) will never be as good a mitigation technique as (3) or (5). That is because (4) relies only on prior observations, while (3) and (5) exploit observations (GPS or external) available exactly at or near to the epoch and location at which one wishes to mitigate the delay.

Therefore (4) taps into the applications for which (3) and (5) are not possible or desired [Schüler, 2006]. An example is during the ambiguity initialization time of systems for real-time kinematic surveying. Another example are applications for which dm-level accuracies suffice, such as navigation in general. Technique (4) is especially useful in validating other mitigation techniques, since it is independent.

(5) Determining the delay with in situ external observations: This technique is a type of prediction (section 2.3.1), as it does not rely on GPS observations. Yet it is distinct from technique (4), in that it relies on in situ observations. Examples are the use of radiometers to measure sky brightness temperature, barometers to measure surface pressure, and radiosondes to measure profiles of pressure, temperature, and humidity. Such observations are coupled to prior models that allow one to convert the measured quantities to the delay (usually only components of the total delay).

The present technique (5) relies on additional equipment, sometimes costly and bulky. That may not always be available nor desirable. Even in cases for which it applies, such as permanent tracking stations, technique (5) is usually replaced by the inexpensive (3).

(6) Interpolating the delay from values at nearby stations: This technique exploits the availability of delay values at nearby locations, obtained through techniques (3) or (5). We expect the error to be due to the interpolation itself plus errors in obtaining the original values at nearby locations [Dai et al., 2003; Zhang and Lachapelle, 2001].

The present technique (6) is becoming increasingly more popular, as part of services offered commercially by companies with networks of tracking stations for real-time kinematic surveying applications. The greater the distance between interpolation point and nearby stations, the greater will be the interpolation error, therefore the denser the network the better (6) will perform, obviously.

It must have become apparent that any mitigation technique is rarely employed alone. A good example of the synergy among different techniques is that involving (3), (4), and (5): only the non-hydrostatic portion of the delay is estimated through (3), the hydrostatic component being part of the modeled counterpart to the observations (i.e., as a correction to the computed geometric range between receiver and satellite), after being predicted through (4) or determined through (5); furthermore, techniques (3) and (4) might provide only zenith delays, in which case the slant delay requires a slant factor predicted through technique (4).

Another general aspect regarding the classification above is that certain examples have characteristics of multiple mitigation techniques. For example, at first sight a numerical weather model (NWM) might be considered a ready-to-use prediction model, thus an instance of technique (4) (predicting the delay with ready-to-use models). At closer look, if the NWM assimilated observations collected nearby the location and epoch at which delay predictions are sought, then the NWM might be considered an instance of technique (5) (determination of delay using in situ observations). Finally, if the NWM has assimilated GPS-estimated zenith delays, it might be considered as instance of technique (6)

(interpolating the delay from nearby stations).

Based on the analysis above and also on expert advice, we selected precise point positioning to be later evaluated at experiments, because it is very vulnerable to the neutral atmosphere delay.

Chapter 5

Experiments

Let us start by recalling the ray-path and atmospheric models presented in Chapter 3; their diagram is repeated in Figure 5.1. In the present chapter, the above-mentioned models are assessed in experiments. In the first experiment (section 5.1) we assess those models in terms of slant delays. In the second and third experiments (section 5.2 and section 5.3) we assess the same models in terms of their impact in GPS positioning, more specifically, in terms of GPS precise point positioning (PPP), a decision based on discussion in Chapter 4. The last two experiments differ between themselves by the way that the delay is mitigated in PPP: following the classification introduced in section 2.3.1, in the second experiment (section 5.2) the delay is only predicted; in the third one (section 5.3), it is estimated.

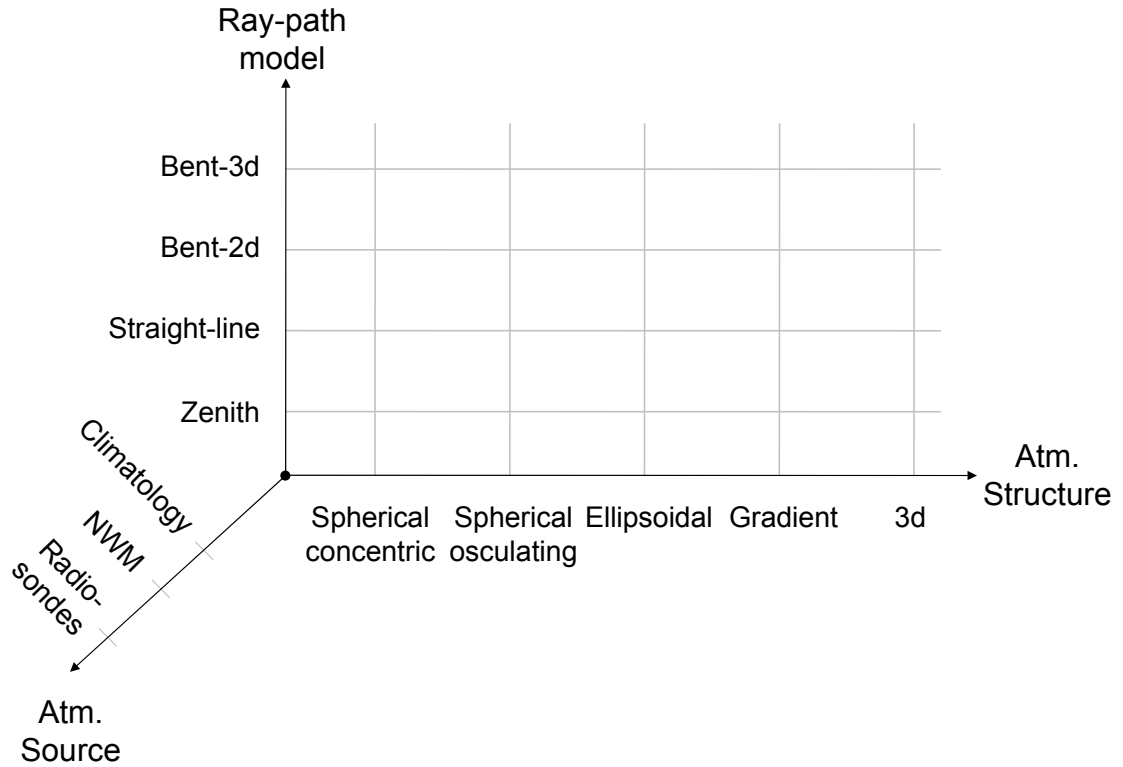


Figure 5.1: Diagram of options available in ray-tracing.

5.1 Delay

There are numerous combinations of ray-path and atmospheric structure. Some combinations, albeit different, are expected to give the same result. For example, the more complicated combination of a bent-3d ray-path model and a spherical atmospheric structure (either concentric or osculating) shall give the same result as the simpler combination of a bent-2d ray-path model and the same spherical atmospheric structure. In contrast, for most other combinations, we do expect some discrepancy. In the experiment reported in the present section, we have quantified that discrepancy in integrated neutral atmosphere delay due to different models for the ray-path and the atmosphere.

We make pair-wise comparisons between combinations that are equal in all but one aspect; in such a way we can isolate the contribution of that single aspect. First we keep

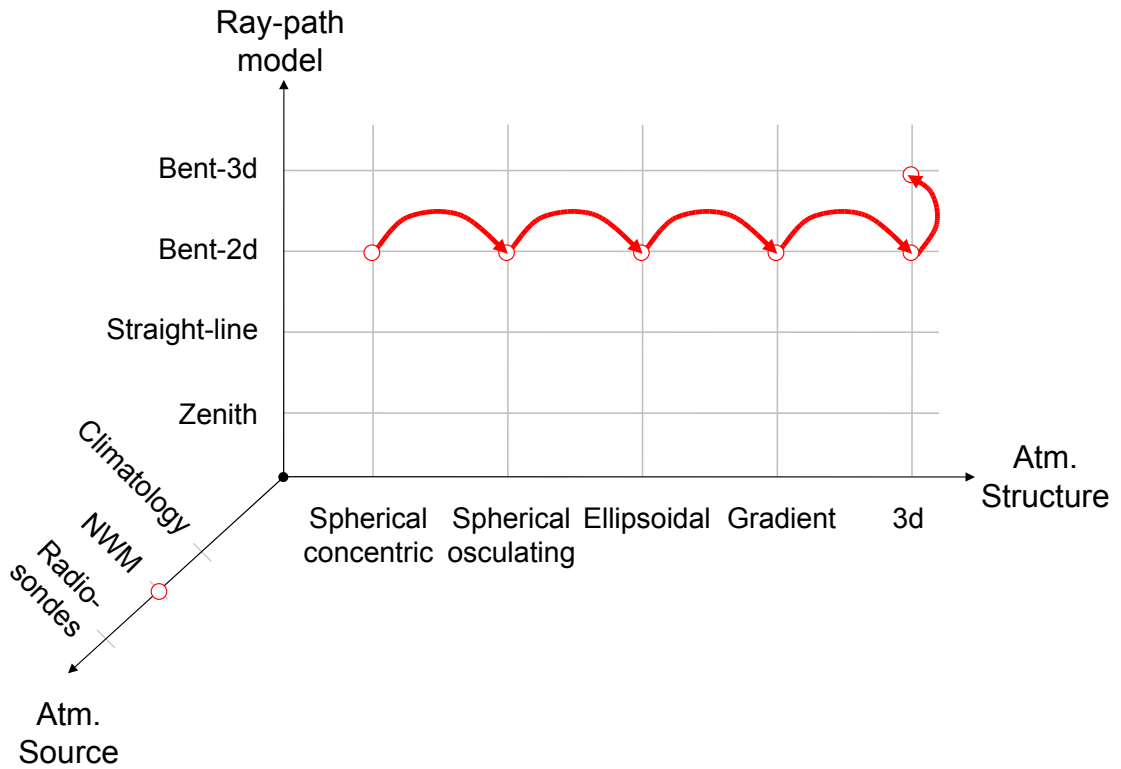


Figure 5.2: Diagram of sequence of pair-wise comparisons.

the ray-path model fixed to the bent-2d case, and then we vary the atmospheric structure, one step at a time. Once we have reached the most realistic atmospheric structure, we keep it fixed and then we vary the ray-path model. Such a sequence is summarized in Figure 5.2, where each arched arrow represents a pair-wise comparison.

The atmospheric source is always fixed to NWM; therefore we are not analyzing the discrepancy between different atmospheric sources — say, a weather model and a climatology. We used the Canadian regional weather model, which has a horizontal resolution of 15km. We do expect the results to depend upon that resolution.

The delay discrepancies represent a snapshot, taken at a single epoch and location. The exact epoch and location (indicated in Table 5.1) were chosen to correspond to a near worst case scenario. That scenario is given by a category 4 hurricane, the 2004 Hurricane

Charley; it was the strongest hurricane that we could find within the spatial and time extents of the NWM archive that we were using. The rationale for that choice was the understanding that, if we did not find significant discrepancies in such a scenario, then we should not expect anything worse under typical conditions.

Table 5.1: Epoch and location of first experiment.

Epoch:	August 14th, 2004, 12 h UTC
Latitude:	+33°49'01"
Longitude:	−76°06'28"

We show the delay discrepancies in the form of a skyplot (e.g., Figure 5.3) — north is to the top, south is to the bottom, east to the right, and west to the left. Samples were taken at directions regularly spaced in azimuth (every 45°) and regularly spaced in $1/\sin(\varepsilon)$ (at 90°, 14°, 8°, 6°, 4°, 3°), which is approximately the way that the delay grows with elevation angle ε (see, e.g., Figure 5.6, on p. 90). Please be aware that those skyplots do *not* cover the entire sky, contrary to what they might suggest at first sight. They were intentionally distorted to emphasize the region near the horizon — if they were drawn to scale, the chosen elevation angles would form only a thin band near the horizon. (Due to that intentional distortion, the center of each skyplot is not exactly at zenith, only close to it.) The discrepancy in each of the delay components (namely, hydrostatic, non-hydrostatic, and geometric), in addition to the total, is shown separately in different skyplots. The sign of the discrepancies is depicted in different hues (blue for negative and red for positive), while the magnitude of the discrepancies is depicted both as the intensity of the color and the radius of the balls at each sampled direction. Bear in mind that, even though we always sample all $8 \times 6 = 48$ directions, sometimes the sample at a particular direction happens to be invisible in the skyplot because its value is too small, when depicted as balls of proportional radius (see, e.g., E-W direction in Figure 5.3, or NE, SE, SW, NW directions in Figure 5.10).

Details about the ray-tracing procedure are given in Appendix I; results reported here were obtained with a tolerance of 0.1 mm.

The first comparison is between a spherical concentric and a spherical osculating atmosphere. The first thing to notice in the resulting discrepancy (Figure 5.3) is its large magnitude, reaching meters and decimeters in the hydrostatic (Figure 5.3(a)) and non-hydrostatic (Figure 5.3(b)) components, respectively. The second aspect is a clear trend, in the north–south direction.

Figure 5.3 has important consequences with regard to azimuthal symmetry, i.e., whether a given atmospheric model yields the same delay values at the same elevation angle but different azimuths. At this point it should be apparent that at least one of the two spherical atmospheres is *not* azimuthally symmetric. In other words, a spherical atmosphere does not necessarily imply azimuthal symmetry. That is in contrast with the usage of those two expressions in the literature — it is not uncommon to find them being used interchangeably. Having checked both individually (see next paragraph), we can assert that the spherical osculating one *is* azimuthally symmetric, while the spherical concentric one *is not*.

Let us make a quick digression to demonstrate the azimuthal symmetry in a spherical osculating atmosphere. In Figures 5.4 and 5.5 we show, respectively, the spherical osculating delay itself and the discrepancy between the delay along each azimuth and the delay along the north azimuth. In Figures 5.6 and 5.7 we inspect the total delay, the same one shown in Figures 5.4(d) and 5.5(d), at this time in a more detailed view. We can notice that the total delay itself is practically the same for the same elevation angle but different azimuths (Figure 5.6). Furthermore, the discrepancy in total delay along each azimuth with respect to the delay along the north azimuth (Figure 5.7) — a measure of the azimuthal asymmetry in the spherical osculating atmosphere — is almost zero, and well

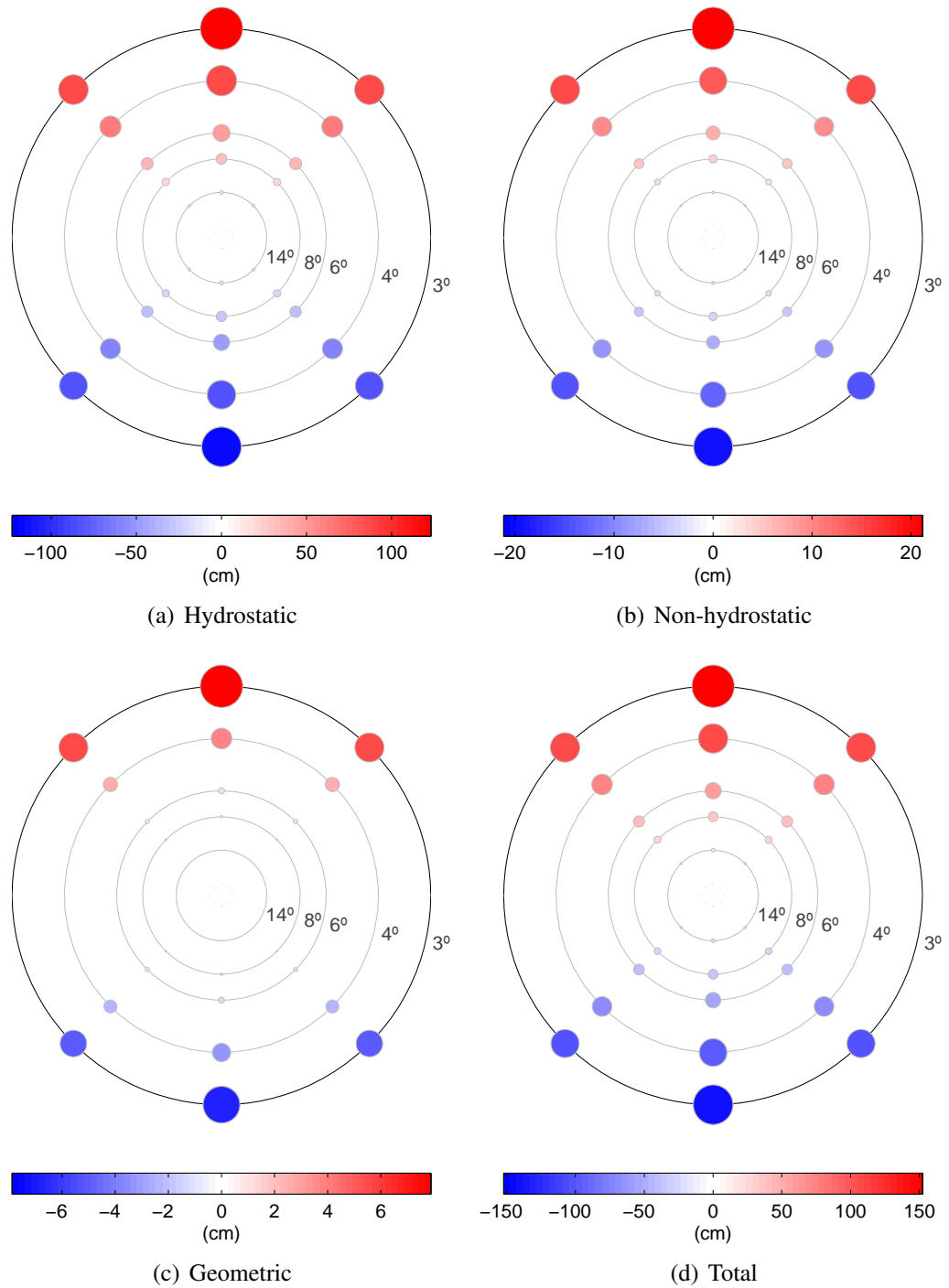


Figure 5.3: Results of first comparison: spherical concentric minus spherical osculating.

explained considering the tolerance set for the ray-tracing procedure, which was 0.1 mm.¹ Albeit negligible, the self-discrepancy in spherical osculating total delay is not random. We interpret the patterns in azimuthal asymmetry exhibited by the computed spherical osculating atmosphere as mere numerical artifacts without physical meaning. Furthermore, we speculatively attribute the cause of the patterns to the loss of floating-point precision after the numerous coordinate transformations involved in ray-tracing. For example, notice in Figure 5.7 (top panel) the discrepancy along the line of different azimuths at the same 90° elevation angle: one could expect that line to be perfectly horizontal, because, mathematically, the zenith direction is exactly the same for any azimuth; that mathematical equivalence, though, is not necessarily valid in numerical computations performed at finite precision, the reason why the delay at zenith is slightly different for different azimuths.

Returning to the first comparison, between spherical concentric and spherical osculating, we attribute the magnitude and trend of the discrepancy shown in Figure 5.3 to the tilting of the spherical concentric horizon with respect to the spherical osculating horizon. Recall that the elevation angle ε is typically reckoned from the ellipsoidal horizon (at least in GPS practice, that is *always* the case). On one hand, the horizon of the osculating sphere coincides with the ellipsoidal horizon at the base point. That coincidence of horizons follows directly from the definition of osculating sphere, in which its radial direction is postulated coincident with the ellipsoidal normal direction. On the other hand, the horizon of the concentric sphere in general does not coincide with the ellipsoidal horizon. The two coincide only at the poles and along the equator; at such latitudes the ellipsoidal normal and the geocentric radial direction coincide.

At this point we would like to put forth a peculiar conclusion regarding the adoption of either concentric or osculating spherical atmospheres by authors of mapping functions. To explain that conclusion, we have to recollect a few facts first. In the process of developing

¹0.1 mm is the tolerance for a single ray-trace; the tolerance for the discrepancy between two ray-traces, assuming them uncorrelated, is thence $\sqrt{(0.1 \text{ mm})^2 + (0.1 \text{ mm})^2} \approx 0.14 \text{ mm}$.

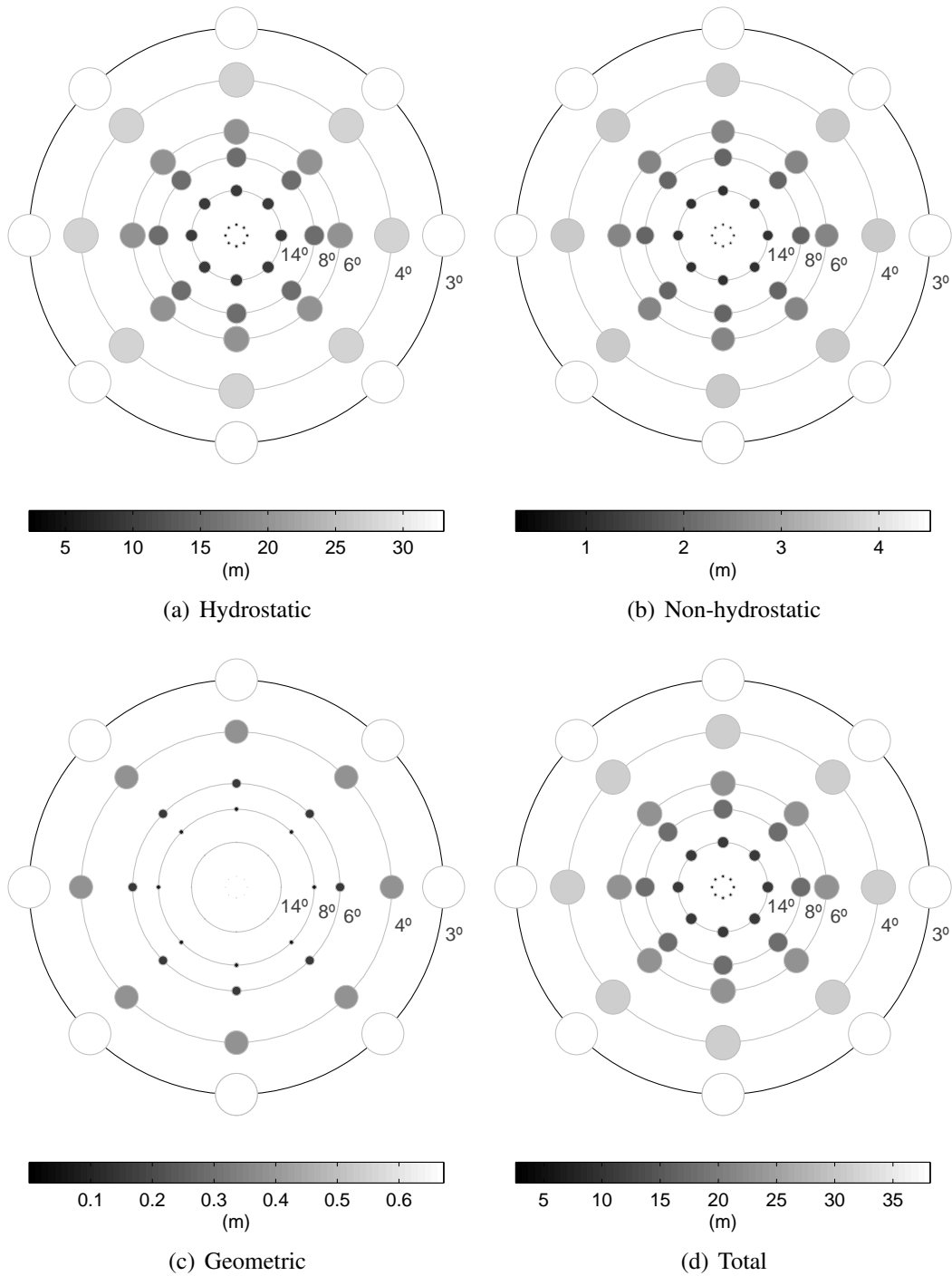


Figure 5.4: Additional results of first comparison: spherical concentric delay (itself).

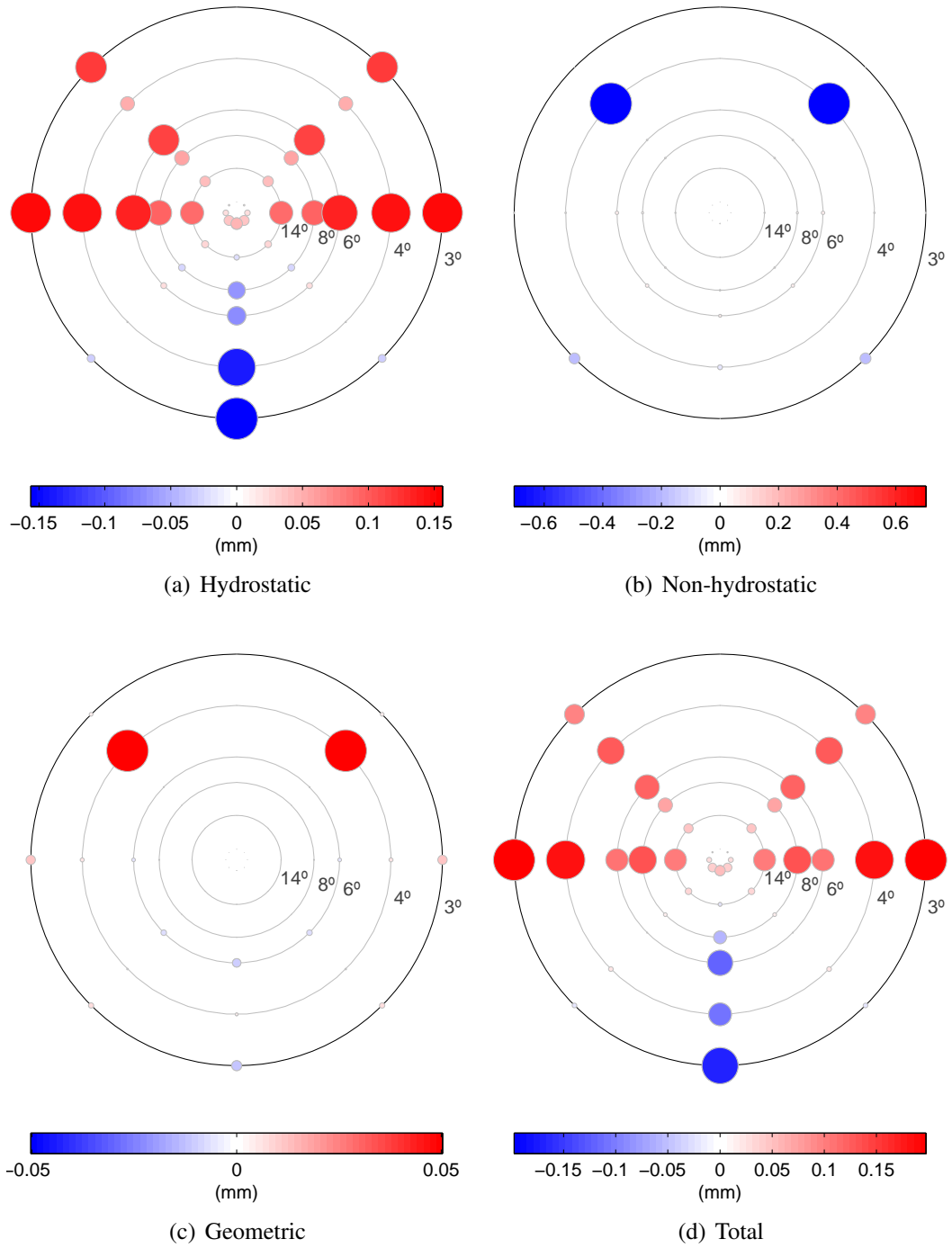


Figure 5.5: Additional results of first comparison: spherical concentric delay (self-discrepancy with respect to north azimuth).

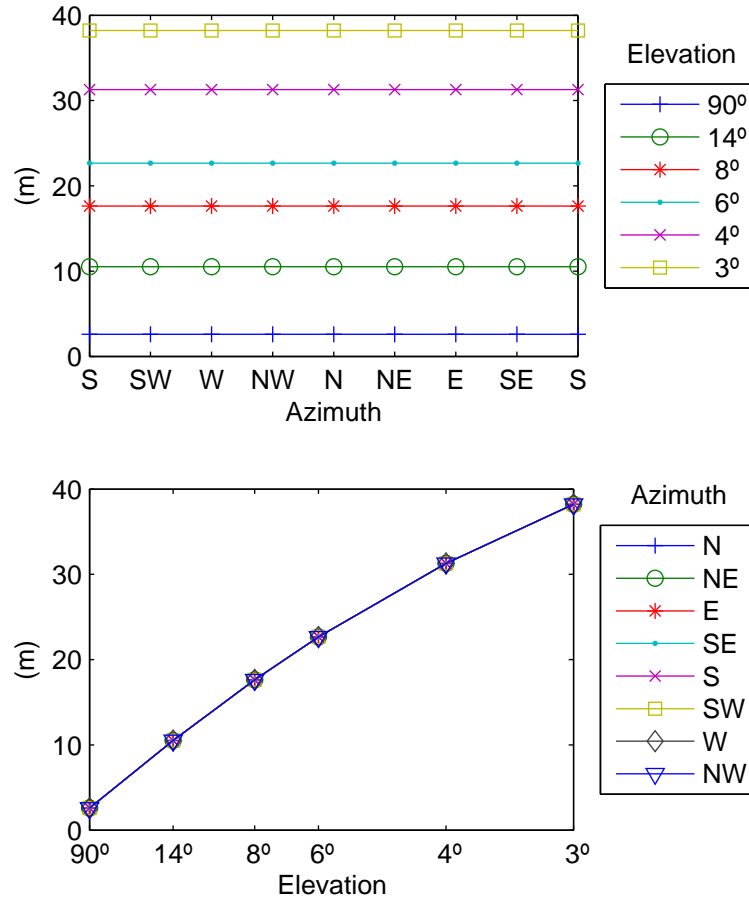


Figure 5.6: Additional results of first comparison: spherical concentric total delay (itself).

a mapping function, typically, the center of the spherical atmosphere is postulated to lie along the direction from which zenith angles are reckoned; in other words, the radial direction of such a spherical atmosphere is postulated to coincide with the zenith direction. But the connection between that zenith direction and the real world is not made at the time of development of a mapping function, only at the time of its usage. Let us give a few examples. Figure 5.9 illustrates two different usage scenarios: in the first scenario a user inputs to a mapping function a zenith angle z reckoned from the ellipsoidal normal (always the case in GPS practice); in a second, distinct, scenario, the user inputs instead zenith angles z' reckoned from the geocentric radial direction. Each of the two scenarios leads to a different orientation of the mapping function's underlying spherical atmosphere: an oscu-

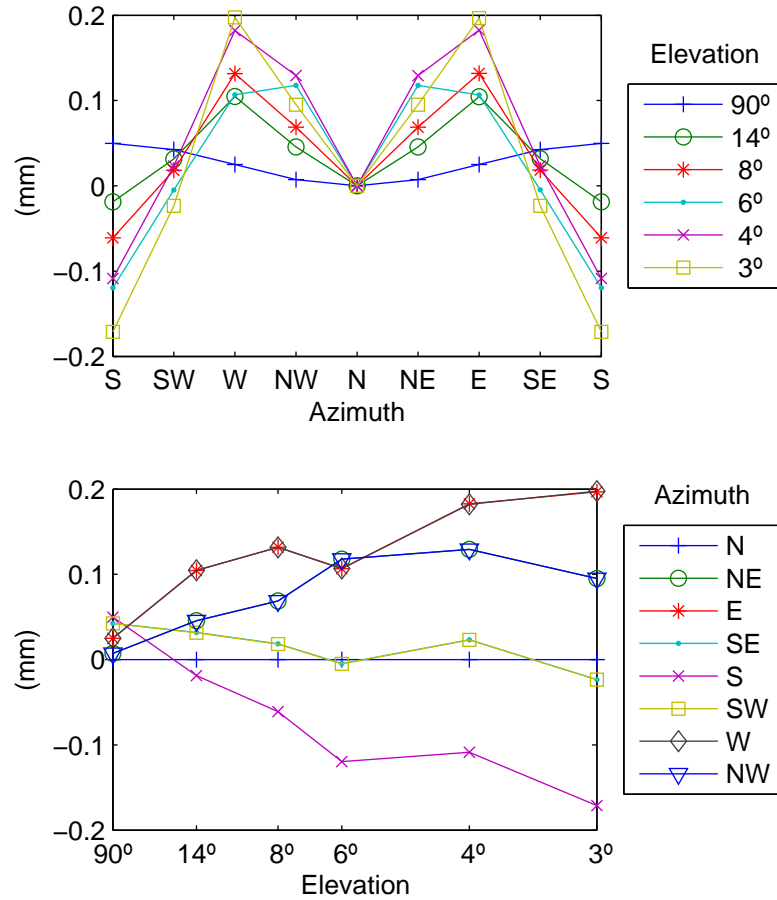


Figure 5.7: Additional results of first comparison: spherical concentric total delay (self-discrepancy with respect to north azimuth).

lating sphere in the first case, and a concentric sphere in the second case. The conclusion, then, is: whether a particular mapping function’s underlying spherical atmosphere is concentric or osculating, that is dictated by the usage of that mapping function — regardless of the author’s intention or awareness when developing it.

We offer the conclusion above as a way to reconcile the following two facts. On one hand, to the best of our knowledge, nowhere in the literature of mapping functions is made the distinction between concentric and osculating spherical atmospheres.² Furthermore, it is not uncommon to find authors calling “geocentric” the zenith angles involved in their

²In contrast, this distinction is well-known in the literature of radio occultation, ever since the earlier studies of Jupiter’s atmosphere from Pioneer spacecraft occultations in the mid-1970’s [Syndergaard, 1998].

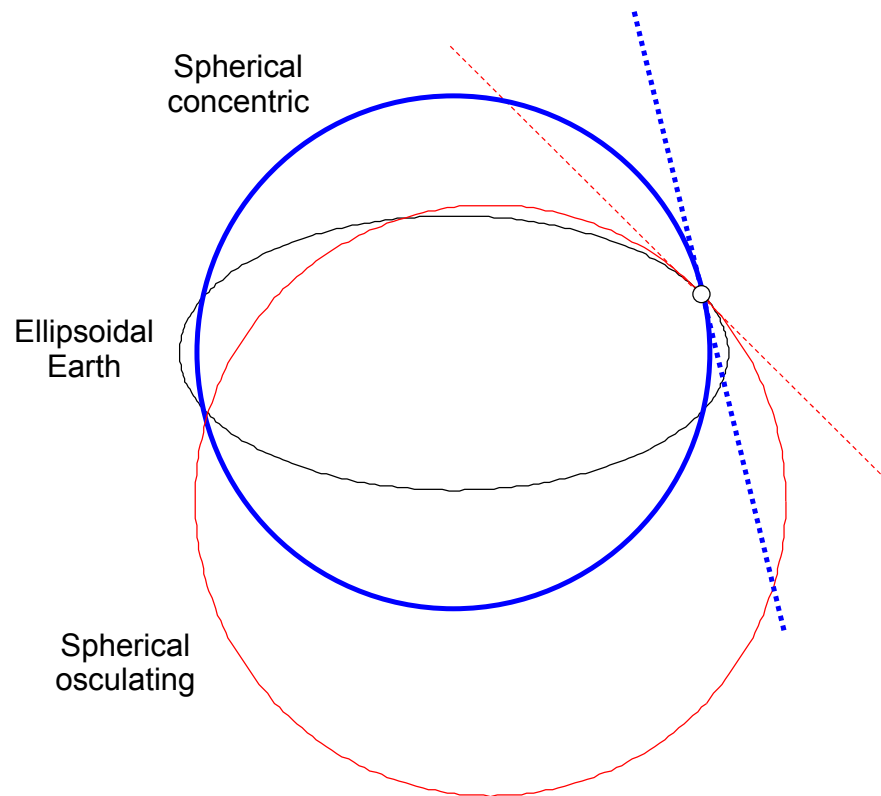


Figure 5.8: Horizons of each concentric (thick blue line) and osculating (thin red line) spheres — base point is at $\varphi = 45^\circ$.

mapping functions (e.g., Mendes [1999, p. 274], Boehm and Schuh [2003, p. 141], Hullely and Pavlis [2007, Fig. 1], Chen and Herring [1997, Fig. A1]), which we interpret as an intention to employ a spherical concentric atmosphere (or a gradient atmosphere augmenting a spherical concentric one). On the other hand, typical results from GPS practice (see sections 5.2 and 5.3) are *not* consistent with a mapping function based on a spherical concentric atmosphere. Restating our conclusion: mapping functions used in GPS are typically of type spherical osculating, not spherical concentric — if not by design, by virtue of their usage.

The next comparison is that between a spherical osculating atmosphere and an ellip-

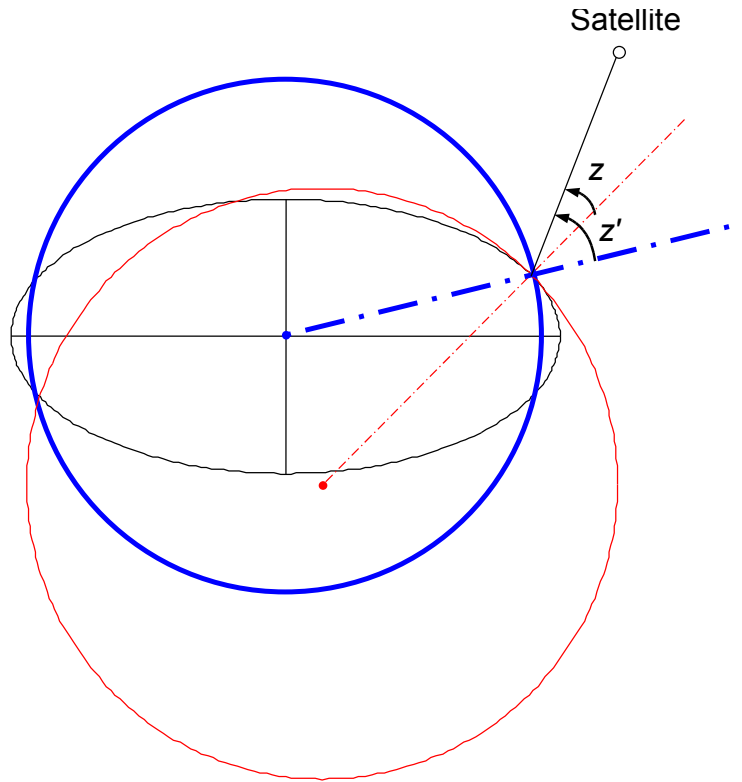


Figure 5.9: Radial directions of each concentric (thick blue lines) and osculating (thin red lines) spheres — base point is at $\varphi = 45^\circ$.

soidal atmosphere. Their discrepancies, shown in Figure 5.10, have much smaller magnitude than in the previous comparison; in fact, it is negligibly small in all but the hydrostatic component (Figure 5.10(a)). We also notice a clear trend in all components: it reaches its maximum along the north–south direction, its minimum along the east–west direction, and zero at the mid-directions.

We attribute that trend to the fact that, whereas the radius of the osculating sphere is constant:

$$R = \sqrt{MN},$$

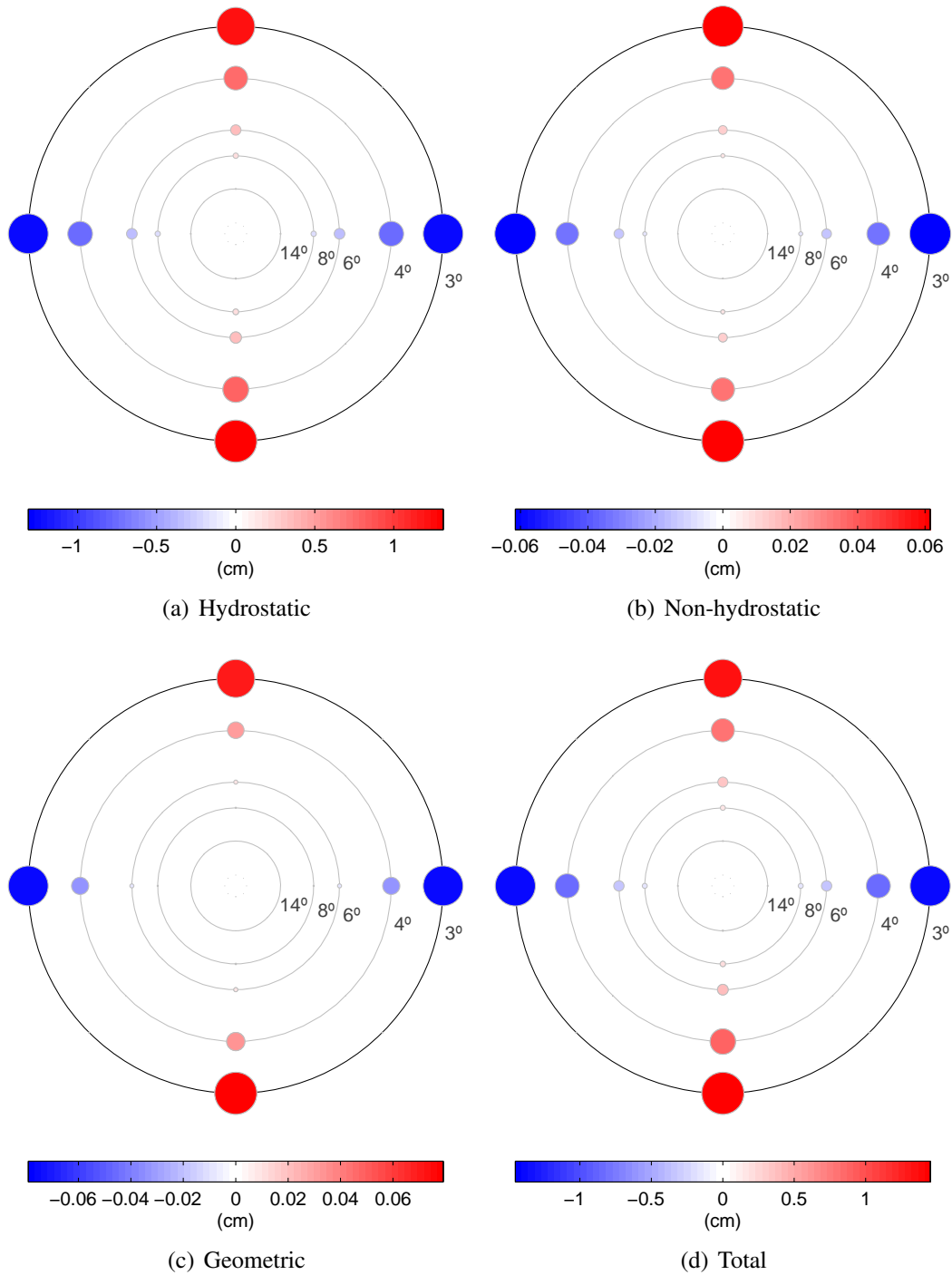


Figure 5.10: Results of second comparison: spherical osculating minus ellipsoidal.

the radius of curvature of the ellipsoid varies with azimuth:

$$R_{\alpha} = \left(\frac{\cos^2 \alpha}{M} + \frac{\sin^2 \alpha}{N} \right)^{-1}.$$

More specifically, in Figure 5.11 we show that the ellipsoidal shell is inside the spherical osculating shell along the north–south rhumb ($R < R_{\alpha} = M$), but it is outside along the east–west rhumb ($R > R_{\alpha} = N$), and the two surfaces intersect in between.

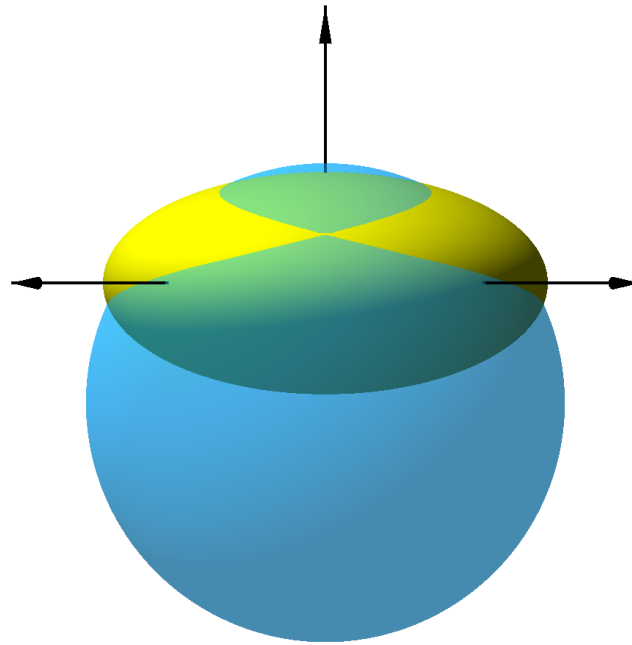
However small is the magnitude of the azimuthal asymmetry exhibited by an ellipsoidal atmosphere, the danger in neglecting it lies in the fact that it is a systematic effect. Therefore it will not be canceled out by data randomization (e.g., processing numerous days of GPS observations); rather, we expect it to persist thus biasing position estimates in an inescapable way. Niell [1996] illustrates that danger:

...if a set of mapping functions has on average a positive bias, then the estimated positions of the stations will be too far from the center of the earth, and a comparison of the reference frame determined by that solution to one using the ‘correct’ mapping function will indicate a scale factor greater than 1.0.

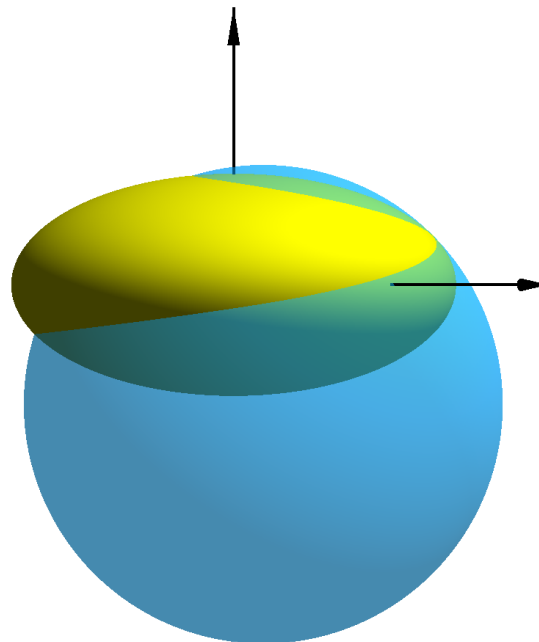
At the risk of being repetitive, let us stress that such a bias would be the spherical osculating atmosphere’s fault, not the ellipsoidal’s — i.e., the latter being more faithful to reality, any discrepancy between the two is due to a deficiency in the former.

The next comparison is that between the ellipsoidal atmosphere and a gradient atmosphere (Figure 5.12). Please recall from section 3.2.4 that we allow for one independent horizontal gradient for each pressure, temperature, and humidity atmospheric field. Consequently, the azimuthal asymmetry exhibited in terms of delay by a gradient atmosphere may be more complex than a single main direction (e.g., SW–NE in Figure 5.12(a)) and may have different direction and magnitude in each delay component (contrast Figures 5.12(a) and 5.12(b)).

It is important to realize that the discrepancy shown in Figure 5.12 is *not* all of the azimuthal asymmetry of a gradient atmosphere: it is only the contribution from a gen-

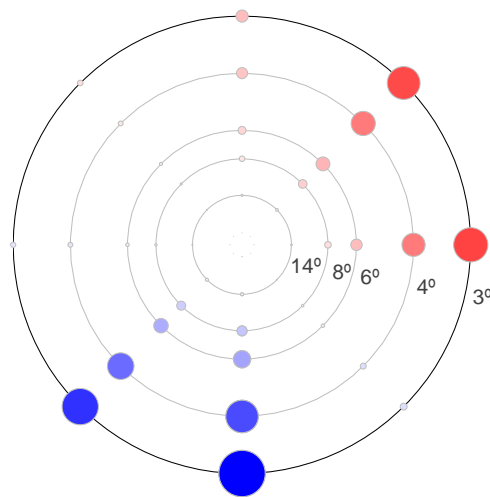


(a) Frontal view.

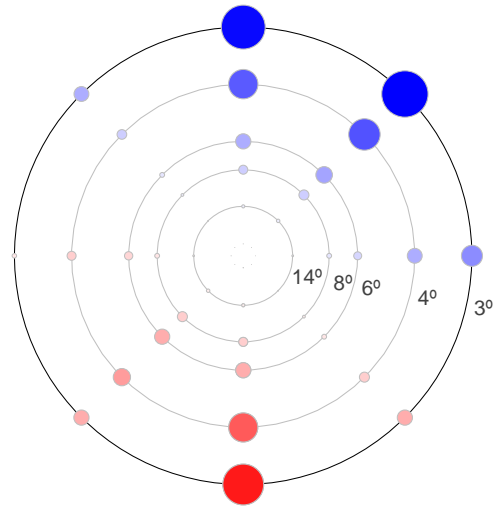


(b) Side view (see Figure 3.3, on p. 61, for the corresponding cross-section).

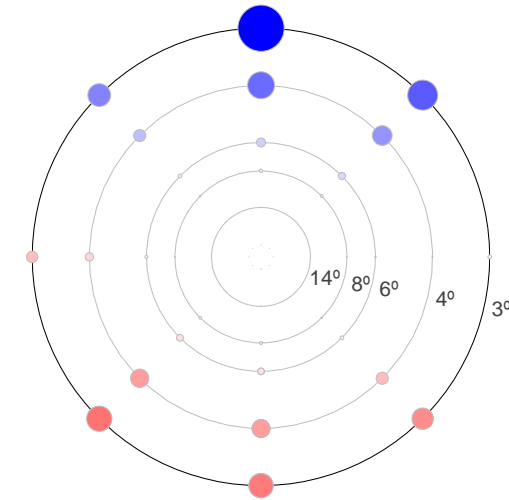
Figure 5.11: Constant radius of osculating sphere (transparent blue shell) vs. azimuth-varying radius of ellipsoid (yellow opaque shell); base point is at $\varphi = 45^\circ$, $\lambda = 45^\circ$; arrows represent axes of global Cartesian coordinate system (X, Y, Z) .



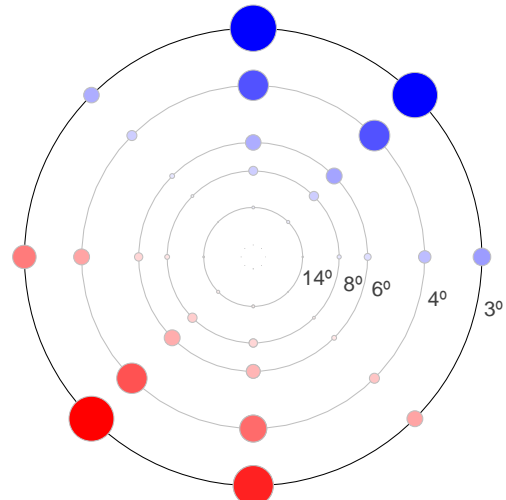
(a) Hydrostatic



(b) Non-hydrostatic



(c) Geometric



(d) Total

Figure 5.12: Results of third comparison: ellipsoidal minus gradient.

uine horizontal gradient acting locally in the atmosphere. To obtain the full azimuthal asymmetry one must add the large-scale characteristics of the ellipsoidal atmosphere upon which the gradient atmosphere is based. We show that in Figure 5.13 (contrast, e.g., Figures 5.12(a) and 5.13(a)).

Following the sequence summarized in Figure 5.2, the last comparison among the comparisons of different atmospheric structures is that between the gradient atmosphere and the 3d atmosphere (Figure 5.14). First we notice that the discrepancy (Figures 5.14(a) to 5.14(d)) tends to be smallest along the main direction previously represented by the gradient atmosphere (e.g., SW–NE in Figure 5.12(a)). That fact implies that the gradient atmosphere does a good job modeling the main direction of azimuthal asymmetry present in the 3d atmosphere. Nonetheless there remains non-negligible, secondary, directions of azimuthal asymmetry present in the 3d atmosphere that the gradient atmosphere is simply not able to capture.

The good agreement between gradient and 3d atmospheres along the main direction of azimuthal asymmetry intriguingly fails at the south portion of the skyplot. I.e., notice how the south portion is significant in the gradient atmosphere (Figure 5.12(a) and does not cancel out in the discrepancy between gradient and 3d atmospheres (Figure 5.14(a)). That might be an indication that a local gradient taken at the base point might not be representative of the conditions too far away from the base point — at such low elevation angles the ray travels several hundred kilometers before exiting the neutral atmosphere (i.e., before it reaches a height of roughly 80 km, as per definition discussed in Appendix I.3.2, p. 166).

Finally, let us say that we expect the discrepancy between gradient and 3d atmospheres to be strongly dependent on the resolution of the NWM. I.e., the lower the resolution is, the smoother the representation of the atmosphere is, and the more similar a 3d atmospheric model is to a gradient one.

Having finished the sequence of comparisons among different atmospheric structures,

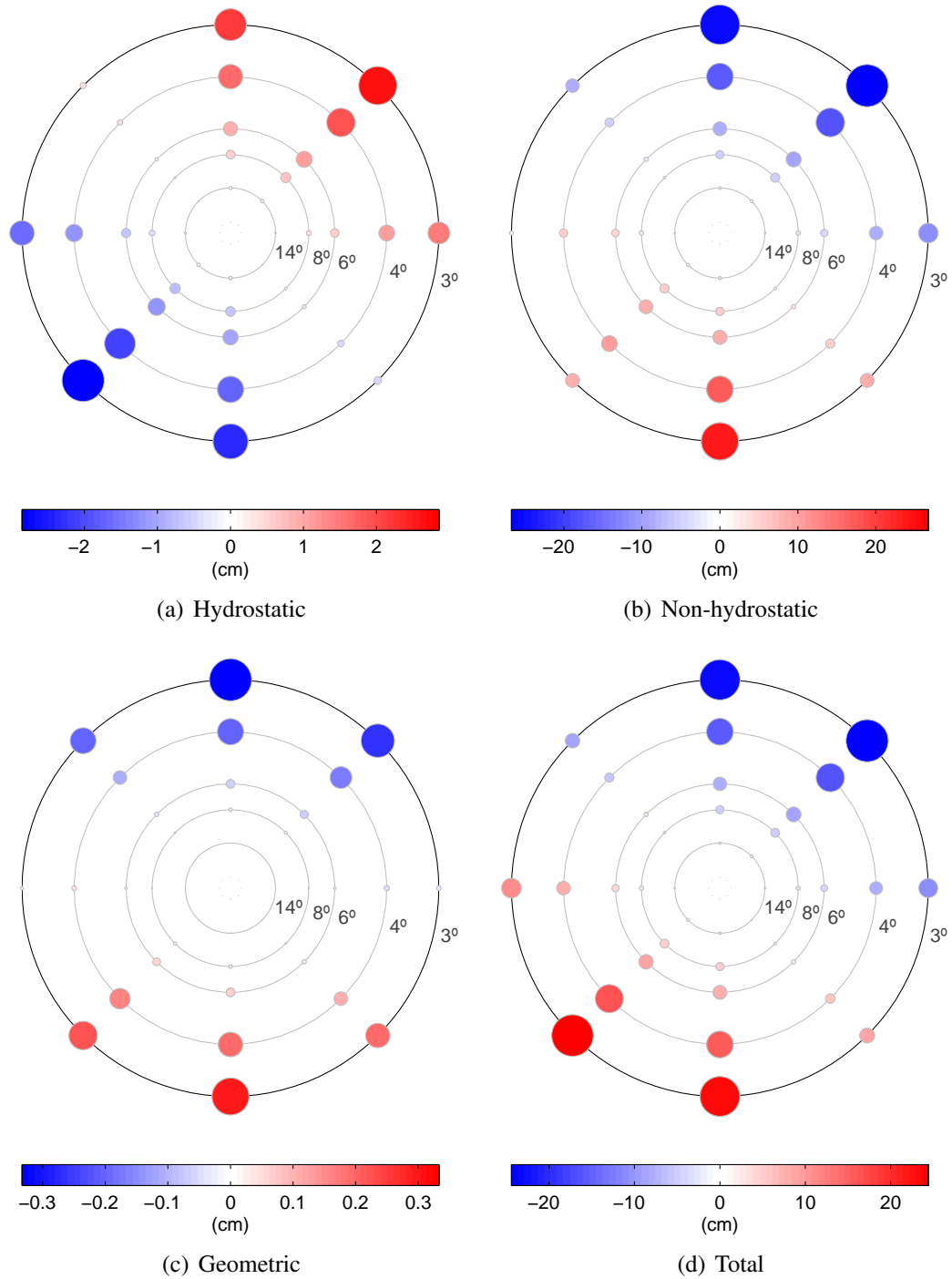


Figure 5.13: Additional results of third comparison: spherical osculating minus gradient.

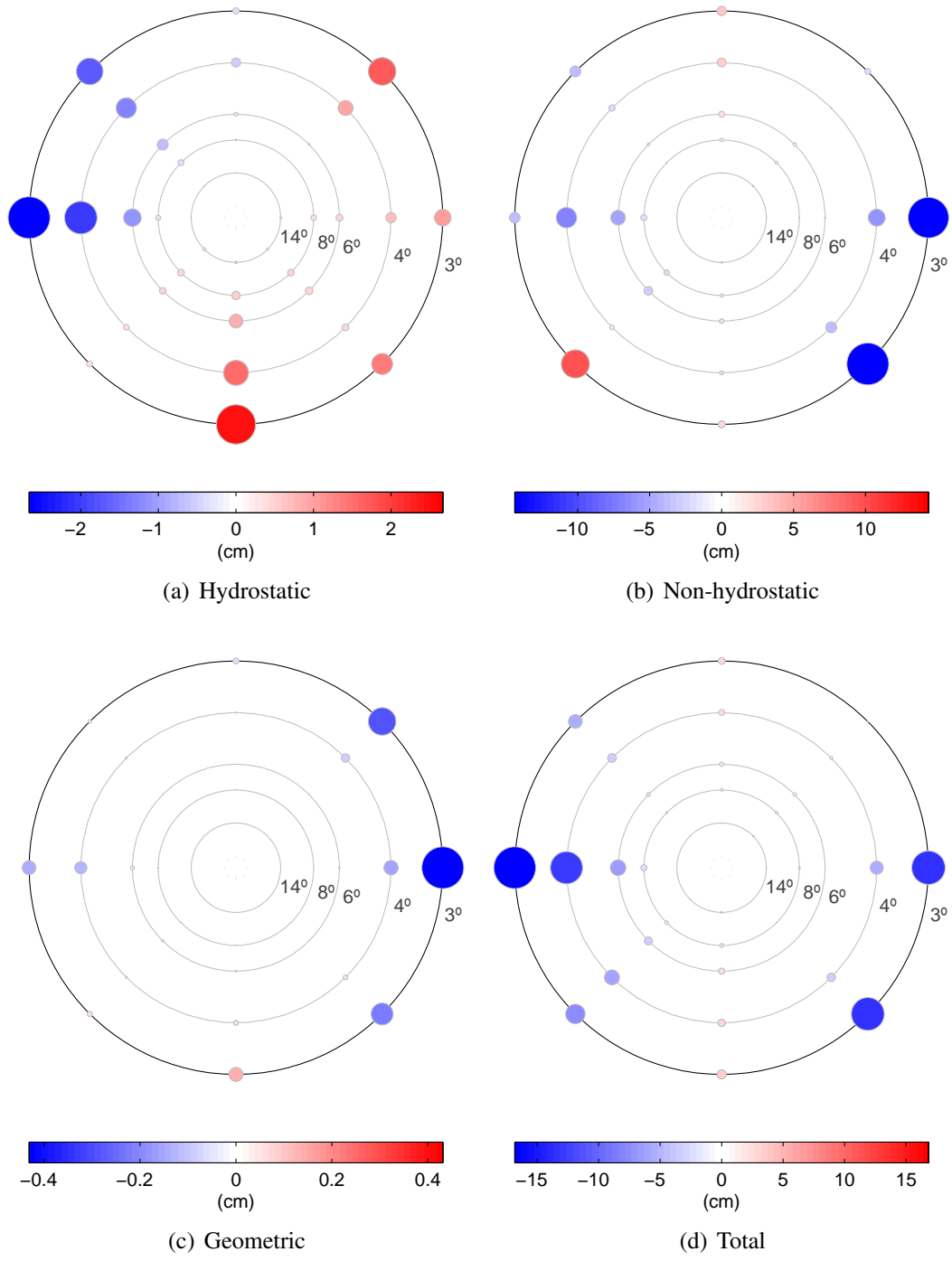


Figure 5.14: Results of fourth comparison: gradient minus 3d.

now we keep the atmospheric structure fixed to the most realistic one (3d), and then we compare the bent-3d and bent-2d ray-path models. What is most striking about their discrepancy, shown in Figure 5.15, is the fact that the sum of the two along-path delays, hydrostatic (Figure 5.15(a)) and non-hydrostatic (Figure 5.15(b)), has nearly the same magnitude and opposite sign as the geometric delay (Figure 5.15(c)). As a consequence, when we add them together to make up the total (Figure 5.15(d)), they nearly cancel each other out.

But the discrepancy in total delay is not exactly zero. At first sight, Figure 5.15(d) may seem completely random. Indeed, it is partially random, due to numerical noise, because the discrepancy values are close to the tolerance set for the numerical integration routine (namely, 0.1 mm). But if we inspect it in detail (Figure 5.16), we realize that the discrepancy in total delay is always positive, *within the above tolerance*. The discrepancy (in the order bent-2d minus bent-3d) being always positive means that the total delay given by the bent-2d ray-path model is always greater than the delay given by the bent-3d ray-path model. Therefore the ray always travels faster with the bent-3d model, a model that allows the ray to bend in whatever way the gradient of refraction directs, instead of forcing the ray to be a plane curve, as the bent-2d model does. In yet another words, the bent-3d model follows more closely Fermat's least time principle, in a 3d atmosphere.

We conclude that the small magnitude of the discrepancies between bent-3d and bent-2d ray-path models, shown in Figure 5.15, warrants the use of the simpler bent-2d ray-path model instead of the more complicated bent-3d one, even if used in conjunction with a 3d atmosphere that does contain horizontal gradients. As a caveat, that conclusion may not be valid with a NWM of higher horizontal resolution than the one we were using (15 km) and for ray-tracing studies interested in products other than the delay (e.g., bending angle).

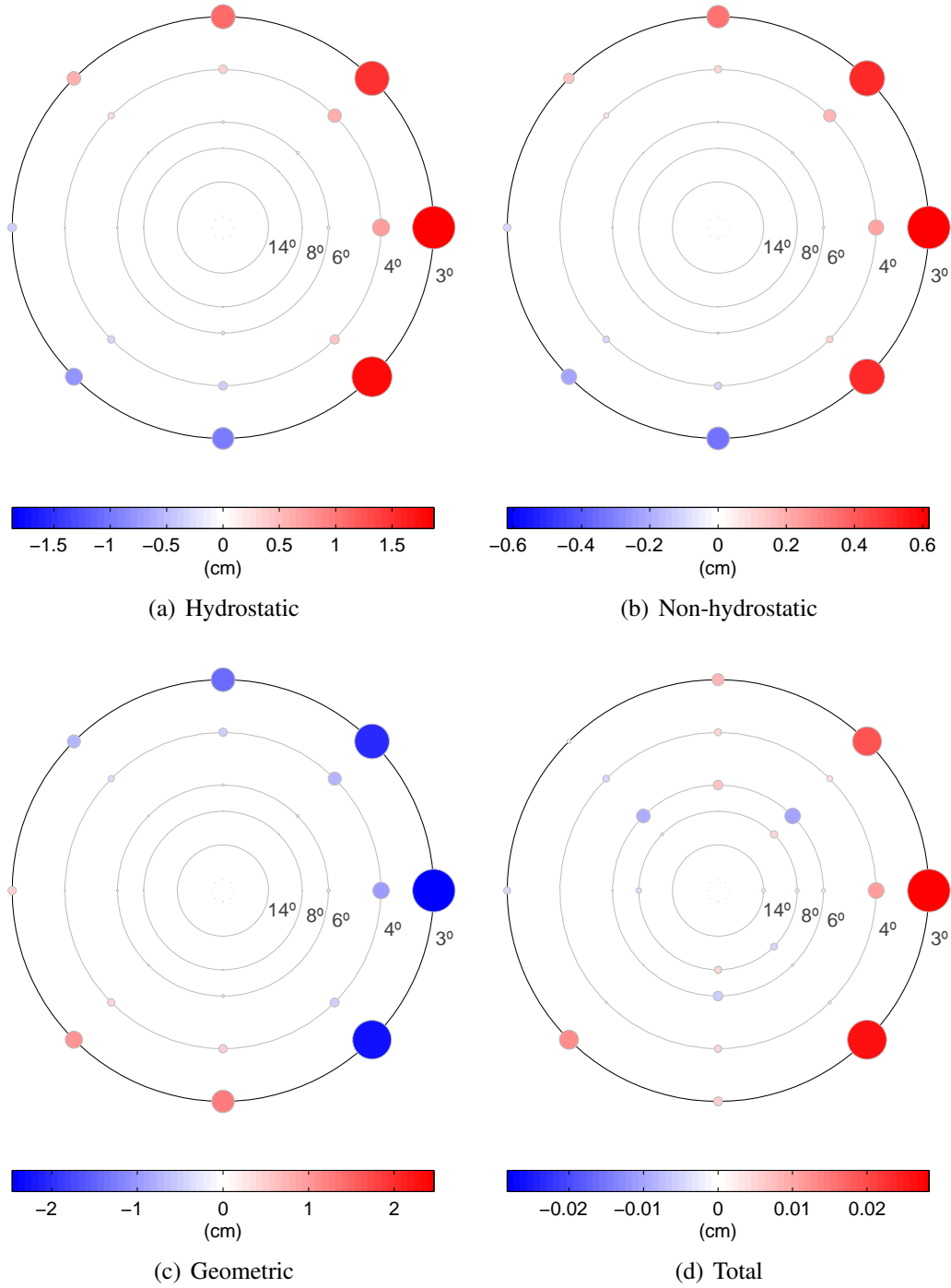


Figure 5.15: Results of fifth comparison: bent-2d minus bent-3d.

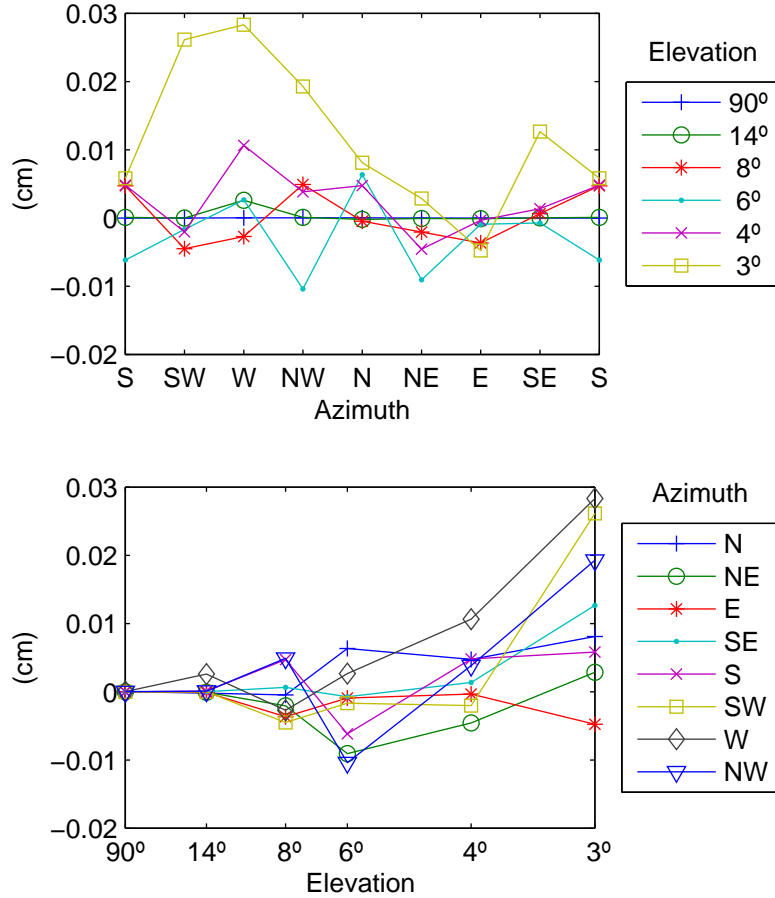


Figure 5.16: Additional results of fifth comparison: bent-2d minus bent-3d (total delay).

5.2 GPS positioning: delay mitigation via prediction only

Having concluded the first experiment (section 5.1), we wished to assess the performance of the same delay models in GPS positioning. For this assessment, we designed two experiments: a preliminary one, in which the delay is mitigated by predicting it only; and a more thorough experiment, in which the delay is mitigated by estimating it.³ We shall describe the preliminary experiment in the present section; the thorough one is described in the next section.

We were specially interested in corroborating or refuting the hypothesis that the spher-

³We use the names “preliminary” and “thorough” to recall the fact that prediction is less accurate a mitigation technique than estimation, as discussed in section 2.3.1 and section 4.2, p. 77.

ical concentric atmosphere is a poor model; the discrepancies with respect to a more realistic spherical osculating atmosphere, obtained in section 5.1, were so large (metre-level) that they should be evident in the GPS observation residuals. Contrarily, the difference among some of the other models was so subtle that we did not expect it to be detectable in this preliminary experiment. Therefore, we set aside the ellipsoidal and gradient models and kept only the spherical concentric, spherical osculating, and 3d models.

All of the slant delay models above share the same zenith delay; they differ only in the slant factor model. Therefore we may say that we are assessing different mapping functions. As a reference for comparison, we included a state-of-art mapping function (Boehm et al.'s [2006a] GMF), coupled with the same zenith delay model (NWM) used in our mapping functions.

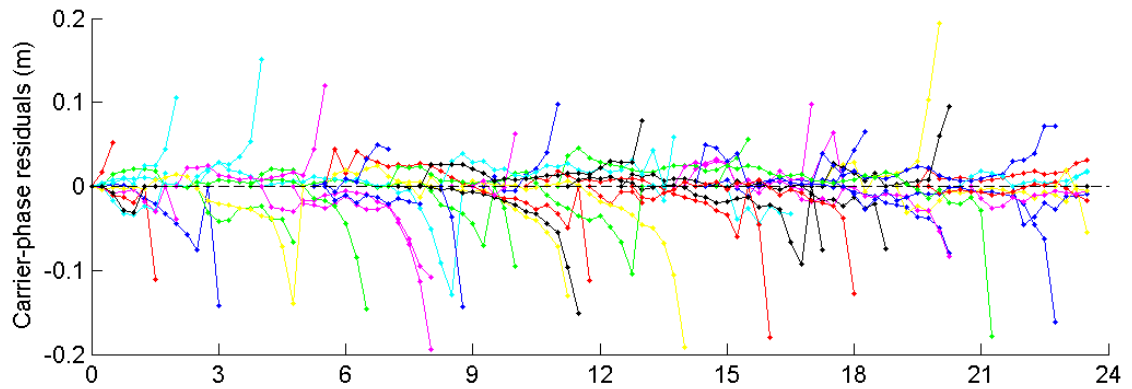
We wanted any mis-modeled delay to be reflected, as much as possible, directly in the observation residuals; in other words, we wanted to prevent as much as possible extraneous parameters from soaking up any mis-modeled delay. Therefore we did not estimate receiver position (we kept it fixed to its IGS determination) nor zenith tropospheric delay; we estimated only receiver clock error and satellite ambiguities. Ionospheric delay was taken care of through the use of dual-frequency observations. Since the delay is the harder to model the lower is the elevation angle, we lowered the cut-off elevation angle, from the usual $10\text{--}15^\circ$ to zero. That does not artificially exacerbate the impact of mis-modeled tropospheric delay, because we weighted the observations' standard deviation by the factor $1/\sin \epsilon$.

We selected an IGS station within the spatial extents of the NWM we were using, with low multipath, and collecting data below 10 degrees of elevation angle: station ALGO, located in Algonquin Park, Canada. We processed only one day of data at the sampling interval of 15 min; the date was chosen arbitrarily as July 28, 2007. The processing mode is precise point positioning (as implemented in the GAPS software [Leandro et al., 2007]);

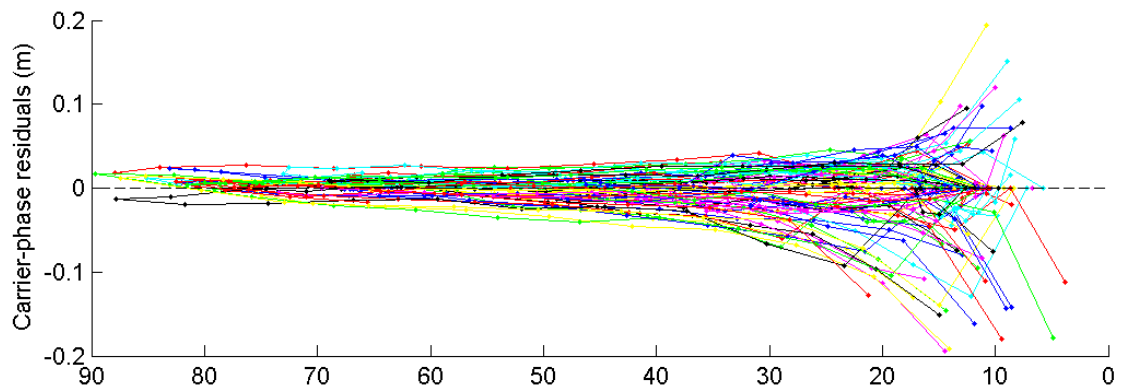
the estimation procedure is carried out sequentially, epoch-by-epoch in a forward-only filter; therefore observations at subsequent epochs are not exploited in the estimation of current epoch's parameters.

The final results are residuals binned by elevation angle; more specifically, their mean and standard deviation per bin. We operated on carrier-phase residuals only, because pseudo-range residuals are inherently too noisy for our purposes here. Before we show final results, though, we would like to briefly describe how they were obtained, including the important decision of whether to discard any data. We will illustrate intermediary results for only one mapping function; rest assured that the final results include all mapping functions. We started with individual residuals for each satellite, at each epoch; Figure 5.17(a) shows the residuals versus time, while Figure 5.17(b) shows the same residuals versus elevation angle. It is apparent the occurrence of larger residuals for setting satellites, both as lines branching off in time (Figure 5.17(a)) and the fan-shaped tail at lower elevation angles (Figure 5.17(b)).

Why we do not notice rising satellites, too, in the carrier-phase residuals? The answer, in a nutshell, is because a satellite ambiguity is initially allowed by its a priori standard deviation to soak up all of that satellite's misclosure, leaving almost nothing of the misclosure to be left as residual. In equations: starting with ℓ , $\tilde{\mathbf{x}}$, \mathbf{C}_x , and \mathbf{A} (respectively, the vector of observations, the vector of approximate parameters, the a priori covariance matrix of the parameters, and the design matrix), then we define $\mathbf{w} = \mathbf{A}\tilde{\mathbf{x}} - \ell$ as the misclosure, part of which is absorbed by the estimated parameters $\hat{\mathbf{x}} = (\mathbf{A}^T\mathbf{A} + \mathbf{C}_x^{-1})^{-1}\mathbf{A}^T\mathbf{w}$. The rest of the misclosure \mathbf{w} , unaccounted for by the estimated parameters $\hat{\mathbf{x}}$ (notice that $\hat{\mathbf{x}}$ is not the same as $\tilde{\mathbf{x}}$), is left over as estimated residual $\hat{\mathbf{r}} = -(\mathbf{A}\hat{\mathbf{x}} + \mathbf{w})$. The ambiguities a priori standard deviation is very large initially (i.e., at the beginning of the tracking session or just after a satellite rises or a cycle slip occurs), but it gets smaller and smaller as the session elapses and the estimated parameters decorrelate. The decorrelation occurs



(a) Residuals vs. time (h).



(b) Residuals vs. elevation angle (degrees).

Figure 5.17: Raw residuals (spherical concentric atmosphere).

due to the accumulation of additional observations (especially, observations collected in different receiver–satellite directions) and also because of constraints applied to some of the parameters.

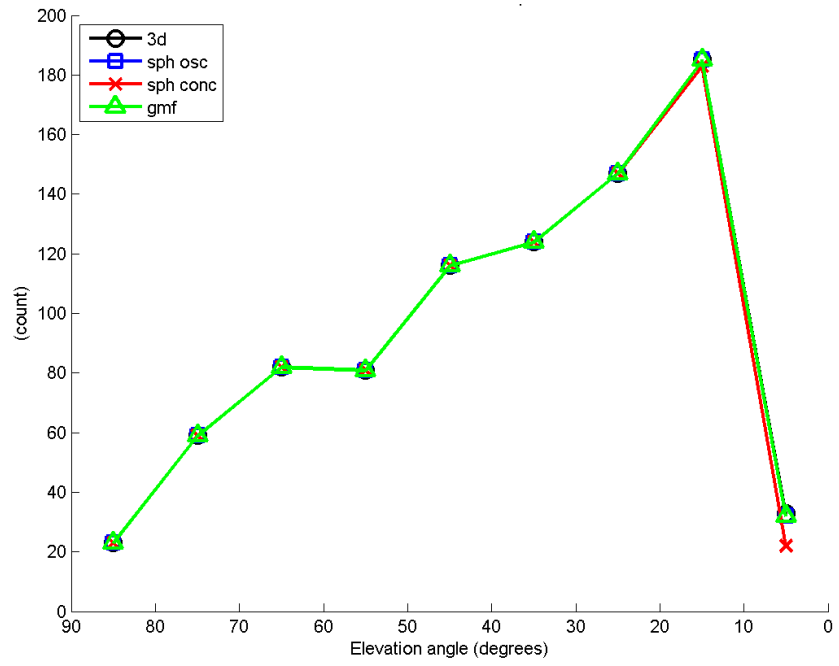
For the statistics shown below, we decided to discard the residuals before ambiguities have converged, as those residuals are unrealistically small. Ideally, we would have adopted as a cut-off criterion the ambiguities' standard deviation. As such an information was not readily available, we adopted instead the following criterion: discard residuals up to one hour after the occurrence of a zero-valued residual (which marks an ambiguity start or reset). The period of one hour usually provides sufficient variation in the receiver–satellite geometry for ambiguities to converge [Leandro, 2008]. Figures 5.18(a) and 5.18(b) show, respectively, the number of residuals per elevation angle bin, before and after the discard; most discarded residuals occur at lower elevation angles. The residuals remaining after the discard are shown in Figures 5.19 to 5.21, for all mapping functions.

Figure 5.23 shows the main result of the present assessment, namely, the standard deviation of residuals binned by elevation angle (each bin spans 10° in elevation angle), as well as the limits $[\sigma^2] < \sigma^2 < [\sigma^2]$ of a 95% confidence interval for the standard deviation of each bin, shown as error bars. The lower $[\sigma^2]$ and upper $[\sigma^2]$ limits are computed according to:

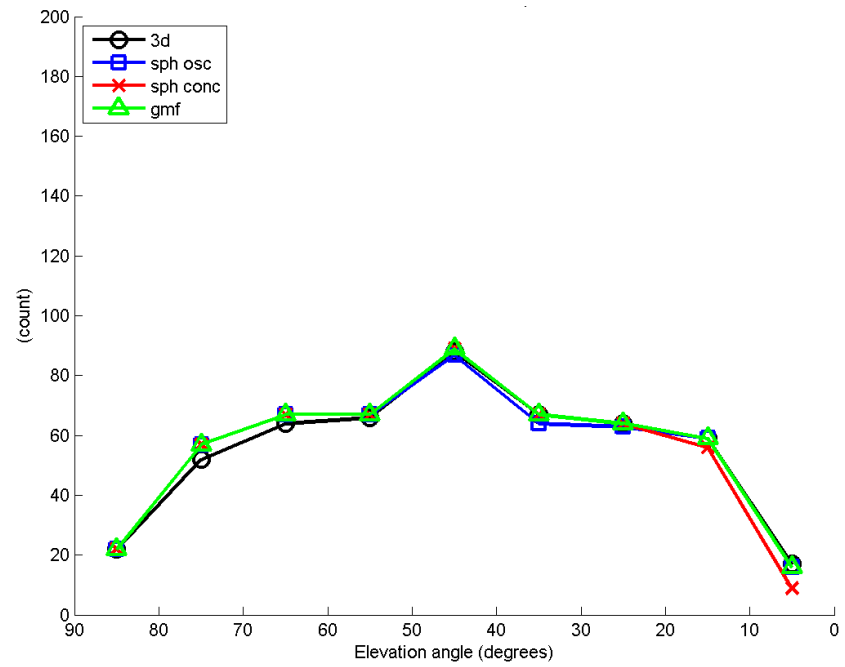
$$[\sigma^2] = s^2 \nu / P_{\chi^2}^{-1}([l], \nu), \quad (5.1a)$$

$$[\sigma^2] = s^2 \nu / P_{\chi^2}^{-1}([\!l], \nu), \quad (5.1b)$$

where σ^2 is the unknown population variance (expected to be between the computed limits), s^2 is the sample variance, $P_{\chi^2}^{-1}()$ is the χ^2 inverse cumulative distribution function, ν is the degrees-of-freedom ($\nu = n - 2$, where n is the number of residuals in a bin and 2 comes from the fact that we are estimating both the mean and variance from the sam-



(a) Before discard (see text).



(b) After discard (see text).

Figure 5.18: Number of observations per bin.

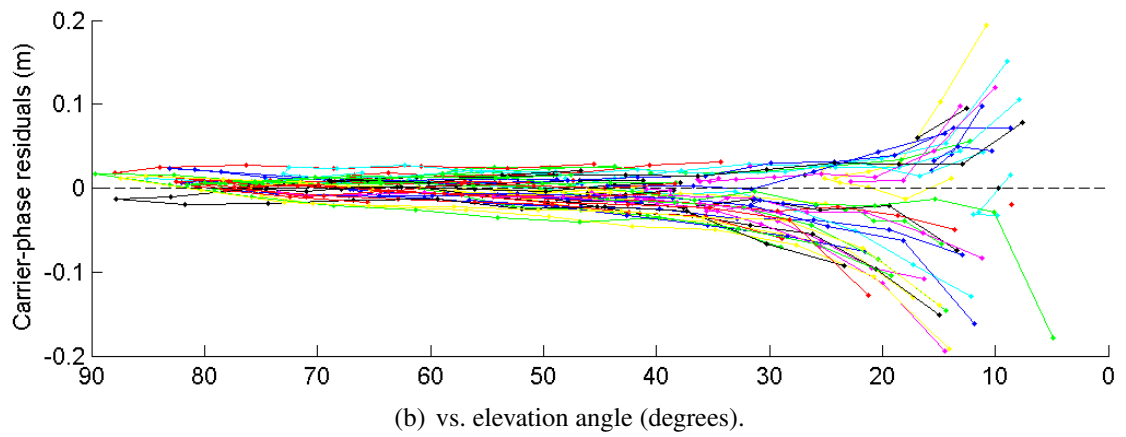
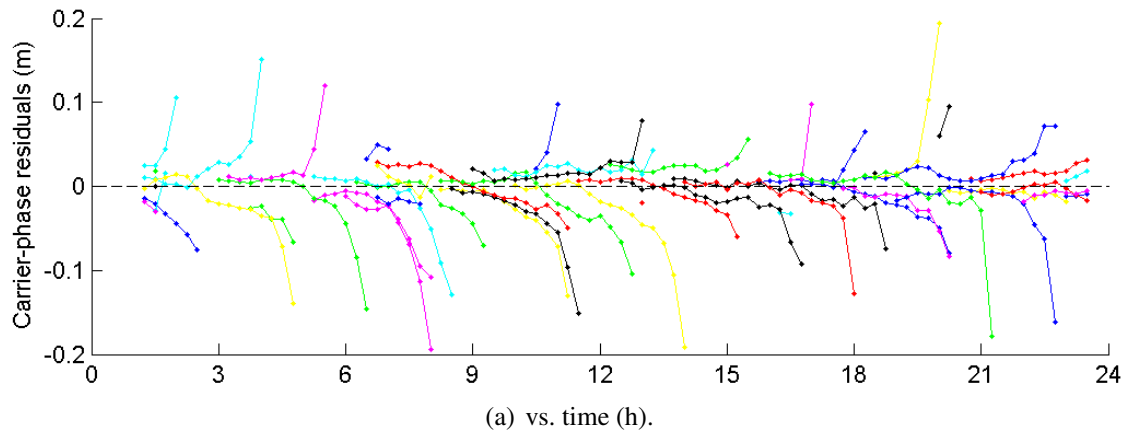


Figure 5.19: Remaining residuals after discard (spherical concentric atmosphere).

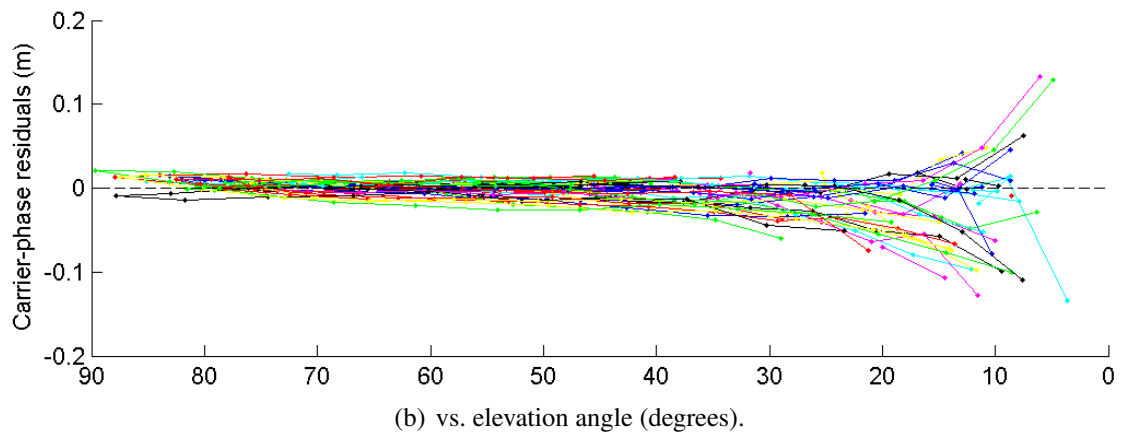
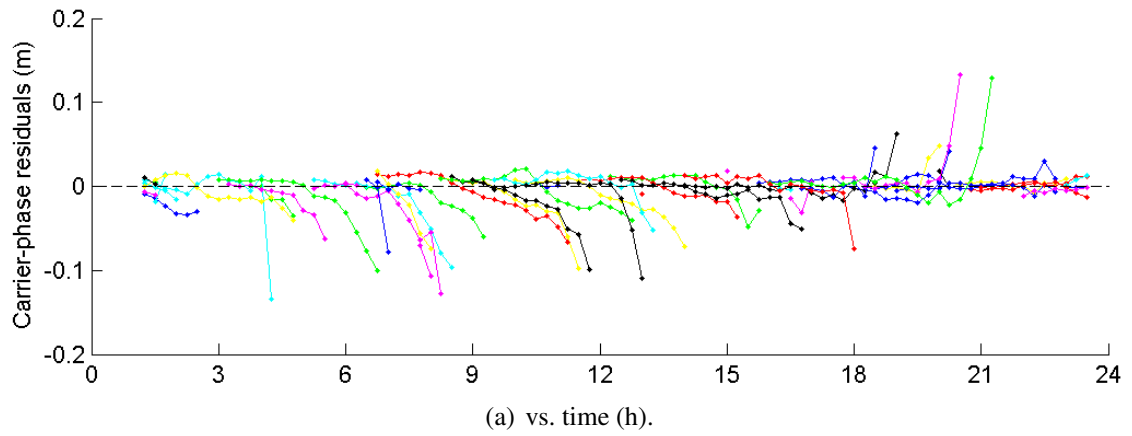


Figure 5.20: Remaining residuals after discard (spherical osculating atmosphere).

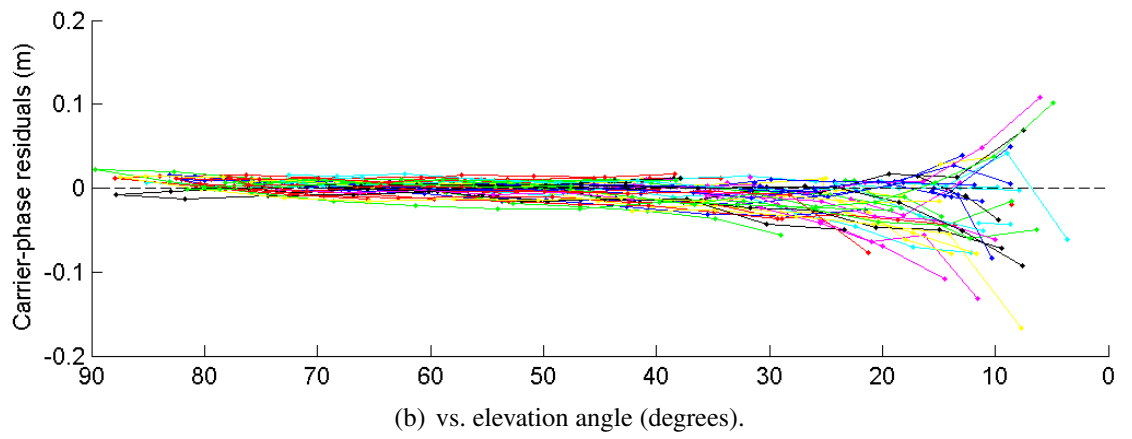
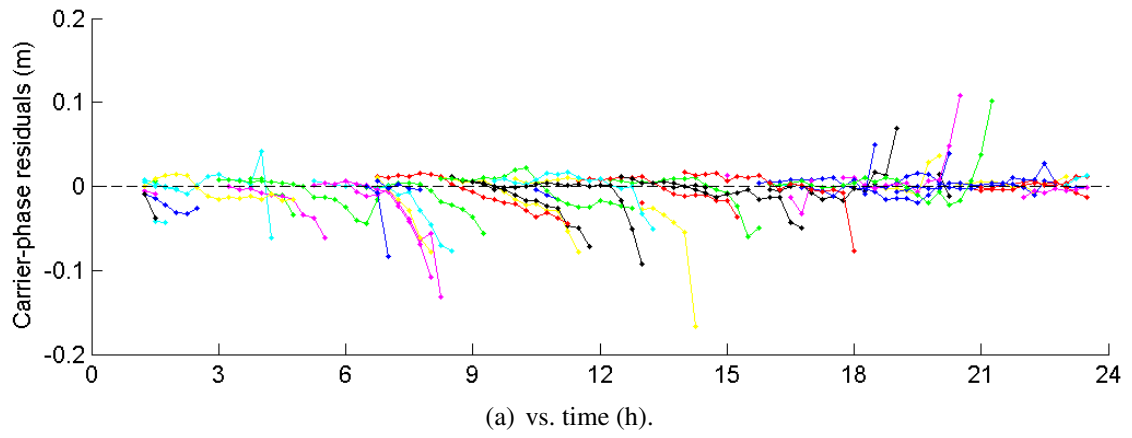


Figure 5.21: Remaining residuals after discard (3d atmosphere).

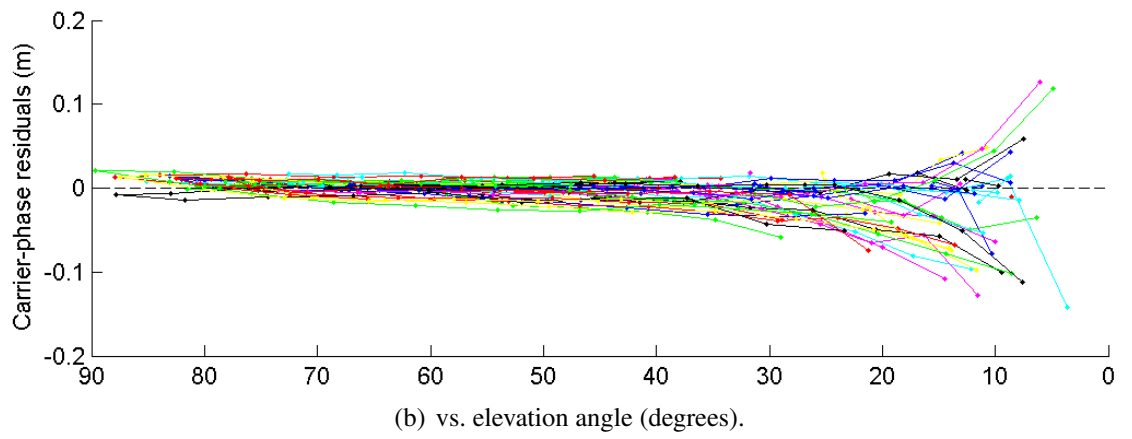
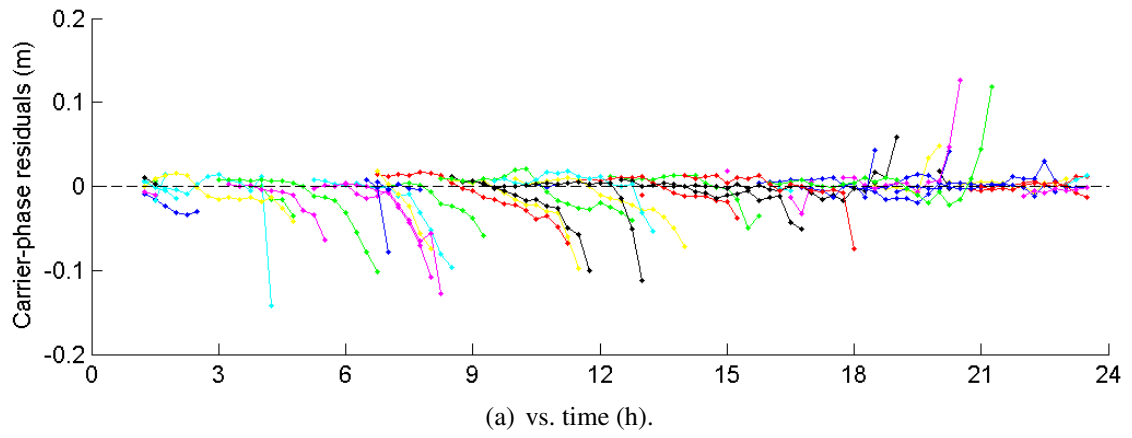


Figure 5.22: Remaining residuals after discard (GMF).

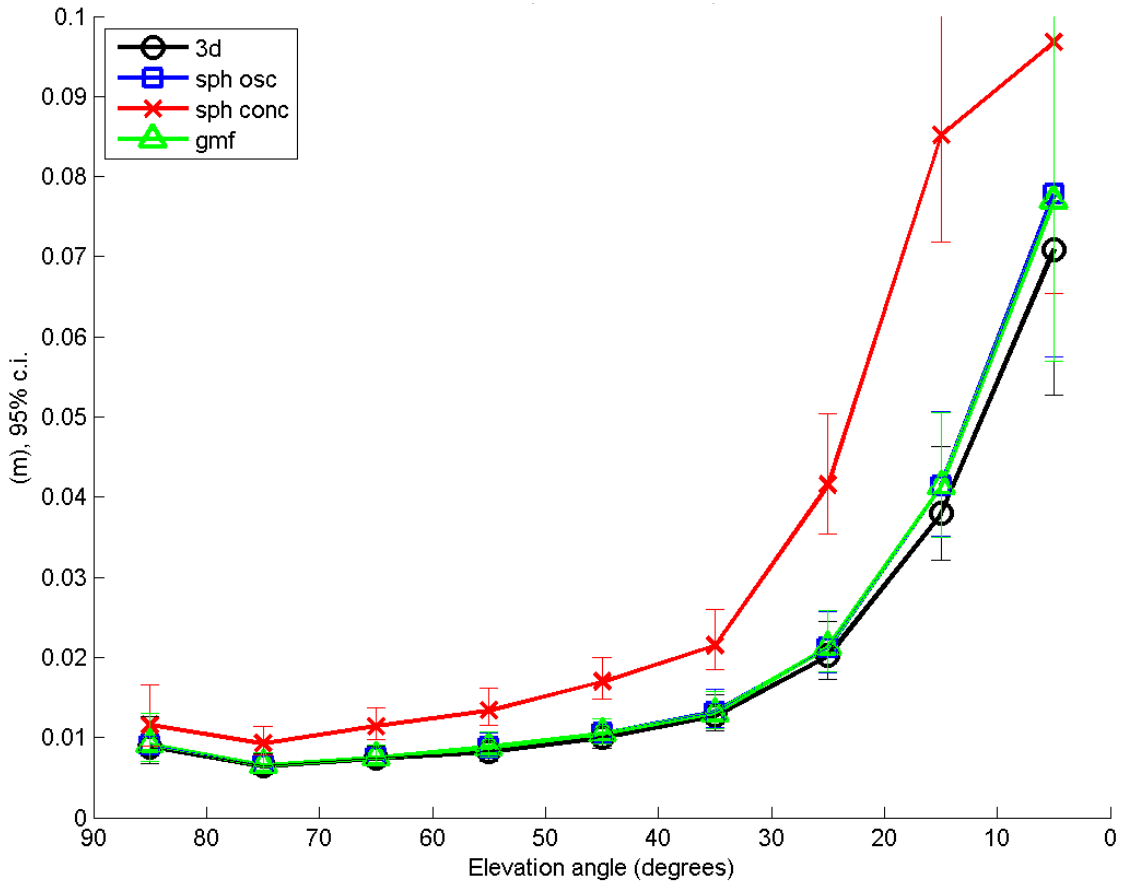


Figure 5.23: Standard deviation of residuals per bin, as well as their respective 95% confidence interval (shown as error bars).

ple), and $[l] = s$, $\lceil l \rceil = 1 - s$ are, respectively, the lower and upper limits of a two-tailed confidence interval with confidence level $c = 95\%$ and significance level $s = (1 - c)/2$.

It stands out clearly in Figure 5.23 that the spherical concentric atmospheric model is significantly worse than the other models. We interpret those results as strong evidence that the Earth’s oblateness — from which the distinction between spherical osculating and spherical concentric atmospheres originates — has a significant impact in mapping functions. The distinction among the other models is not detectable in the present experiment; that is certainly the case between our spherical osculating model and GMF, and to a lesser extent also between those two and the 3d model — notice how their error bars greatly

overlap. In section 5.3 we describe a more thorough attempt to detect the subtle difference among those other models.

With that we end the main results for the present experiment. In the remainder of the present section we discuss a secondary result for the present experiment. Even though we do not expect that secondary result to provide further evidence of the distinction among the mapping functions, we show it for completeness and also as a check on the experiment as a whole. The secondary result is the mean of residuals binned by elevation angle, shown in Figure 5.24, as well as the limits $\lfloor \mu \rfloor < \mu < \lceil \mu \rceil$ of a 95% confidence interval for the mean of each bin, shown as error bars. The lower $\lfloor \mu \rfloor$ and upper $\lceil \mu \rceil$ limits are computed according to:

$$\lfloor \mu \rfloor = \bar{x} - \sqrt{s^2/nP_t^{-1}(\lfloor l \rfloor, \mu)} \quad (5.2a)$$

$$\lceil \mu \rceil = \bar{x} + \sqrt{s^2/nP_t^{-1}(\lceil l \rceil, \mu)} \quad (5.2b)$$

where μ is the unknown population mean (expected to be between the computed limits), \bar{x} is the sample mean, $P_t^{-1}()$ is Student's inverse cumulative distribution function, and the rest is as in eq. (5.1).

The first aspect to notice in Figure 5.24 is the non-zero value for the mean residuals at most bins. At first sight, non-zero mean residuals might seem contrary to what is expected for residuals — after all, the principle of least squares guarantees zero-mean residuals. At closer look, though, we realize that the condition of zero-mean residuals applies only to all residuals as a whole, not to subsets of them, like the subsets we form after binning residuals by elevation angle.

Nonetheless, it remains intriguing the existence of an almost linear trend in the mean residuals per bin shown in Figure 5.24; one might even suspect that such a trend is a consequence of unaccounted for systematic error sources. In the remainder of this section, we

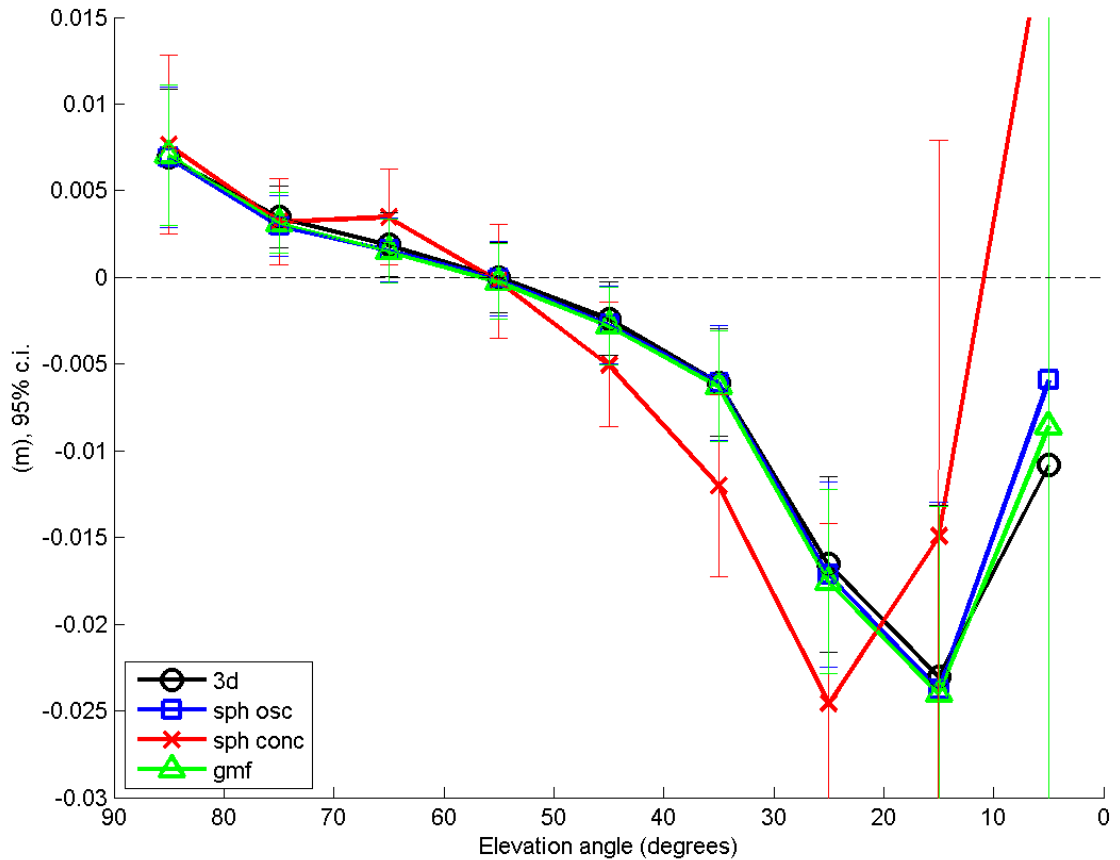


Figure 5.24: Mean of residuals per bin, as well as their respective 95% confidence interval (shown as error bars).

would like to explain how that trend can be replicated. We will be speaking of misclosures and residuals, as defined on p. 105, as well as error, especially (tropospheric) slant delay error, here defined as modeled slant delay (predicted or estimated) minus true or actual slant delay (usually unknown).

Assume that the observations are affected by a slant delay error (i.e., a slant delay unaccounted for by the prediction or estimation model). The slant delay error could be either due to a zenith delay error or a slant factor error (or both, obviously); when we do not estimate a zenith delay parameter, as in the present experiment, the zenith delay error tends to dominate over the slant factor error; so we will consider only zenith delay errors here. Let us further assume a typical zenith delay error of 1 cm. The corresponding slant

delay error, obtained by multiplying 1 cm by the slant factor, is shown in Figure 5.25.

Comparing the mean residuals that we wish to replicate (Figure 5.24) with the assumed slant delay error that we intend to serve as a replica (Figure 5.25), we notice the following matches and mismatches:⁴ (i) the absolute value of both mean residuals and slant delay errors increase with decreasing elevation angle; (ii) direction is reversed in the two plots: slant delay errors “go up”, mean residuals “go down”; (iii) slant delay error has constant sign, while mean residuals are both positive and negative (the zero crossing is near 55 degrees); (iv) trend in errors is $1/\sin \epsilon$; trend in mean residuals is close to linear between 90 and 30 degrees, then decreases more rapidly up until 20 degrees. Now let us see how we can correct each of the mismatches above to obtain a better replica for the mean residuals per bin.

The mismatch mentioned under (ii) is easily taken care of by reversing the sign of the assumed zenith delay error (-1 cm); see Figure 5.26: now slant delay errors “go down”, just like the mean residuals (Figure 5.27) do.

Mismatch (iii) can be explained mainly by the estimation of receiver clock error: any non-zero mean slant error across all satellites at the same epoch will be estimated as part of receiver clock error, as discussed in section 4.1, p. 69. If we take the mean of the slant delay errors (Figure 5.26), weighted by the number of observations per elevation bin (Figure 5.18(b)), we obtain the value -2.08 cm; the leaking of that mean slant delay error into the receiver clock error pushes up the plot in Figure 5.26, bringing to zero the bin with most observations. So now there are both negative and positive values in our replica, just like in the original mean residuals per bin (Figure 5.27).

Mismatch (iii) can be further improved by considering the weighting applied to the observations; the observations’ standard deviation was multiplied by the factor $1/\sin \epsilon$. Computing the weighted mean of slant delay errors (weighted at this time by both the

⁴We ignore the two lowest elevation bins because their extremely large uncertainty disallow any meaningful analysis.

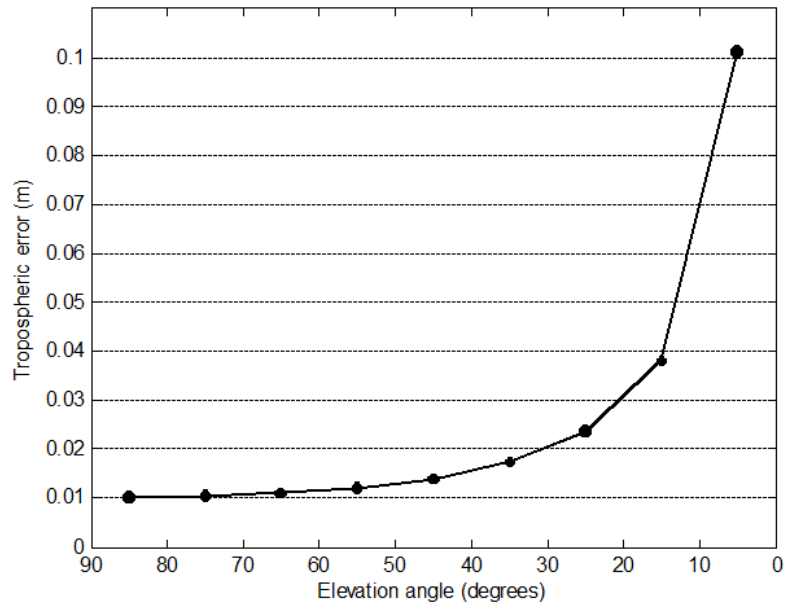


Figure 5.25: Slant delay error due to nominal zenith delay error (1 cm).

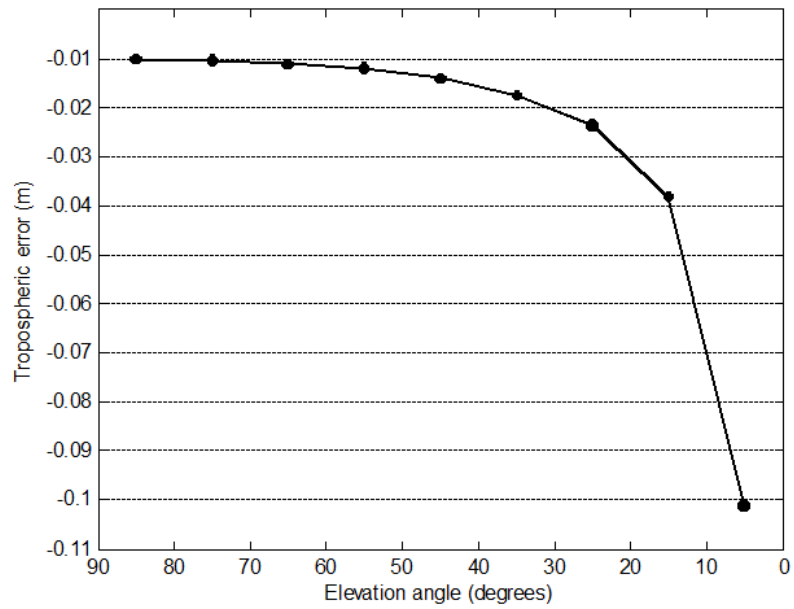


Figure 5.26: Slant delay error due to nominal zenith delay error (-1 cm).

number of observations per bin (Figure 5.18(b)) and their standard deviation) we obtain the value -1.31 cm for the receiver clock error. That pushes up the plot in Figure 5.26 a little differently, so that now the intersection between the zero line and the plot in Figure 5.28 matches better with that intersection in Figure 5.24.

Mismatch (*iv*) is more laborious to be accounted for. We take a number of measures to make the replica more realistic: instead of assuming a nominal zenith delay error, we take the zenith delay error as equal to the difference between its estimated value (obtained from a separate GPS processing) and its predicted value (used in the present experiment) — see Figure 5.29; instead of taking the slant factors at the mid-elevation angle per bin, we take the slant factors at the exact elevation angle at which observations were made — the resulting slant delay errors are shown in Figure 5.30; instead of estimating a single clock error throughout the session, we will estimate a different clock error per epoch — then we obtain the replica of individual residuals shown in Figure 5.31.

Figure 5.32 shows the final replica of mean residuals per bin, obtained from the replica of individual residuals shown in Figure 5.31. Figure 5.32 is a good match for the original Figure 5.24, especially considering that we have neglected all but one possible error source: zenith delay error. The fact that we were able to explain the perhaps intriguing trend in the secondary results (Figure 5.24), represents an extra assurance on the main results (Figure 5.23) and on the present experiment as a whole.

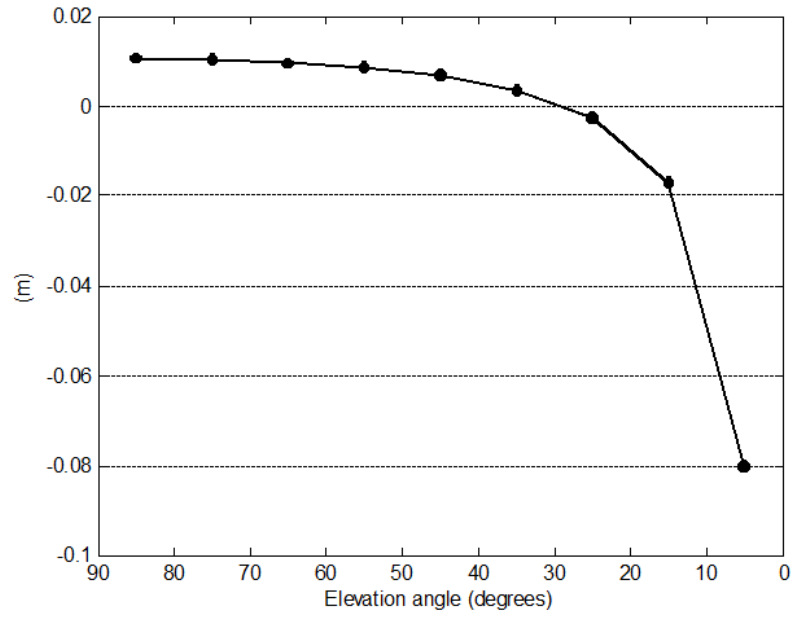


Figure 5.27: Replicated residuals (see text for details).

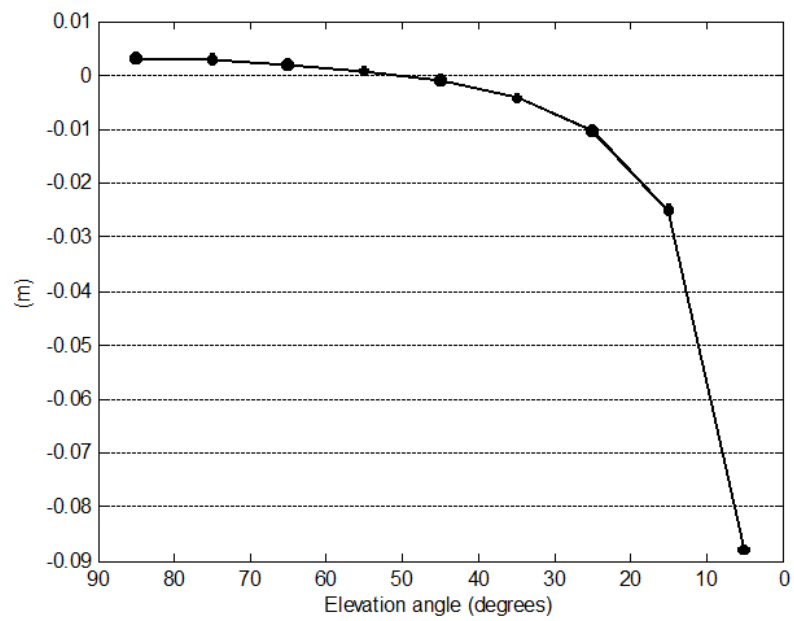


Figure 5.28: Replicated residuals (see text for details).

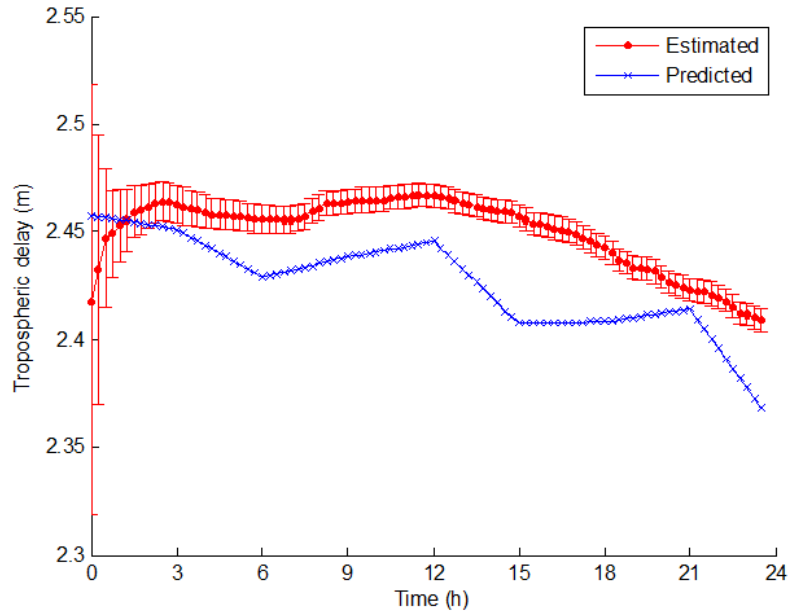


Figure 5.29: Estimated and predicted zenith delay.

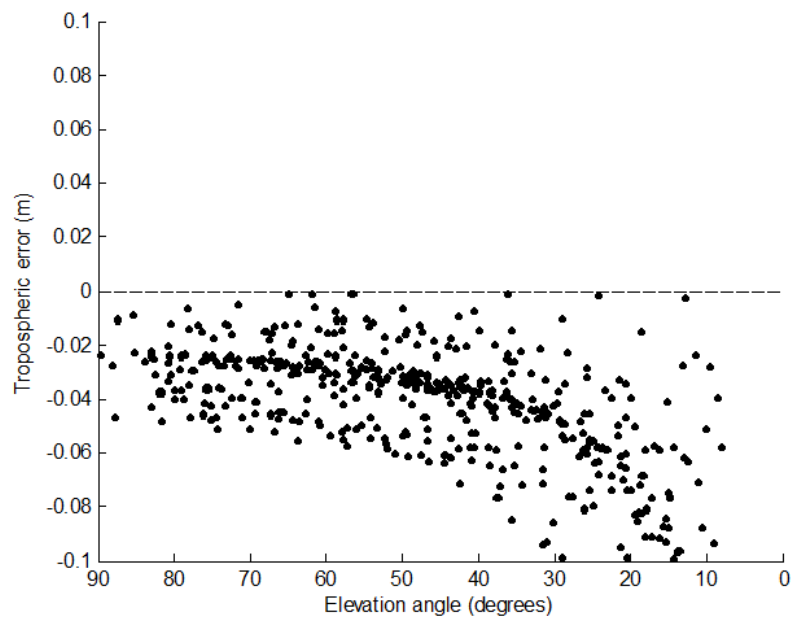


Figure 5.30: Slant delay error due to realistic zenith delay error.

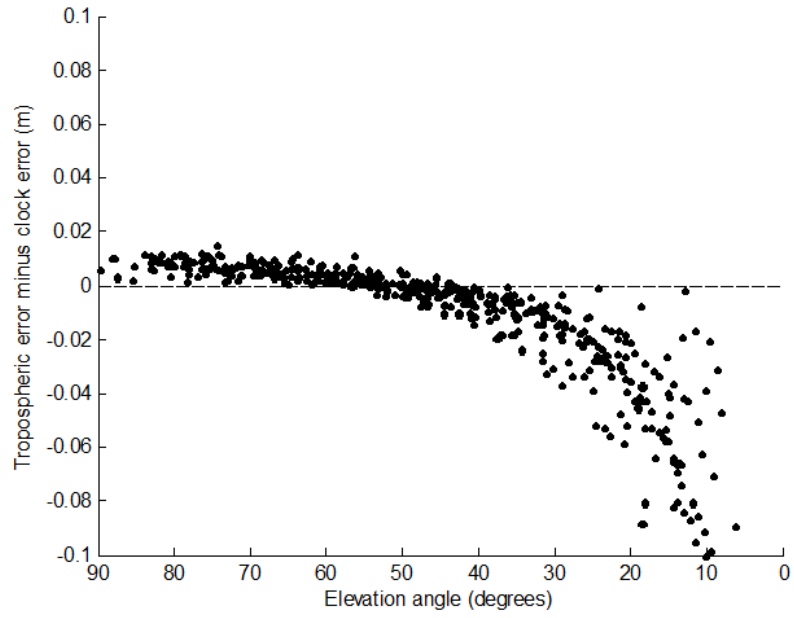


Figure 5.31: Replicated residuals (see text for details).

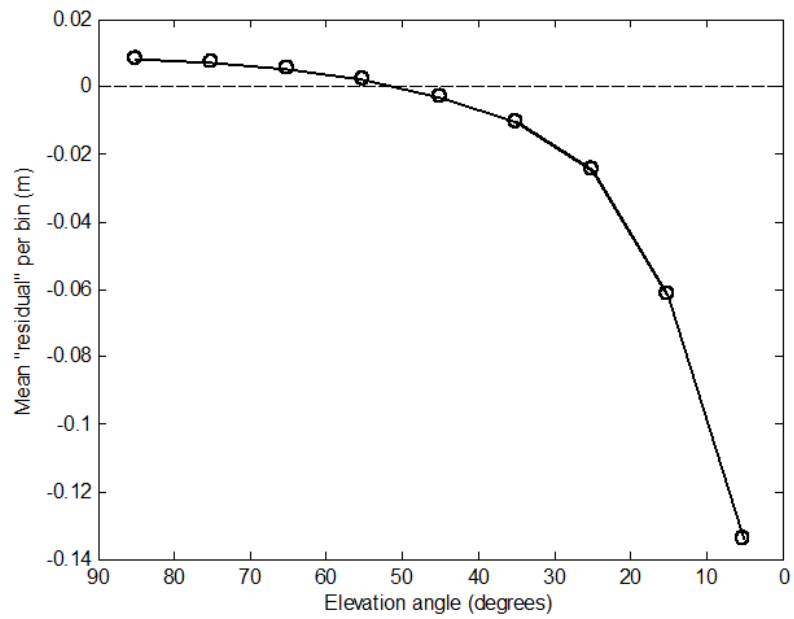


Figure 5.32: Replicated residuals (see text for details).

5.3 GPS positioning: delay mitigation via estimation

In the experiment described in section 5.2 we have mitigated the delay only by predicting it. In the present experiment we mitigate the delay mainly by estimating it. As discussed in section 2.4, the latter is a superior mitigation technique (provided its requirements are fulfilled, as discussed in section 4.2). That superiority is the reason why we called the previous experiment preliminary and the present one “more thorough”.

We follow the delay separation motivated in section 2.5: we predict the large, well-behaved component of delay (the hydrostatic one), both its zenith delay and slant factors; as for the more variable component (non-hydrostatic), we predict only its slant factors and estimate its zenith delay.

The prediction of zenith hydrostatic delay was performed using the UNB3m model [Leandro et al., 2006a, 2008]. This model employs a look-up table, based on a low-resolution climatology, to obtain the pressure value at the receiver position, which is then input into Saastamoinen’s [1972] model (as revised by Davis [1986]), to obtain the corresponding zenith hydrostatic delay. The estimation of zenith non-hydrostatic delay was modeled stochastically, letting that parameter free to vary from epoch to epoch, under a prescribed arbitrary yet reasonable (based on practice) variance rate of $(5 \text{ mm})^2$ per hour.

The tacit assumption underlying the handling of zenith delays described above is that the zenith hydrostatic delay is sufficiently well predicted so that the remaining zenith delay is mostly non-hydrostatic; in the event of that assumption breaking down, the estimated zenith delay parameter would represent all of the non-hydrostatic delay plus some of hydrostatic delay, corresponding to the error in the hydrostatic zenith delay prediction model; that unwanted mixing of both components in a single parameter would lead to an error in slant delay, due to the fact that one would be inconsistently multiplying a portion of the hydrostatic delay with a non-hydrostatic mapping function [Tregoning and Herring, 2006].

Another difference with the experiment described in section 5.2 is that, instead of constraining the receiver position to its IGS solution, in the present experiment we estimate it. In total, we estimate receiver clock, satellite ambiguities, receiver position, and zenith non-hydrostatic delay. The estimation procedure is still carried out sequentially, epoch-by-epoch in a forward-only filter. The receiver position solution that we take for analysis is the one given at the end of the session, obtained after accumulating the entire session's data; hence we speak of static, instead of kinematic, positioning.

A few more processing details. We processed data collected at 30s sampling interval, decimated to 15min (to reduce the processing load), over the course of a day (chosen arbitrarily as July 28, 2007), at the IGS station ALGO (which is within the spatial extents of the NWM we were using, suffers low multi-path, and logs data below 10 degrees of elevation angle). Ionospheric delay was taken care of through the use of dual-frequency observations, forming the so-called iono-delay-free combination. The cut-off elevation angle was set to zero, and the observations' standard deviation were weighted by the factor $1/\sin \epsilon$. The processing mode is precise point positioning (as implemented in the GAPS software [Leandro et al., 2007]).

It is important to realize that, if on one hand the estimation of additional parameters certainly yields a reduction in the observation residuals, on the other hand it might inadvertently result in biased parameter estimates. In other words, surely additional parameters represent an additional opportunity for effects embedded in the misclosures to be absorbed by parameters before being left over as residuals (see similar discussion on p. 105). But the danger lies in that absorption taking place in extraneous parameters, unrelated to the error source affecting the observations. For example, a genuine vertical displacement due to ocean or atmospheric loading, if not corrected with prediction models nor allowed to be reflected in the receiver position, will end up corrupting estimates of parameters correlated with vertical position, such as zenith delay.

To assess the impact of different mapping functions, instead of inspecting the residuals as done in the previous experiment, we will be inspecting the estimated position and zenith delay; more specifically, the *bias* in those estimated parameters. The definition of a bias requires a benchmark or reference solution; ideally such a reference should be significantly more accurate than any of the test solutions (against which it is being compared), so that it is safe to attribute any discrepancy between the two to an error in the less accurate test solution. We have strived to find such a reliable reference, but we have to admit that, at the accuracy levels we are discussing in this experiment (from tenths to tens of mm), that is no easy task.

The reference solution for the receiver position is the IGS weekly solution, which itself is a combination of individual solutions submitted by several processing groups (more or less independent — some share common stations and the same software in their solutions). The IGS also offers a weekly cumulative solution, which is a combination of the (non-cumulative) weekly solution with the long-term, historic solution. Why did we choose the weekly solution, instead of the cumulative one? The reason lies in the well known fact that estimated positions of IGS stations exhibit variations at, e.g., annual, seasonal periods. The cumulative solution filters out most of that variation; the resulting stability is highly desirable if the objective is the establishment of a reference frame. Yet if the objective is to make a comparison with test solutions that are expected to be affected by the same variations, we better choose a reference solution that intentionally includes those same systematic variations. That was the rationale for our preference for the IGS (non-cumulative) weekly solution.

We were unable to find a reliable external reference solution for the zenith delay. The options considered were radiosondes and water vapor radiometers. Radiosondes have excellent vertical resolution, but poor temporal and horizontal resolution: they are launched only every 12 hours, at scattered locations, hundreds of kilometers apart. Radiometers

have their own limitations, such as contamination due to liquid water (i.e., they do not work reliably when it rains) and noisy data. Surely one could try harder to get more reliable values from a radiometer, but much research has already been devoted to that topic without definitive solutions — otherwise radiometers would be more widely used. Also, in case we had opted for an external reference solution for the zenith delay, we would have to deal with the fact that the estimated zenith delay, contrary to the predicted zenith delay, is not meant to be equal to the slant delay in the zenith direction. Instead, the estimated zenith delay is more correctly interpreted as a weighted average of the slant delays affecting the observations collected [Wolfe and Gutman, 2000; Gutman et al., 2003b].⁵ Given the lack of an external reference for the delay, we looked at the discrepancies among solutions, without attributing that discrepancy to any one solution alone. Such a simpler comparison still allows us to draw useful conclusions, as presented below.

We have included in the present experiment combinations of the bent-2d ray-path with each of the atmospheric structures discussed in section 5.1, namely, spherical concentric, spherical osculating, ellipsoidal, gradient, and 3d. Based on the theory laid down in Chapter 3 and also on the results from prior experiments (sections 5.1 and 5.2), our expectation was that the 3d mapping function would give the best results, and the spherical concentric one the worst results; we also expected that the spherical concentric and the ellipsoidal ones would be indistinguishable.

For comparison, we have also included the following mapping functions from other authors: NMF [Niell, 1996], GMF [Boehm et al., 2006a], and VMF1 Site [Boehm et al., 2006b]. All of the them, to the best of our knowledge, employ a bent-2d ray-path model

⁵Were the sky covered uniformly by satellites, that weighted average would comprise the portion of the atmosphere contained in an inverted cone, having tip at the receiver and aperture given by the cut-off elevation angle. In actuality, the weighted average is biased because the characteristics of the GPS constellation yield a non-uniform sampling of the sky (e.g., northern hole in the sky of high latitude stations).

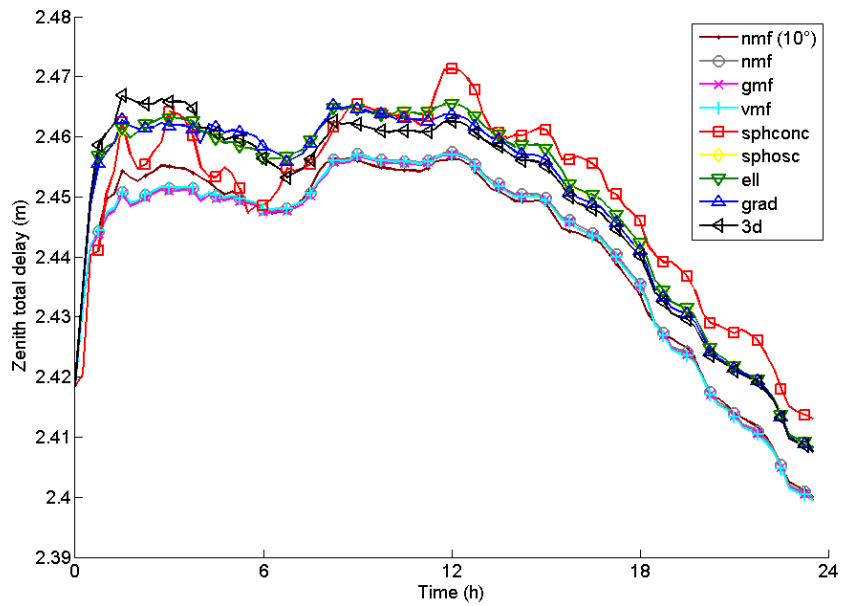
and a spherical atmospheric structure⁶; their main difference, therefore, lies in the atmospheric source. NMF uses as atmospheric source the Supplements to U.S. Standard Atmosphere; it is perhaps the most widely used mapping function, due to both its accuracy and convenience for not depending on actual surface atmospheric conditions⁷ GMF is intended as an update to NMF, based on a more accurate atmospheric source: the state-of-art global NWM developed at the European Centre for Medium-Range Weather Forecasts (ECMWF); to keep GMF as a convenient replacement for NMF and a backup for VMF1 Site, its final temporal and spatial resolution is intentionally not as high as its atmospheric source would allow. Finally, VMF1 Site uses the same atmospheric source as GMF (ECMWF), but at this time at full resolution; the final mapping function was developed specifically at a fixed set of pre-defined sites. Bear in mind that, although VMF1 Site and the mapping functions developed in this work employ state-of-art NWM at full resolution, the NWM employed by VMF1 Site is a global one, whose resolution is not as high as that of the regional NWM used in this work (15 km).

Results from this experiment are summarized in Table 5.2 (position parameters) and in Figure 5.33 (zenith delay parameter). Regarding the position results, the values in Table 5.2 are discrepancies of the final estimated position (static positioning mode) with respect to the reference position solution. As for the zenith delay, the values in Figure 5.33(a) are the total zenith delay itself (the sum of the estimated non-hydrostatic and predicted hydrostatic components), while in Figure 5.33(b) we have the discrepancies of each solution with respect to the 3d solution.

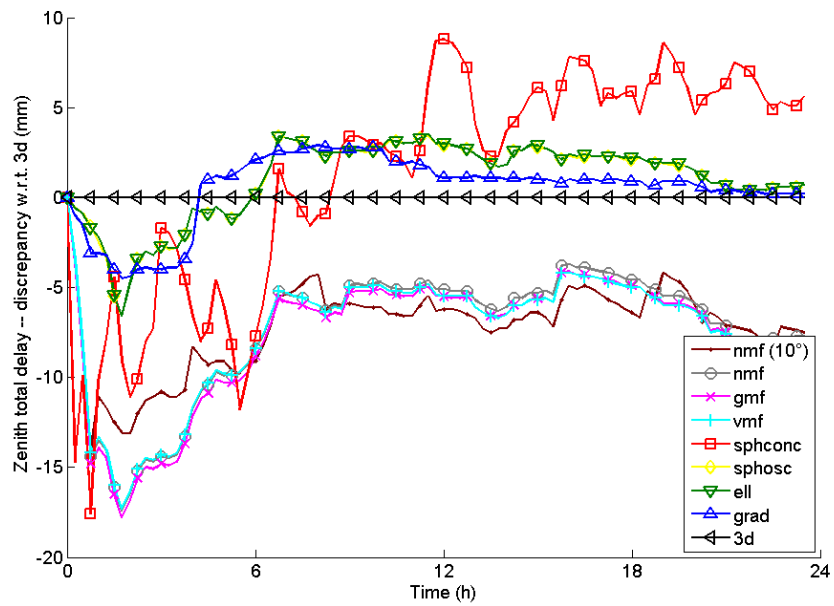
The first thing to notice in Table 5.2 is the very large bias in the north component for the spherical concentric model. That is yet another manifestation of the tilting of the

⁶More specifically, a spherical osculating atmospheric structure, following our usage of those mapping functions, in which we input elevation angles reckoned from the ellipsoidal horizon — see discussion on p. 91.

⁷Using the terminology introduced in section 2.8, NMF is a ready-to-use mapping function, as it does not depend on proxy variables.



(a) Zenith delay itself (non-hydrostatic estimated portion plus hydrostatic predicted portion).



(b) Discrepancy in zenith non-hydrostatic delay with respect to that estimated using 3d model.

Figure 5.33: Zenith delay.

Table 5.2: Bias in static position estimates using various mapping functions.

Model	North (mm)	East (mm)	Up (mm)
NMF (10°)	2.2	-3.0	21.7
NMF	2.4	-2.3	17.5
GMF	2.4	-2.3	18.1
VMF1 Site	2.4	-2.3	18.2
spherical osculating	2.5	-1.9	2.5
ellipsoidal	2.5	-1.9	2.5
gradient	-0.1	-1.3	4.3
3d	0.3	-0.5	4.9
spherical concentric	33.0	-1.6	-3.0

spherical concentric horizon with respect to the ellipsoidal horizon (from which elevation angles are reckoned). We have already seen its nasty consequences in terms of slant delays in section 5.1 (metre-level discrepancies) and in terms of GPS observation residuals in section 5.2. The east component is not as badly affected, which is expected according to the way that the spherical concentric horizon is tilted: the east–west direction acts as an axis of rotation for the tilting along the north–south direction (see Figures 5.8 and 5.9 on p. 92 and 93). The up component is surprisingly not as bad as one might expect, which might seem highly suspicious at first sight. At closer look, though, we have to recall the vulnerability inherent in any estimation of correlated parameters (section 4.1). I.e., even though the vertical position is not too bad, its correlated cousin, the zenith delay, is severely biased: inspecting Figure 5.33(b), it is apparent that the zenith delay solution from a spherical concentric atmosphere deviates significantly from any other solution.

A second result to highlight in Table 5.2 is that denoted as “NMF (10°)”. It is a solution processed with a cut-off elevation angle of 10°, which is more usual than the one we have adopted in all the other cases (zero degrees). Our intention with the inclusion of NMF (10°) is to demonstrate that our processing settings do not unrealistically deteriorate the results, so as to, e.g., artificially exacerbate the difference among different mapping functions. That is demonstrated by the fact that the results from NMF (zero degrees) are no worse

than those from NMF (10°).⁸

A third point to discuss is the likeness among results of NMF, GMF, and VMF1 Site, both in positioning (Table 5.2) and zenith delay (Figure 5.33). We expected that at least VMF1 Site would stand out from the other mapping functions, as it is generated specifically for the receiver location analyzed. We speculate that the difference among these models might clear up with sessions longer than the single day we have analyzed.

A fourth point shown in Table 5.2 is the improvement in the up component, by almost one order of magnitude, of the spherical osculating model over NMF, GMF, and VMF1 Site. Accordingly, their respective zenith delay solutions have a consistent 1 cm bias. What difference between these two groups of mapping functions could explain the improvement found in the up component? Recalling from p. 125, both groups share the same ray-path model (bent-2d) and atmospheric structure (spherical osculating); their main difference lies in the atmospheric source: even though both use a NWM, in this work we use a regional, high-resolution NWM (15 km horizontal resolution), while VMF1 Site uses a global, mid-resolution NWM (ECMWF). Therefore we are tempted to conclude that higher resolution NWM have the potential to offer significant improvements in mapping functions. A definite conclusion would require an analysis involving longer periods and more numerous stations; we strongly recommend further comparisons in that regard, as future work.

A fifth point concerns the ellipsoidal model. Even though we expect it to be more realistic than the spherical osculating one, the reality is that the smallness of the Earth's oblateness makes their difference insignificant, as shown in both Table 5.2 and Figure 5.33. A more positive way of wording the same conclusion is saying that the spherical osculating case does an excellent job as a surrogate for the more rigorous ellipsoidal case, in the

⁸At hindsight, it also demonstrates that our slight misuse of NMF, GMF, and VMF1 — developed for observations down to 3° in elevation angle, but employed here down to 0° — has no impact, probably due to lack of actual observations between 0 and 3° .

vicinity of a specific base point. Again, even though the short observation period in this experiment prevents us from dismissing that possibility, we would be surprised if analyses of longer periods revealed the ellipsoidal model significantly different than the spherical osculating one. Of course, the ellipsoidal model might still be useful in planetary atmospheres other than the Earth's (e.g., Saturn's flattening is 30 times that of the Earth).

A sixth remark is the improvement in horizontal position, from the spherical osculating model to the 3d one: the horizontal bias is reduced from the few-mm level to sub-mm (Table 5.2). That is a very encouraging, although very preliminary, result about the benefits of the most rigorous 3d atmospheric structure over the simpler, profile-based, spherical osculating atmosphere, typically employed in mapping functions. Again, a definite conclusion would require an analysis involving longer periods and more numerous stations; we strongly recommend further comparisons in that direction, as future work.

A seventh point concerns the lack of improvement in the vertical position, from the spherical osculating model to the 3d one, expected to accompany the improvement found in the horizontal position. That unexpected result made us cast a doubt over the reference or benchmark solution, upon which all the biases shown in Table 5.2 are defined. That doubt is reasonable because there remains an unaccounted for error source known to cause variations in the vertical position at the few-mm level:⁹ that error source is atmospheric and oceanic loading, which we now consider with some detail. The displacement caused by loading of the Earth's enveloping fluids upon the Earth's crust has tidal and non-tidal components. The tidal component, albeit larger than the non-tidal one, tends to average out over the course of a day, as the greatest tidal frequencies are daily and twice-daily. Between atmospheric and oceanic non-tidal loading, the former is greater. Therefore, for our day-long observation session, we expect a bias mostly from non-tidal atmospheric pressure loading [Tregoning and van Dam, 2005; Steigenberger et al., 2008]. Figure 5.34

⁹We expect that the current reference solution is still reliable for the comparisons in points first to sixth above.

shows the vertical displacement due to non-tidal atmospheric pressure loading, over the course of the week of July 22–28, 2007 (recall that our reference solution is computed with a week of data), at station ALGO.¹⁰ The peak-to-peak variation is 6.8 mm; the corresponding correction would shift all solutions up or down with respect to the reference solution, possibly decreasing and increasing what we now see as larger and smaller biases in the test solutions. For example, if the true reference height is actually 6.8 mm upwards of where we have it currently, then the bias of the 3d solution would decrease, from the current 4.9 mm to a smaller value of -1.9 mm; in contrast, the bias of the spherical osculating solution would increase, from the current 2.5 mm to a greater value of -4.3 mm (greater in absolute value). Please note that we are not actually correcting our test solution for loading; that would require great care in, e.g., defining the zero-loading epoch and also accounting for the geocenter motion, as loading displacements are given with respect to the center of mass of the entire Earth, solid plus fluids. What we are doing is suggesting that a variation of this order of magnitude plays a role in explaining why the expected best model (3d) does not show results with the smallest bias in the vertical position. For the horizontal position, results are as expected: 3d model shows the smallest bias; not coincidentally, loading does not affect horizontal positions as severely.

The eighth and last remark concerns the gradient atmosphere. As expected, its results show biases with values in between those of spherical osculating and 3d atmospheres, both in positioning (Table 5.2) and zenith delay (Figure 5.33). This is consistent with the fact that the gradient atmosphere has intermediary complexity, between the single profile-based atmospheres (of which the spherical osculating is an example), and the full 3d model. In other words, a gradient atmosphere — whose formulation is based on a profile of horizontal gradients obtained at a particular base location and thereafter postulated the same everywhere else — is better than neglecting horizontal gradients entirely, as the

¹⁰Provided by the Goddard VLBI group at <http://gemini.gsfc.nasa.gov/aplo>

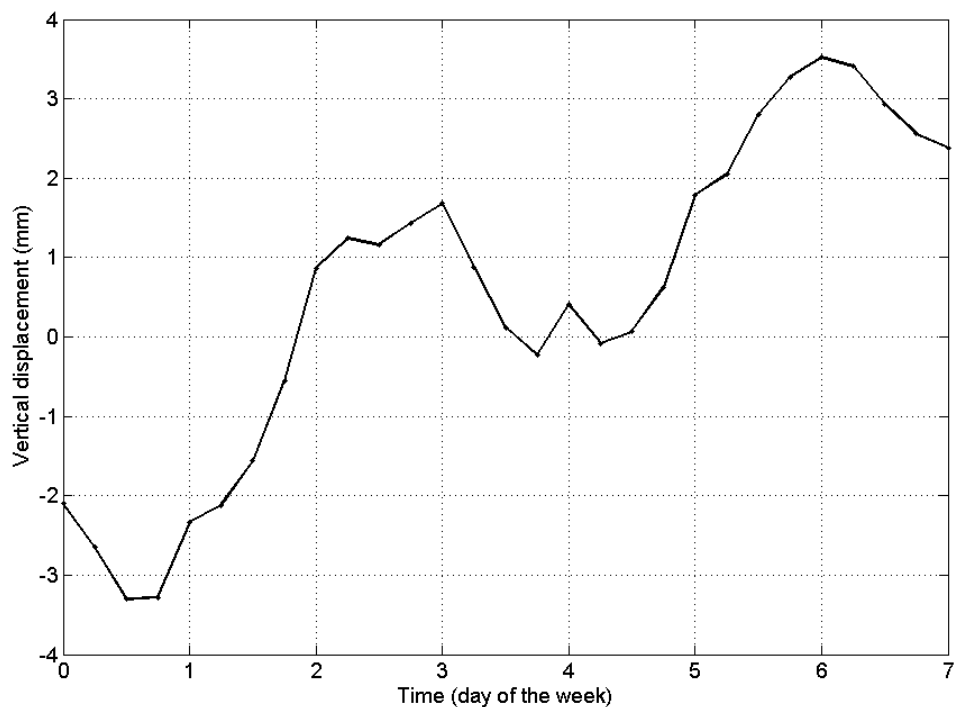


Figure 5.34: Vertical displacement due to non-tidal atmospheric pressure loading, over the week July 22–28, 2007, at station ALGO.

single profile-based atmospheres do, and is worse than allowing for a possibly different horizontal gradient at every position, as the 3d atmosphere does. It is important to stress that the discrepancy between the gradient and 3d atmospheres is dependent upon the resolution of the 3d model, because the lower that resolution is, the smoother the 3d model is, and the easier it is for a gradient atmosphere to act as a decent surrogate for the 3d model. More work is needed to assess at a which resolution full 3d atmosphere starts to be significantly better than simpler gradient atmospheres.

To sum up the conclusions from this experiment: spherical concentric is a poor model (or, in other words, the Earth’s oblateness cannot be neglected); a spherical osculating atmosphere accounts sufficiently well for the Earth’s oblateness; high-resolution NWM are promising, even if only extracting a single profile from it; non-tidal atmospheric pressure

loading should be accounted for; more research is needed to assess the benefits of full 3d atmospheres over simpler profile-based atmospheres.

Chapter 6

Conclusions

To close this work we summarize it, answering the following questions:

- What are the findings of the present work?
- What are the shortcomings of the present work?
- What remains to be learned?
- What are the fruits of the present work?
- How does this research fit in a wider context?

What are the findings of the present work? Following our working proposition (section 1.1, p. 2), we learned that there is more than one reasonable way of performing ray-tracing, sometimes with significantly different results.

On the theoretical side, we introduced the distinction among atmospheric source, atmospheric structure, and ray-path model. We reviewed the existing atmospheric sources, namely, climatologies, radiosondes, and numerical weather models. We systematized the full range of atmospheric structures and ray-path models, from simplest to most rigorous. More specifically, we examined the atmospheric structures called spherical concentric, spherical osculating, ellipsoidal, gradient, and 3d; and the ray-path models bent-3d, bent-2d, straight-line, and zenith.

On the experimental side, we learned that the bent-2d ray-path model, albeit not strictly valid in a 3d atmosphere, introduces only negligible errors, compared to the more rigorous bent-3d ray-path model (in a 15-km horizontal resolution atmospheric model). After completing the comparison of the two main ray-path models above, we turned to the atmospheric structures: we learned that the oblateness of the Earth cannot be neglected when it comes to predict the neutral atmosphere delay, as demonstrated by the poor results of a spherical concentric atmosphere; we learned that the spherical osculating model is the only one exhibiting azimuthal symmetry; we learned that the oblateness of the Earth is adequately accounted for by a spherical osculating model, as demonstrated by the small discrepancy between a spherical osculating and a more rigorous ellipsoidal model; we learned that a gradient atmosphere helps in accounting for the main trend in azimuthal asymmetry exhibited by a 3d atmosphere, but there remains secondary directions of azimuthal asymmetry that only a full 3d atmosphere is able to capture; furthermore, we found experimental evidence confirming the theoretical expectation that gradient and especially 3d atmospheric structures offer promising benefits for GPS positioning. Finally, beyond the comparison of atmospheric structures above, an interesting side conclusion regarding atmospheric sources was that atmospheric models of higher resolution might offer significant improvements in mapping functions.

Concerning the GPS technique itself, we offered a classification of the numerous ways in which the neutral atmosphere delay can be mitigated, which helped us in designing the GPS experiments in this work; we also recalled the importance of non-tidal atmospheric pressure loading when analyzing vertical position GPS solutions at the few-mm level.

What are the shortcomings of the present work? The main shortcoming in the present work is the somewhat small session duration and small number of stations involved in the assessment. Surely one can never prove a hypothesis right, but an analysis involving

longer sessions and a greater number of stations would corroborate — or refute — the conclusions offered here.

What remains to be learned? A major missing theme is the impact of atmospheric models of increasingly higher resolutions — both spatial and temporal. On one hand, we expect our conclusions about the validity of the bent-2d ray-path model in a 3d atmosphere and also about the discrepancy between a gradient and 3d atmospheres to be strongly dependent upon that resolution. On the other hand, we find higher resolution atmospheric models promising, even if used following simpler atmospheric structures based on a single vertical profile extracted from the full 3d NWM.

The missing theme above is in addition to the open problems identified in the background chapter but immediately put outside the scope of this work, namely, the representation of slant factors in a convenient form and the statistical exploitation of correlations in the delay, across nearby stations and nearby satellite directions.

What are the fruits of the present work? This being an exploratory work, it was not our intention to leave the reader with a practical product at hand. We will be satisfied if the reader leaves this work with a greater awareness regarding the different ways of performing ray-tracing, a more critical view about the shortcomings and potentials of the technique, and a sense of scrutiny towards uses of ray-tracing in the literature. The take-home message is: “There is more than one way of doing ray-tracing!”

How does this research fit in a wider context? Ever since the beginning of space geodesy, in the aftermath of the space race fifty years ago, efforts have been put into mitigating the effects of the neutral atmosphere in space geodetic measurements. Great progress has been made over that period, to the point that atmospheric scientists nowadays can rely on GPS for estimates of integrated water vapor on a routine basis. Which is not

to say that the neutral atmosphere delay is a solved case in space geodesy. Into the future, as the computation and communication machinery continues to improve, more abundant space geodetic observations and higher resolution atmospheric models will become available. That availability, in its turn, will surely pose new challenges, because every time we are able to improve upon the modeling of the current major error sources, a new level of smaller yet more difficult error sources is unveiled. Whereas in astrometry it might be well accepted that “To obtain higher accuracies, one must get rid of refraction completely by observing from space [as opposed to observing from the Earth’s surface].” [Kovalevsky and Seidelmann, 2004, p. 129], that option does not apply to geodesy and geophysics, whose phenomena of interest lie under the atmosphere. Ultimately, whether the neutral atmosphere delay will be a curse or a blessing for space geodetic techniques, that will be measured by our own creativity in discovering opportunities for what otherwise would be a nuisance.

References

- AGU (1995). “Water vapor in the climate system.” Special Report, American Geophysical Union (AGU), Public Information Committee, Washington, D.C., December, ISBN 0-87590-865-9, http://www.agu.org/sci_soc/mockler.html.
- Alexander, J. (2004). “Loxodromes: A rhumb way to go.” *Mathematics Magazine*, Vol. 77, No. 5, pp. 349–356, <http://www.cwru.edu/artsci/math/alexander/mathmag349-356.pdf>.
- Auer, L. H., and E. M. Standish (2000). “Astronomical refraction: Computational method for all zenith angles.” *The Astronomical Journal*, Vol. 119, No. 5, pp. 2472–2474, doi:10.1086/301325.
- Bevis, M., S. Businger, S. Chiswell, T. A. Herring, R. A. Anthes, C. Rocken, and R. H. Ware (1994). “GPS Meteorology: Mapping zenith wet delays onto precipitable water.” *Journal of Applied Meteorology*, Vol. 33, No. 3, pp. 379–386, doi:10.1175/1520-0450(1994)033<0379:GMMZWD>2.0.CO;2.
- Bevis, M., S. Chiswell, S. Businger, T. A. Herring, and Y. Bock (1996). “Estimating wet delays using numerical weather analyses and predictions.” *Radio Science*, Vol. 31, No. 3, pp. 477–487, doi:10.1029/96RS00008.
- Blewitt, G. (2007). “GPS and space based geodetic methods.” In *Geodesy*, ed. T. Herring, Vol. 3 of *Treatise on Geophysics*, Academic Press, Oxford, ISBN 978-0-444-52748-6, pp. 351–390, doi:10.1016/B978-044452748-6.00058-4, G. Schubert (Editor-in-Chief), Eleven-volume set.
- Bock, Y. (1998). “Medium distance GPS measurements.” In Kleusberg and Teunissen [1998], Chapter 9, pp. 483–586.
- Boehm, J., A. E. Niell, P. Tregoning, and H. Schuh (2006a). “Global Mapping Function (GMF): A new empirical mapping function based on numerical weather model data.” *Geophysical Research Letters*, Vol. 33, No. L07304, doi:10.1029/2005GL025546.
- Boehm, J., and H. Schuh (2003). “Vienna Mapping Functions.” *Proceedings of the 16th Working Meeting on European VLBI for Geodesy and Astrometry*, eds. W. Schwegmann, and V. Thorandt, Bundesamt für Kartographie und Geodäsie, Verlag des Bundesamtes für Kartographie und Geodäsie, Frankfurt/Leipzig, Germany, May 9–10,

- pp. 131–144. http://www.hg.tuwien.ac.at/Archiv/Papers/2003_16thEuVLBI_boehm_pa_a.pdf.
- Boehm, J., and H. Schuh (2007). “Troposphere gradients from the ECMWF in VLBI analysis.” *Journal of Geodesy*, Vol. 81, No. 6, pp. 403–408, doi:10.1007/s00190-007-0144-2.
- Boehm, J., B. Werl, and H. Schuh (2006b). “Troposphere mapping functions for GPS and very long baseline interferometry from European Centre for Medium-Range Weather Forecasts operational analysis data.” *Journal of Geophysical Research*, Vol. 111, No. B02406, doi:10.1029/2005JB003629.
- Böhm, J., and H. Schuh (2001). “Spherical harmonics as a supplement to global tropospheric mapping functions and horizontal gradients.” *Proceedings of the 15th Working Meeting on European VLBI for Geodesy and Astrometry*, eds. D. Behrend, and A. Rius, Institut d’Estudis Espacials de Catalunya, Barcelona, Spain, September 7–8, pp. 143–148. <ftp://io.ieec.fcr.es/ieec/gps/15wmevga/pdf/boehm.pdf>.
- Bolton, D. (1980). “The computation of equivalent potential temperature.” *Monthly Weather Review*, Vol. 108, No. 7, pp. 1046–1053, doi:10.1175/1520-0493(1980)108<1046:TCOEPT>2.0.CO;2.
- Born, M., and E. Wolf (1999). *Principles of Optics: Electromagnetic Theory of Propagation, Interference and Diffraction of Light*. 7th ed., 986 pp., Cambridge University Press, Cambridge, U.K., ISBN 0521642221, doi:10.2277/0521642221.
- Bosser, P., O. Bock, J. Pelon, and C. Thom (2007). “An improved mean-gravity model for GPS hydrostatic delay calibration.” *IEEE Geoscience and Remote Sensing Letters*, Vol. 4, No. 1, pp. 3–7, doi:10.1109/LGRS.2006.881725.
- Brettle, M. J., and J. F. P. Galvin (2003). “Back to basics: Radiosondes: Part 1 — The instrument.” *Weather*, Vol. 58, No. 9, pp. 336–341, doi:10.1256/wea.126.02A.
- Brunner, F. K. (Ed.) (1984). *Geodetic refraction: effects of electromagnetic wave propagation through the atmosphere*, Springer-Verlag, Berlin, 213 pp., ISBN 0387138307, with contributions by P.V. Angus-Leppan et al.
- Buizza, R. (2002). “Weather prediction: Ensemble prediction.” In *Encyclopedia of Atmospheric Sciences*, eds. J. R. Holton, J. A. Curry, and J. A. Pyle, Academic Press, Oxford, ISBN 9780122270901, pp. 2546–2557, doi:10.1016/B0-12-227090-8/00461-9.
- Červený, V. (2001). *Seismic Ray Theory*. 722 pp., Cambridge University Press, Cambridge, U.K., ISBN 0521366712.

- Chandra, S., E. L. Fleming, M. R. Schoeberl, and J. J. Barnett (1990). “Monthly mean global climatology of temperature, wind, geopotential height and pressure for 0–120 km.” *Advances in Space Research*, Vol. 10, No. 6, pp. 3–12, doi:10.1016/0273-1177(90)90230-W.
- Chen, G., and T. A. Herring (1997). “Effects of atmospheric azimuth asymmetry on the analysis of space geodetic data.” *Journal of Geophysical Research*, Vol. 102, No. B9, pp. 20489–20502, doi:10.1029/97JB01739.
- COESA (1962). “U.S. Standard Atmosphere, 1962 — ICAO standard atmosphere to 20 kilometers, proposed ICAO extension to 32 kilometers, tables and data to 700 kilometers.” U.S. Government Printing Office, United States Committee on Extension to the Standard Atmosphere (COESA), National Aeronautics and Space Administration, United States Air Force, United States Weather Bureau, Washington, D.C., December, 278 pp., <http://handle.dtic.mil/100.2/AD659893>.
- COESA (1976). “U.S. Standard Atmosphere, 1976.” U.S. Government Printing Office NOAA–S/T 76–1562, United States Committee on Extension to the Standard Atmosphere (COESA), National Oceanic and Atmospheric [sic] Administration, National Aeronautics and Space Administration, United States Air Force, Washington, D.C., October, 227 pp., <http://hdl.handle.net/2060/19770009539>.
- Collins, J. P. (1999). *Assessment and Development of a Tropospheric Delay Model for Aircraft Users of the Global Positioning System*. M.Sc.E. thesis, University of New Brunswick, Dept. of Geodesy and Geomatics Engineering, Fredericton, N.B., Canada, September, 174 pp., Technical Report 203, <http://gge.unb.ca/Pubs/TR203.pdf>.
- Colombo, O. L. (2006). “A zenith delay model for precise kinematic aircraft navigation.” *Proceedings of the 19th International Technical Meeting of the Satellite Division of the Institute of Navigation — ION GNSS 2006*, The Institute of Navigation, Fort Worth, Tex., September 26–29, pp. 68–76. <http://bowie.gsfc.nasa.gov/697/staff/colombo/pdfs/colombo.tropo.ion.gnss-2006.pdf>.
- COMET-UCAR (1999). “NWP model fundamentals.”, University Corporation for Atmospheric Research (UCAR), Cooperative Program for Operational Meteorology, Education and Training (COMET), (Lecture Notes), http://www.meted.ucar.edu/topics_nwp.php.
- Côté, J., S. Gravel, A. Méthot, A. Patoine, M. Roch, and A. Staniforth (1998). “The operational CMC–MRB Global Environmental Multiscale (GEM) model. Part I: Design considerations and formulation.” *Monthly Weather Review*, Vol. 126, No. 6, pp. 1373–1395, doi:10.1175/1520-0493(1998)126<1373:TOCMGE>2.0.CO;2.
- Cove, K. (2005). *Improvements in GPS Tropospheric Delay Estimation with Numerical Weather Prediction*. M.Sc.E. thesis, University of New Brunswick, Dept. of Geodesy

- and Geomatics Engineering, Fredericton, N.B., Canada, May, 98 pp., Technical Report 230, <http://gge.unb.ca/Pubs/TR230.pdf>.
- Curry, J. (2002a). "Humidity variables." In *Encyclopedia of Atmospheric Sciences*, eds. J. R. Holton, J. A. Curry, and J. A. Pyle, Academic Press, Oxford, ISBN 9780122270901, pp. 939–941, doi:10.1016/B0-12-227090-8/00162-7.
- Curry, J. A. (2002b). "Thermodynamics: Moist (unsaturated) air." In *Encyclopedia of Atmospheric Sciences*, eds. J. R. Holton, J. A. Curry, and J. A. Pyle, Academic Press, Oxford, ISBN 9780122270901, pp. 2274–2278, doi:10.1016/B0-12-227090-8/00405-X.
- Dabberdt, W. F., R. Shellhorn, H. Cole, A. Paukkunen, J. Horhammer, and V. Antikainen (2002). "Radiosondes." In *Encyclopedia of Atmospheric Sciences*, eds. J. R. Holton, J. A. Curry, and J. A. Pyle, Academic Press, Oxford, ISBN 9780122270901, pp. 1900–1913, doi:10.1016/B0-12-227090-8/00344-4.
- Dach, R., and R. Dietrich (2000). "Influence of the ocean loading effect on GPS derived precipitable water vapor." *Geophysical Research Letters*, Vol. 27, No. 18, doi:10.1029/1999GL010970.
- Dai, L., S. Han, J. Wang, and C. Rizos (2003). "Comparison of interpolation algorithms in network-based GPS techniques." *Navigation*, Vol. 50, No. 4, pp. 277–293.
- Davis, J. L. (1986). *Atmospheric propagation effects on radio interferometry*. Ph.D. thesis, Massachusetts Institute of Technology, Dept. of Earth, Atmospheric, and Planetary Sciences, Cambridge, Mass., 284 pp., Also published as AFGL Technical Report 86-0243, U.S. Air Force Geophysics Laboratory, Hanscom AFB, Mass., <http://hdl.handle.net/1721.1/27953>.
- Davis, J. L., G. Elgered, A. E. Niell, and C. E. Kuehn (1993). "Ground-based measurement of gradients in the "wet" radio refractivity of air." *Radio Science*, Vol. 28, No. 6, pp. 1003–1018, doi:10.1029/93RS01917.
- Davis, J. L., T. A. Herring, I. I. Shapiro, A. E. E. Rogers, and G. Elgered (1985). "Geodesy by radio interferometry: Effects of atmospheric modeling errors on estimates of baseline length." *Radio Science*, Vol. 20, No. 6, pp. 1593–1607, doi:10.1029/RS020i006p01593.
- Dodson, A. H., W. Chen, N. T. Penna, and H. C. Baker (2001). "GPS estimation of atmospheric water vapour from a moving platform." *Journal of Atmospheric and Solar-Terrestrial Physics*, Vol. 63, No. 12, pp. 1331–1341, doi:10.1016/S1364-6826(00)00251-0.
- Dow, J. M., R. E. Neilan, and G. Gendt (2005). "The international GPS service: Celebrating the 10th anniversary and looking to the next decade." *Advances in Space Research*, Vol. 36, No. 3, pp. 320–326, doi:10.1016/j.asr.2005.05.125.

- Eresmaa, R., and H. Jarniven (2005). “Estimation of spatial global positioning system zenith delay observation error covariance.” *Tellus A*, Vol. 57, No. 2, pp. 194–203, doi:10.1111/j.1600-0870.2005.00096.x.
- Espelid, T. O. (2007). “Algorithm 868: Globally doubly adaptive quadrature — reliable matlab codes.” *ACM Transactions on Mathematical Software (TOMS)*, Vol. 33, No. 3, Article No. 21, doi:10.1145/1268769.1268775.
- Essen, L., and K. D. Froome (1951). “The refractive indices and dielectric constants of air and its principal constituents at 24,000 Mc/s.” *Proceedings of the Physical Society, Section B*, Vol. 64, pp. 862–875, doi:10.1088/0370-1301/64/10/303.
- Fleming, E. L., S. Chandra, M. R. Schoeberl, and J. J. Barnet (1988). “Monthly mean global climatology of temperature, wind, geopotential height and pressure for 0–120 km.” Technical Memorandum 100697, NASA Goddard Space Flight Center, Greenbelt, Md., February, 91 pp., <http://hdl.handle.net/2060/19880013119>.
- Gardner, C. S. (1976). “Effects of horizontal refractivity gradients on the accuracy of laser ranging to satellites.” *Radio Science*, Vol. 11, No. 12, pp. 1037–1044, doi:10.1029/RS011i012p01037.
- Ge, S. (2006). *GPS radio occultation and the role of atmospheric pressure on spaceborne gravity estimation over Antarctica*. Ph.D. thesis, Ohio State University, Geodetic Science and Surveying, Columbus, Ohio, July, 215 pp., Also published as OSU Report No. 479, http://rave.ohiolink.edu/etdc/view?acc_num=osu1149070384.
- Ghoddousi-Fard, R., and P. Dare (2007). “A climatic based asymmetric mapping function using a dual radiosonde raytracing approach.” *Proceedings of the 20th International Technical Meeting of the Satellite Division of the Institute of Navigation — ION GNSS 2007*, The Institute of Navigation, Fort Worth, Tex., September 25–28, pp. 2870–2879. <http://gauss.gge.unb.ca/papers.pdf/iongnss2007.ghoddousi-fard.pdf>.
- Glickman, T. S. (Ed.) (2000). *Glossary of Meteorology*. 2nd ed., 855 pp., American Meteorological Society, Boston, ISBN 1-878220-34-9, <http://amsglossary.allenpress.com/>.
- Golding, B. W. (2002). “Weather prediction: Regional prediction models.” In *Encyclopedia of Atmospheric Sciences*, eds. J. R. Holton, J. A. Curry, and J. A. Pyle, Academic Press, Oxford, ISBN 9780122270901, pp. 2557–2562, doi:10.1016/B0-12-227090-8/00253-0.
- Gutierrez, D., F. J. Seron, A. Munoz, and O. Anson (2006). “Simulation of atmospheric phenomena.” *Computers & Graphics*, Vol. 30, No. 6, pp. 994–1010, doi:10.1016/j.cag.2006.05.002.

- Gutman, S. I., T. Fuller-Rowell, and D. Robinson (2003a). “Using NOAA atmospheric models to improve ionospheric and tropospheric corrections.” U.S. Coast Guard Differential GPS Symposium, Portsmouth, Va., June 19. (Oral presentation), http://www.gpsmet.noaa.gov/jsp/downloads/NOAA_Atmospheric_Corrections.ppt.
- Gutman, S. I., S. Sahm, J. Stewart, S. Benjamin, T. L. Smith, and B. Schwartz (2003b). “A new composite observing system strategy for ground based GPS meteorology.” *Proceedings of 12th Symposium on Meteorological Observations and Instrumentation*, American Meteorological Society, Long Beach, Calif., February 8–13. <http://ams.confex.com/ams/pdfpapers/54168.pdf>.
- Guyot, G. (1998). *Physics of the Environment and Climate*. Wiley-Praxis series in atmospheric physics and climatology, 3rd ed., 632 pp., John Wiley & Sons/Praxis Publishing, Chichester, N.Y., ISBN 0471968285.
- Hartmann, D. L. (2002). “Climate: Overview.” In *Encyclopedia of Atmospheric Sciences*, eds. J. R. Holton, J. A. Curry, and J. A. Pyle, Academic Press, Oxford, ISBN 9780122270901, pp. 403–411, doi:10.1016/B0-12-227090-8/00024-5.
- Hase, F., and M. Höpfner (1999). “Atmospheric ray path modeling for radiative transfer algorithms.” *Appl. Opt.*, Vol. 38, No. 15, pp. 3129–3133, doi:10.1364/A0.38.003129.
- Hayya, J., D. Armstrong, and N. Gressis (1975). “A note on the ratio of two normally distributed variables.” *Management Science*, Vol. 21, No. 11, pp. 1338–1341, <http://www.jstor.org/stable/2629897>.
- Herring, T. A. (1992). “Modelling atmospheric delays in the analysis of space geodetic data.” *Proceedings of the Symposium Refraction of Transatmospheric Signals in Geodesy*, eds. J. C. de Munck, and T. A. TH. Spoelstra, No. 36, Netherlands Geodetic Commission, The Hague, The Netherlands, May 19–22, ISBN 9789061322436, pp. 157–164. <http://www.ncg.knaw.nl/eng/publications/Geodesy/36DeMunck.html>.
- Hobiger, T., R. Ichikawa, Y. Koyama, and T. Kondo (2008a). “Fast and accurate ray-tracing algorithms for real-time space geodetic applications using numerical weather models.” *Journal of Geophysical Research*, Vol. 113, No. D20, doi:10.1029/2008JD010503.
- Hobiger, T., R. Ichikawa, T. Takasu, Y. Koyama, and T. Kondo (2008b). “Ray-traced troposphere slant delays for precise point positioning.” *Earth, Planets and Space*, Vol. 60, No. 5, pp. e1–e4, <http://www.terrapub.co.jp/journals/EPS/pdf/2008e/6005e001.pdf>.
- Hopfield, H. S. (1969). “Two-quartic tropospheric refractivity profile for correcting satellite data.” *Journal of Geophysical Research*, Vol. 74, No. 18, pp. 4487–4499.

- Hulley, G. C., and E. C. Pavlis (2007). "A ray-tracing technique for improving satellite laser ranging atmospheric delay corrections, including the effects of horizontal refractivity gradients." *Journal of Geophysical Research*, Vol. 112, No. B06417, doi:10.1029/2006JB004834.
- Ifadis, I. M. (1986). *The atmospheric delay of radio waves: modeling the elevation dependence on a global scale*. Licentiate thesis, Chalmers University of Technology, School of Electrical and Computer Engineering, Gothenburg, Sweden, 115 pp., Technical Report 38L.
- Ifadis, I. M. (2000). "A new approach to mapping the atmospheric effect for GPS observations." *Earth, Planets and Space*, Vol. 52, No. 10, pp. 703–708, <http://www.terrapub.co.jp/journals/EPS/pdf/5210/52100703.pdf>.
- Ifadis, I. M., and P. Savvaidis (2001). "Space to earth geodetic observations: approaching the atmospheric effect." *Physics and Chemistry of the Earth, Part A*, Vol. 26, No. 3, pp. 195–200, doi:10.1016/S1464-1895(01)00046-1.
- Ingensand, H. (2002). "Concepts and solutions to overcome the refraction problem in terrestrial precision measurement." *FIG XXII International Congress*, International Federation of Surveyors (FIG), Washington, D.C., April 19–26. http://www.fig.net/pub/fig_2002/Js28/JS28_ingensand.pdf.
- IPCC (2007). *Climate Change 2007 — The Physical Science Basis*. 1009 pp., Cambridge University Press, Cambridge, U.K., ISBN 9780521705967, Working Group I Contribution to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change (IPCC), <http://www.ipcc.ch/ipccreports/ar4-wg1.htm>.
- Iwabuchi, T., C. Rocken, J. Johnson, T. Van Hove, and I. Naito (2003). "Estimation of GPS site height by using dynamic hydrostatic and wet mapping functions based on radiosonde refractivity profile." Fall Meeting 2008, American Geophysical Union, San Francisco, Calif., December 8–12, Eos Trans. AGU Fall Meet. Suppl., Abstract G52B-0044. (Poster presentation).
- Jensen, A. B. O. (2005). "Numerical weather predictions for GPS positioning." In *A Window on the Future of Geodesy — Proceedings of the IAG 2003 General Assembly, Sapporo, Japan, June 30 – July 11*, ed. F. Sansò, Vol. 128 of *International Association of Geodesy Symposia*, Springer, Berlin, ISBN 978-3-540-24055-6, pp. 65–70, doi:10.1007/3-540-27432-4_12.
- Johnson, D. L., B. C. Roberts, and W. W. Vaughan (2002). "Reference and standard atmosphere models." *10th Conference on Aviation, Range, and Aerospace Meteorology*, Portland, Oreg., May 13–16. <http://hdl.handle.net/2060/20020092087>.

- Kalnay, E., M. Kanamitsu, R. Kistler, W. Collins, D. Deaven, L. Gandin, M. Iredell, S. Saha, G. White, J. Woollen, Y. Zhu, A. Leetmaa, R. Reynolds, M. Chelliah, W. Ebisuzaki, W. Higgins, J. Janowiak, K. Mo, C. Ropelewski, J. Wang, R. Jenne, and D. Joseph (1996). “The NCEP/NCAR 40-year reanalysis project.” *Bulletin of the American Meteorological Society*, Vol. 77, No. 3, pp. 437–471, doi:10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2.
- Kirchengast, G., J. Hafner, and W. Poetzi (1999). “CIRA86aQ_UoG model: An extension of the CIRA–86 monthly tables including humidity tables and a Fortran95 global moist air climatology model.” Technical Report for ESA/ESTEC No. 8/1999, Institute for Meteorology and Geophysics, University of Graz, Graz, Austria, November 18, 18 pp., <http://www.uni-graz.at/igam-arsclisys/publications/publ1999/GK-IMGTechRepfESA-18p-n8y1999.pdf>.
- Kleusberg, A., and P. J. G. Teunissen (Eds.) (1998). *GPS for Geodesy*, Vol. 60 of *Lecture Notes in Earth Sciences*. 2nd ed., 650 pp., Springer, Berlin, ISBN 3540636617, compl. rev. and extend. ed.
- Kouba, J. (2008). “Implementation and testing of the gridded Vienna Mapping Function 1 (VMF1).” *Journal of Geodesy*, Vol. 82, No. 4, pp. 193–205, doi:10.1007/s00190-007-0170-0.
- Kovalevsky, J., and P. K. Seidelmann (2004). *Fundamentals of Astrometry*. 404 pp., Cambridge University Press, Cambridge, U.K., ISBN 0521642167.
- Kravtsov, Y. A. (2005). *Geometrical Optics in Engineering Physics*. 370 pp., Alpha Science, Harrow, UK, ISBN 1842651595.
- Krueger, E., T. Schüler, G. W. Hein, A. Martellucci, and G. Blarzino (2004). “Galileo tropospheric correction approaches developed within GSTB-V1.” *Proceedings of GNSS 2004 – European Navigation Conference*, Rotterdam, The Netherlands, May 17–19. <http://forschung.unibw.de/berichte/2006/11i3llunkzpngxhfxagnqvlvoxxkkw.pdf>.
- Kuang, D., B. E. Schutz, and M. M. Watkins (1996). “On the structure of geometric positioning information in GPS measurements.” *Journal of Geodesy*, Vol. 71, No. 1, pp. 35–43, doi:10.1007/s001900050073.
- Kursinski, E. R., G. A. Hajj, S. S. Leroy, and B. Herman (2000). “The GPS radio occultation technique.” *Terrestrial, Atmospheric and Oceanic Sciences*, Vol. 11, No. 1, pp. 53–114. <http://hdl.handle.net/2014/14027>
- Kursinski, E. R., G. A. Hajj, J. T. Schofield, R. P. Linfield, and K. R. Hardy (1997). “Observing Earth’s atmosphere with radio occultation measurements using the Global Positioning System.” *Journal of Geophysical Research*, Vol. 102, No. D19, pp. 23429–23465, doi:10.1029/97JD01569.

- Lambert, W. D. (1949). "Some notes on the calculation of the geopotential.", unpublished manuscript.
- Leandro, R., M. Santos, and R. B. Langley (2006a). "UNB neutral atmosphere models: Development and performance." *Proceedings of the 2006 National Technical Meeting of the Institute of Navigation — ION NTM 2006*, The Institute of Navigation, Monterey, Calif., January 18–20, pp. 564–573. <http://gauss.gge.unb.ca/papers.pdf/ionntm2006.leandro.pdf>.
- Leandro, R. F. (2008). Personal communication. Department of Geodesy and Geomatics Engineering, University of New Brunswick, Fredericton, N.B., Canada.
- Leandro, R. F., R. B. Langley, and M. C. Santos (2008). "UNB3m-pack: a neutral atmosphere delay package for radiometric space techniques." *GPS Solutions*, Vol. 12, No. 1, pp. 65–70, doi:10.1007/s10291-007-0077-5.
- Leandro, R. F., M. C. Santos, and R. B. Langley (2006b). "Wide area neutral atmosphere models for GNSS applications." *Proceedings of the 19th International Technical Meeting of the Satellite Division of the Institute of Navigation — ION GNSS 2006*, The Institute of Navigation, Fort Worth, Tex., September 26–29, pp. 1910–1924. <http://gauss.gge.unb.ca/papers.pdf/iongnss2006.leandro.tropo.pdf>.
- Leandro, R. F., M. C. Santos, and R. B. Langley (2007). "GAPS: The GPS analysis and positioning software – a brief overview." *Proceedings of the 20th International Technical Meeting of the Satellite Division of the Institute of Navigation — ION GNSS 2007*, The Institute of Navigation, Fort Worth, Tex., September 25–28, pp. 1807–1811. <http://gauss.gge.unb.ca/papers.pdf/iongnss2007.leandro.gaps.pdf>.
- Lehn, W. H., and S. van der Werf (2005). "Atmospheric refraction: a history." *Applied Optics*, Vol. 44, No. 27, pp. 5624–5636, doi:10.1364/AO.44.005624.
- Leick, A. (2003). *GPS Satellite Surveying*. 3rd ed., 464 pp., Wiley, Hoboken, N.J., ISBN 0471059307.
- Lemoine, F. G., S. C. Kenyon, J. K. Factor, R. Trimmer, N. K. Pavlis, D. S. Chinn, C. M. Cox, S. M. Klosko, S. B. Luthcke, M. H. Torrence, Y. M. Wang, R. G. Williamson, E. C. Pavlis, R. H. Rapp, and T. R. Olson (1998). "The development of the joint NASA GSFC and the National Imagery and Mapping Agency NIMA geopotential model EGM96." TP 1998-206861, NASA Goddard Space Flight Center, Greenbelt, Md., July, 575 pp., <http://hdl.handle.net/2060/19980218814>.
- List, R. J. (1951). *Smithsonian Meteorological Tables*, Vol. 114 of *Smithsonian Miscellaneous Collections*. 6th ed., 527 pp., The Smithsonian Institution, Washington, D.C., Publication 4014, third reprint, 1966.

- Lorenc, A. C. (2002). "Weather prediction: Data assimilation." In *Encyclopedia of Atmospheric Sciences*, eds. J. R. Holton, J. A. Curry, and J. A. Pyle, Academic Press, Oxford, ISBN 9780122270901, pp. 2542–2546, doi:10.1016/B0-12-227090-8/00458-9.
- MacMillan, D. S., and C. Ma (1998). "Using meteorological data assimilation models in computing tropospheric delays at microwave [*sic*] frequencies." *Physics and Chemistry of the Earth*, Vol. 23, No. 1, pp. 97–102, doi:10.1016/S0079-1946(97)00249-8.
- Marchand, E. W. (1978). *Gradient Index Optics*. 166 pp., Academic Press, New York, ISBN 0124707505.
- Marini, J. W. (1972). "Correction of satellite tracking data for an arbitrary tropospheric profile." *Radio Science*, Vol. 7, No. 2, pp. 223–231, doi:10.1029/RS007i002p00223.
- McCarthy, D. D., and G. Petit (Eds.) (2004). *IERS Conventions (2003)*. 127 pp., Frankfurt am Main: Verlag des Bundesamts für Kartographie und Geodäsie, Frankfurt, ISBN 3-89888-884-3, IERS Technical Note 32, <http://www.iers.org/MainDisp.csl?pid=46-25776>.
- Mendes, V. B. (1999). *Modeling the Neutral-Atmosphere Propagation Delay in Radiometric Space Techniques*. Ph.D. thesis, University of New Brunswick, Dept. of Geodesy and Geomatics Engineering, Fredericton, N.B., Canada, April, 349 pp., Technical Report 199, <http://gge.unb.ca/Pubs/TR199.pdf>.
- Misra, P. N. (1996). "The role of the clock in a GPS receiver." *GPS World*, Vol. 7, No. 4, pp. 60–66.
- Murphy, D. M., and T. Koop (2005). "Review of the vapour pressures of ice and super-cooled water for atmospheric applications." *Quarterly Journal of the Royal Meteorological Society*, Vol. 131, No. 608, pp. 1539–1565, doi:10.1256/qj.04.94.
- Nener, B. D., N. Fowkes, and L. Borredon (2003). "Analytical models of optical refraction in the troposphere." *J. Opt. Soc. Am. A*, Vol. 20, No. 5, pp. 867–875, doi:10.1364/JOSAA.20.000867.
- Niell, A. E. (1996). "Global mapping functions for the atmosphere delay at radio wavelengths." *Journal of Geophysical Research*, Vol. 101, No. B2, pp. 3227–3246, doi:10.1029/95JB03048.
- Niell, A. E. (2003). "The IMF mapping functions." *International Workshop on GPS Meteorology*, Tsukuba, Japan, January 14–17. http://www.haystack.mit.edu/geo/pubs/the_imf_mapping_functions_rev2.pdf.
- Nievinski, F., K. Cove, M. Santos, D. Wells, and R. Kingdon (2005). "Range-extended GPS kinematic positioning using numerical weather prediction model." *Proceedings of the ION 61st Annual Meeting — ION AM 2005*, The Institute of Navigation,

- Cambridge, Mass., June 27–29, pp. 902–911. <http://gge.unb.ca/Research/GRL/PrincessOfAcadia/Publications/E2-2-Nievinski.pdf>.
- OFCM (2007). “Rawinsonde and pibal observations.” In *Federal Meteorological Handbook No. 3*, Office of the Federal Coordinator for Meteorology, Silver Spring, Md., FCM-H3-1997, <http://www.ofcm.gov/fmh3/fmh3.htm>.
- Pany, T., P. Peseć, and G. Stangl (2001). “Elimination of tropospheric path delays in GPS observations with the ECMWF numerical weather model.” *Physics and Chemistry of the Earth, Part A: Solid Earth and Geodesy*, Vol. 26, No. 6-8, pp. 487–492, doi:10.1016/S1464-1895(01)00089-8.
- Petrov, L., and J. P. Boy (2004). “Study of the atmospheric pressure loading signal in very long baseline interferometry observations.” *Journal of Geophysical Research*, Vol. 109, No. B03405, doi:10.1029/2003JB002500.
- Pharr, M., and G. Humphreys (2004). *Physically Based Rendering: From Theory to Implementation*. 1019 pp., Elsevier/Morgan Kaufman, San Francisco, ISBN 012553180X.
- Rademacher, P. (1997). “Ray tracing: graphics for the masses.” *ACM Crossroads*, Vol. 3, No. 4, pp. 3–7, doi:10.1145/270955.270962.
- Rawlinson, N., J. Hauser, and M. Sambridge (2008). “Seismic ray tracing and wavefront tracking in laterally heterogeneous media.” In *Advances in Geophysics*, ed. R. Dmowska, Vol. 49, Academic Press/Elsevier, Amsterdam, ISBN 9780123742315, pp. 203–267, doi:10.1016/S0065-2687(07)49003-3.
- Rocken, C., S. Sokolovskiy, J. M. Johnson, and D. Hunt (2001). “Improved mapping of tropospheric delays.” *Journal of Atmospheric and Oceanic Technology*, Vol. 18, No. 7, pp. 1205–1213, doi:10.1175/1520-0426(2001)018<1205:IMOTD>2.0.CO;2.
- Rüeger, J. M. (2002). “Refractive index formulae for radio waves.” *FIG XXII International Congress*, International Federation of Surveyors (FIG), Washington, D.C., April 19–26. http://www.fig.net/pub/fig_2002/Js28/JS28_rueger.pdf.
- Saastamoinen, J. (1972). “Atmospheric correction for the troposphere and stratosphere in radio ranging of satellites.” In *The Use of Artificial Satellites for Geodesy*, eds. S. W. Henriksen, A. Mancini, and B. H. Chovitz, Vol. 15 of *Geophysical Monograph Series*, American Geophysical Union, Washington, D.C., ISBN 0-87590-015-1, pp. 247–251.
- Salomon, D. (2007). *Data Compression: The Complete Reference*. 4th ed., 1092 pp., Springer, London, ISBN 1846286026, (with contributions by G. Motta and D. Bryant).
- Salstein, D. A., D. M. Kann, A. J. Miller, and R. D. Rosen (1993). “The sub-bureau for atmospheric angular momentum of the International Earth Rotation Service: A meteorological data center with geodetic applications.” *Bulletin of the American Meteorological Society*, Vol. 74, No. 1, pp. 67–80, doi:10.1175/1520-0477(1993)074<0067:TSBFAA>2.0.CO;2.

- Schüler, T. (2006). “Impact of systematic errors on precise long-baseline kinematic GPS positioning.” *GPS Solutions*, Vol. 10, No. 2, pp. 108–125, doi:10.1007/s10291-005-0012-6.
- Seko, H., H. Nakamura, and S. Shimada (2004). “An evaluation of atmospheric models for GPS data retrieval by output from a numerical weather model.” *Journal of the Meteorological Society of Japan*, Vol. 82, No. 1B, pp. 339–350, doi:10.2151/jmsj.2004.339.
- Seron, F. J., D. Gutierrez, G. Gutierrez, and E. Cerezo (2005). “Implementation of a method of curved ray tracing for inhomogeneous atmospheres.” *Computers & Graphics*, Vol. 29, No. 1, pp. 95–108, doi:10.1016/j.cag.2004.11.010.
- Sovers, O. J., J. L. Fanselow, and C. S. Jacobs (1998). “Astrometry and geodesy with radio interferometry: experiments, models, results.” *Reviews of Modern Physics*, Vol. 70, No. 4, pp. 1393–1454, doi:10.1103/RevModPhys.70.1393.
- Spencer, G. H., and M. V. R. K. Murty (1962). “General ray-tracing procedure.” *J. Opt. Soc. Am.*, Vol. 52, No. 6, pp. 672–676, doi:10.1364/JOSA.52.000672.
- Spilker, J. J., Jr. (1996). “Tropospheric effects on GPS.” In *Global Positioning System: Theory and Applications*, eds. B. W. Parkinson, and J. J. Spilker, Jr., Vol. I of *Progress in Astronautics and Aeronautics*, Chapter 13, American Institute of Aeronautics & Astronautics (AIAA), Washington, D.C., ISBN 156347106X, pp. 517–546.
- Stam, J., and E. Languénoü (1996). “Ray tracing in non-constant media.” *Proceedings of the Eurographics Rendering techniques '96 workshop*, eds. X. Pueyo, and P. Schröder, Springer-Verlag, Porto, Portugal, June 17–19, ISBN 3-211-82883-4, pp. 225–234. <http://www.dgp.toronto.edu/people/stam/reality/Research/pdf/wobble.pdf>.
- Steigenberger, P., J. Boehm, and V. Tesmer (2008). “Comparison of GMF/GPT with VMF1/ECMWF and implications for atmospheric loading.” *Journal of Geodesy*, (submitted).
- Stoew, B., and G. Elgered (2005). “Spatial and temporal correlations of the GPS estimation errors.” Tech. Rep., Onsala Space Observatory, Chalmers University of Technology, Onsala, Sweden, March 16, Work carried out within the TOUGH Project (Targeting Optimal Use of GPS Humidity Measurements in Meteorology), <http://web.dmi.dk/pub/tough/deliverables/d19-report.pdf>.
- Syndergaard, S. (1998). “Modeling the impact of the Earth’s oblateness on the retrieval of temperature and pressure profiles from limb sounding.” *Journal of Atmospheric and Solar–Terrestrial Physics*, Vol. 60, No. 2, pp. 171–180, doi:10.1016/S1364-6826(97)00056-4.

- Syndergaard, S., Y.-H. Kuo, and M. S. Lohmann (2006). "Observation operators for the assimilation of occultation data into atmospheric models: A review." In *Atmosphere and Climate — Studies by Occultation Methods*, eds. U. Foelsche, G. Kirchengast, and A. Steiner, Part 5, Springer, Berlin, ISBN 978-3-540-34116-1, pp. 205–224, doi:10.1007/3-540-34121-8_18.
- Thayer, G. D. (1974). "An improved equation for the radio refractive index of air." *Radio Science*, Vol. 9, No. 10, pp. 803–807, doi:10.1029/RS009i010p00803.
- Thessin, R. N. (2005). *Atmospheric signal delay affecting GPS measurements made by space vehicles during launch, orbit and reentry*. Master's thesis, Massachusetts Institute of Technology, Dept. of Aeronautics and Astronautics, Cambridge, Mass., 182 pp., <http://hdl.handle.net/1721.1/33211>.
- Thorpe, A. J. (2005). "Climate change prediction: A challenging scientific problem." Institute of Physics, London, http://www.iop.org/activity/policy/Publications/file_4147.pdf.
- Torge, W. (2001). *Geodesy*. 3rd ed., 416 pp., de Gruyter, Berlin, ISBN 978-3-11-017072-6, compl. rev. and extend. ed.
- Tregoning, P., and T. van Dam (2005). "Atmospheric pressure loading corrections applied to GPS data at the observation level." *Geophysical Research Letters*, Vol. 32, No. L22310, doi:10.1029/2005GL024104.
- Tregoning, P., and T. A. Herring (2006). "Impact of a priori zenith hydrostatic delay errors on GPS estimates of station heights and zenith total delays." *Geophysical Research Letters*, Vol. 33, No. L23303, doi:10.1029/2006GL027706.
- Treuhaft, R. N. (1992). "Tropospheric and charged particle propagation errors in very long baseline interferometry." *Proceedings of the Symposium Refraction of Transatmospheric Signals in Geodesy*, eds. J. C. de Munck, and T. A. TH. Spoelstra, No. 36, Netherlands Geodetic Commission, The Hague, The Netherlands, May 19–22, ISBN 9789061322436, pp. 45–53. <http://www.ncg.knaw.nl/eng/publications/Geodesy/36DeMunck.html>.
- Uppala, S. M., P. W. Kållberg, A. J. Simmons, U. Andrae, V. Da Costa Bechtold, M. Fiorino, J. K. Gibson, J. Haseler, A. Hernandez, G. A. Kelly, X. Li, K. Onogi, S. Saari-
nen, N. Sokka, R. P. Allan, E. Andersson, K. Arpe, M. A. Balmaseda, A. C. M. Beljaars, L. van de Berg, J. Bidlot, N. Bormann, S. Caires, F. Chevallier, A. De-
thof, M. Dragosavac, M. Fisher, M. Fuentes, S. Hagemann, E. Hólm, B. J. Hoskins, L. Isaksen, P. A. E. M. Janssen, R. Jenne, A. P. McNally, J.-F. Mahfouf, J.-J. Mor-
crette, N. A. Rayner, R. W. Saunders, P. Simon, A. Sterl, K. E. Trenberth, A. Untch, D. Vasiljevic, P. Viterbo, and J. Woollen (2005). "The ERA-40 re-analysis." *Quarterly Journal of the Royal Meteorological Society*, Vol. 131, No. 612, pp. 2961–3012, doi:10.1256/qj.04.176.

- Urschl, C., R. Dach, U. Hugentobler, S. Schaer, and G. Beutler (2005). “Validating ocean tide loading models using GPS.” *Journal of Geodesy*, Vol. 78, No. 10, pp. 616–625, doi:10.1007/s00190-004-0427-9.
- Vedel, H. (2000). “Conversion of WGS84 geometric heights to NWP model HIRLAM geopotential heights.” Scientific Report 00-04, Danish Meteorological Institute, Copenhagen, Denmark, ISBN 8774784072, <http://www.dmi.dk/dmi/sr00-04.pdf>.
- Vedel, H., K. S. Mogensen, and X. Y. Huang (2001). “Calculation of zenith delays from meteorological data comparison of NWP model, radiosonde and GPS delays.” *Physics and Chemistry of the Earth, Part A*, Vol. 26, No. 6-8, pp. 497–502, doi:10.1016/S1464-1895(01)00091-6.
- Vey, S., E. Calais, M. Llubes, N. Florsch, G. Woppelmann, J. Hinderer, M. Amalvict, M. F. Lalancette, B. Simon, F. Duquenne, and J. S. Haase (2002). “GPS measurements of ocean loading and its impact on zenith tropospheric delay estimates: a case study in Brittany, France.” *Journal of Geodesy*, Vol. 76, No. 8, pp. 419–427, doi:10.1007/s00190-002-0272-7.
- van der Wal, A. D. (1995). *Evaluation of Strategies for Estimating Residual Neutral-Atmosphere Propagation Delay in High Precision Global Positioning System Data Analysis*. M.Sc.E. thesis, University of New Brunswick, Dept. of Geodesy and Geomatics Engineering, Fredericton, N.B., Canada, November, 106 pp., Technical Report 177, <http://gge.unb.ca/Pubs/TR177.pdf>.
- Wallace, J. M., and P. V. Hobbs (2006). *Atmospheric Science: An Introductory Review*. 2nd ed., 504 pp., Academic Press, Amsterdam, ISBN 978-0-12-732951-2.
- Weisstein, E. W. (2008). “Gaussian curvature.” (Online), MathWorld — A Wolfram Web Resource. <http://mathworld.wolfram.com/GaussianCurvature.html>
- van der Werf, S. Y. (2003). “Ray tracing and refraction in the modified US1976 atmosphere.” *Appl. Opt.*, Vol. 42, No. 3, pp. 354–366, doi:10.1364/AO.42.000354.
- Wittmann, A. D. (1997). “Astronomical refraction: formulas for all zenith distances.” *Astronomische Nachrichten*, Vol. 318, No. 5, pp. 305–312, doi:10.1002/asna.2113180507.
- Wolfe, D. E., and S. I. Gutman (2000). “Developing an operational, surface-based, GPS, water vapor observing system for NOAA: Network design and results.” *Journal of Atmospheric and Oceanic Technology*, Vol. 17, No. 4, pp. 426–440, doi:10.1175/1520-0426(2000)017<0426:DA0SBG>2.0.CO;2.
- Yan, H. (1996). “A new expression for astronomical refraction.” *Astronomical Journal*, Vol. 112, No. 3, pp. 1312–1316, doi:10.1086/118102.

- Yan, H., and J. Ping (1995). "The generator function method of the tropospheric refraction corrections." *The Astronomical Journal*, Vol. 110, No. 2, pp. 934–939, doi:10.1086/117574.
- Yan, H., and G. Wang (1999). "New consideration of atmospheric refraction in laser ranging data." Vol. 307, No. 3, pp. 605–610, doi:10.1046/j.1365-8711.1999.02618.x.
- Yan, H., G. Zhang, P. Guo, and Z. Hong (2002). "Discussion and comparison of the mapping functions in radio frequencies." *Terrestrial, Atmospheric and Oceanic Sciences*, Vol. 13, No. 4, pp. 563–575, <http://tao.cgu.org.tw/pdf/v134p563.pdf>.
- Yatsenko, A. Yu. (1995). "Exact theory of astronomical refraction for the real atmosphere." *Astronomy and Astrophysics Supplement*, Vol. 111, pp. 579–586, <http://adsabs.harvard.edu/abs/1995A&AS..111..579Y>.
- Young, A. T. (2004). "Sunset Science. IV. Low-altitude refraction." *The Astronomical Journal*, Vol. 127, No. 6, pp. 3622–3637, doi:10.1086/420806.
- Young, A. T. (2006). "Understanding astronomical refraction." *The Observatory*, Vol. 126, pp. 82–115, <http://adsabs.harvard.edu/abs/2006Obs...126...82Y>.
- Zhang, J., and G. Lachapelle (2001). "Precise estimation of residual tropospheric delays using a regional GPS network for real-time kinematic applications." *Journal of Geodesy*, Vol. 75, No. 5, pp. 255–266, doi:10.1007/s001900100171.

Appendices

“The Truth is in the details.”
— (anonymous)

Appendix I

Ray-tracing

In the present chapter we discuss the “nuts and bolts” of calculating the neutral atmosphere propagation delay. We aim at providing enough detail so that the interested reader may replicate the results presented throughout this work.

I.1 Delay formulation

The total delay can be formulated in two mathematically equivalent ways: eqs. (2.10) and (2.11). That mathematical equivalence does not necessarily carry over to the realm of numerical calculations performed at finite precision, though. In the two subsections that follow, we discuss the numerical calculation of the total delay according to each of eqs. (2.10) and (2.11).

I.1.1 Calculating the total delay as apparent ray-path length minus geometric distance

Following section 2.2, the total delay is defined as eq. (2.10):

$$d = \int_{\text{bent ray-path}} n(\ell) d\ell - \int_{\text{straight line}} 1 d\ell, \quad (2.10)$$

Let's see how to calculate each term in eq. (2.10).

The second term in eq. (2.10) is much easier to calculate than the first one, because (i) it is an integral whose integrand is unity and (ii) its entire integration path is known beforehand; therefore the integral simplifies to the difference in its limits:

$$\int_{\text{straight line}} 1 d\ell = \ell_{\text{max,sl}} - \ell_{\text{min,sl}}. \quad (I.1)$$

If we arbitrarily set the lower limit as zero:

$$\ell_{\text{min,sl}} \equiv 0, \quad (I.2)$$

then the upper limit becomes, very conveniently, the straight-line distance between receiver and satellite, which (assuming that the bent ray-path does contain both receiver and satellite) can be readily calculated as:

$$\ell_{\text{max,sl}} = \left| \mathbf{r}_{\text{sat}}^{\text{inertial}}(t_{\text{trans}}) - \mathbf{r}_{\text{rec}}^{\text{inertial}}(t_{\text{rec}}) \right|, \quad (I.3)$$

where $\mathbf{r}_{\text{rec}}^{\text{inertial}}$ and $\mathbf{r}_{\text{sat}}^{\text{inertial}}$ are the position vectors of receiver and satellite, respectively.

Two notes need to be put in place regarding eq. (I.3). The first one is that, even under Newtonian physics, the model of the ray-path in vacuum as a straight-line requires that

the position vectors be expressed in an inertial reference frame. That is opposed to common practice in GPS, in which one employs a non-inertial — Earth-centered, Earth-fixed (i.e., co-rotating) — frame. Here we assume that a suitable coordinate transformation was performed.¹ Otherwise, e.g., the straight-line ray-path would become a spiral when represented in a rotating reference frame. The second note concerns the fact that the receiver's and satellite's positions are to be taken at different epochs, $t_{\text{trans}} \neq t_{\text{rec}}$. More specifically, the satellite's position is to be taken at the signal's transmission epoch t_{trans} , which is obviously earlier than the receiver's reception epoch, t_{rec} . The bulk of the propagation time $t_{\text{rec}} - t_{\text{trans}}$ is due simply to the physical distance between satellite and receiver, and also, to a minor extent, due to numerous atmospheric and electric delays and advances. However small might be the propagation time, it is more than enough for the orbiting satellite to displace significantly. In what follows, we assume that the two notes above have been either properly taken care of whenever relevant or that the need for them has been avoided otherwise (see Appendix I.1.2). Therefore we may unclutter the notation in eq. (I.3) rewriting it as:

$$\ell_{\text{max,sl}} = |\mathbf{r}_{\text{sat}} - \mathbf{r}_{\text{rec}}|. \quad (\text{I.4})$$

Going back to eq. (2.10), its first term is much harder to calculate than the second one, because of primary and secondary difficulties. The primary difficulty is that the integrand in the first term, $n(\ell)$ is unknown beforehand; in other words, we make no assumption about the variation of the index of refraction n over the along-path distance (other than the assumption that $n(\ell)$ is continuous). Therefore, that integral cannot be simplified to a closed-form expression and requires numerical methods to be evaluated. A secondary difficulty, different yet closely related to the primary one, is the fact that the path over which the integral is to be evaluated is generally unknown beforehand. In other words, in

¹Considering the magnitude of the velocities and masses involved, the transformation is more rigorously obtained according to the theory of general relativity [McCarthy and Petit, 2004, Chapter 10].

general we do not know the position $\mathbf{r}(\ell)$ of a point along the ray-path at a given distance ℓ ; in its turn, that position $\mathbf{r}(\ell)$ is required before we can obtain the refractivity n at the given distance ℓ . That difficulty will require numerical methods to solve the differential equations describing the ray evolution, subject to specified boundary conditions. In Appendices I.2 and I.3 we discuss, respectively, how to overcome the primary and secondary difficulties related to the first term in eq. (2.10).

I.1.2 Calculating the total delay as along-path delay plus geometric delay

Still following section 2.2, the same total delay eq. (2.10) can be formulated as eq. (2.11):

$$d = d_a + d_g, \quad (2.11)$$

where d_a is the along-path delay:

$$d_a = 10^{-6} \int_{\text{bent ray-path}} N(\ell) d\ell, \quad (2.12)$$

and d_g is the geometric delay:

$$d_g = \int_{\text{bent ray-path}} 1 d\ell - \int_{\text{straight line}} 1 d\ell, \quad (2.13)$$

There are similarities between eqs. (2.10) and (2.11). The along-path delay d_a suffers the same difficulties as the first term in eq. (2.10), namely, the lack of prior knowledge about the integrand variation and about the integration path; therefore d_a shares the same remedies for those two difficulties, namely, numerical methods for integrals and for differ-

ential equations. The geometric delay d_g is similar to the second term in eq. (2.10), in the sense that it is expressed in integrals whose integrands are unity; therefore, such integrals can be simplified to the difference in their limits:

$$\int_{\text{straight line}} 1 \, d\ell = \ell_{\max,\text{sl}} - \ell_{\min,\text{sl}}, \quad (\text{I.5})$$

$$\int_{\text{bent ray-path}} 1 \, d\ell = \ell_{\max,\text{rp}} - \ell_{\min,\text{rp}}. \quad (\text{I.6})$$

Notice that the integral limits are different for different integration paths; in other words, even if we set both lower limits to zero:

$$\ell_{\min,\text{sl}} \equiv 0,$$

$$\ell_{\min,\text{rp}} \equiv 0,$$

the upper limits will still be different, in general; in fact:

$$\ell_{\max,\text{rp}} \geq \ell_{\max,\text{sl}},$$

as discussed in section 2.2.2. Notice also that the path over which the first integral in the geometric delay d_g is evaluated is the actual, bent ray-path, more complicated than the straight-line path over which the second integral in d_g is evaluated. That does not represent any additional effort, though, because we can readily reuse the same bent ray-path previously obtained for use in the along-path delay d_a (see Appendix I.5.2).

There are also dissimilarities between eqs. (2.10) and (2.11). To discuss them, it helps to realize that the actual, bent ray-path can be split up into two parts, each lying in the atmosphere and in outer space. That splitting is possible and desirable because the Earth's

neutral atmosphere eventually vanishes for all practical purposes at a certain height, far below the GPS satellites (not to mention far below the quasar sources employed in VLBI, at practically infinite distances from the Earth). Therefore, with some care, we may avoid having to explicitly integrate over the outer space portion of the ray-path.

Recall from section 2.1 that in outer space (near vacuum), the index of refraction is virtually unity, $n = 1$, while refractivity is zero, $N = 0$. Consequently, while the apparent ray-path length in eq. (2.10) includes a potentially very large part due to the outer space portion of the ray-path, the along-path delay in eq. (2.11) intentionally neglects that part (or, more correctly, leaves it to be taken care of by the geometric delay, as per the rearranging of terms in eq. (2.11)). Such a rearranging brings significant benefits in terms of the numerical calculations involved: if our ultimate interest is the delay, not the path lengths upon which it is defined, then integrating the tiny delay is much more numerically stable than it would be to integrate the huge path length. By numerical stability we mean whether the integration algorithm is able to provide a sufficiently accurate approximate solution, given a tolerance as stop criterion. Equation (2.10) also suffers from a secondary numerical malady, namely, the loss of precision after subtracting two very large quantities, which is also avoided in the along-path delay in eq. (2.11).

Now turning to the geometric delay d_g , at first sight it might seem that it suffers from the same loss of precision due to a subtraction, as eq. (2.10); at closer look, though, we realize that in outer space both index of refraction and refractivity are uniform or, in other words, their gradients are null. Consequently, the outer space portion of the bent ray-path is a straight-line segment, which we may avoid having to calculate explicitly. The subtraction in d_g can involve, after reasonable assumptions, much smaller operands than the ones involved in eq. (2.10), namely, only the portion of the bent ray-path *not* in outer space and a suitable corresponding portion of the straight-line path; see Appendix I.5.2 for details.

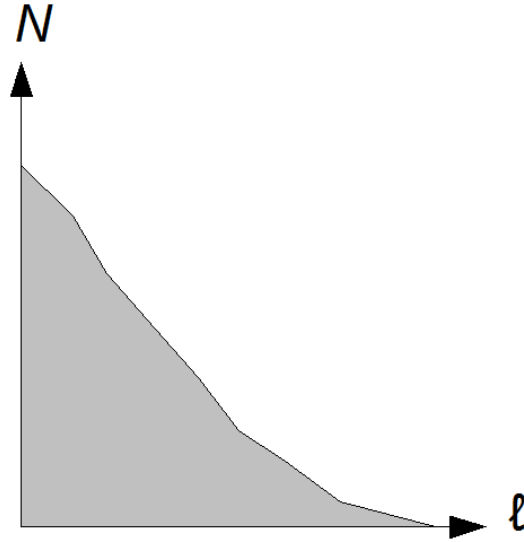


Figure I.1: Along-path delay as the area under the curve of refractivity vs. along-path distance.

I.2 Solver

In the present section we would like to discuss how the options available for ray-tracing are used together in practice, as part of a general procedure for the computation of the along-path delay. Let us start recalling that the along-path delay (eq. (2.12)) represents the area under the curve of refractivity N versus along-path distance ℓ , between upper and lower limits, as shown in Figure I.1. The numerical integration routine therefore requires as input (i) a function representing the integrand $N(\ell)$ and (ii) the integration limits. Given those inputs, the numerical integration routine would construct the curve by requesting to the integrand function $N(\ell)$ for samples of refractivity N at specific along-path distance values ℓ , i.e., $N_i = N(\ell_i)$. The sampling strategy is discussed in Appendix I.5.1. The choice of integration limits is discussed in Appendix I.3.1. Below we would like to discuss how the integrand function $N(\ell)$ can be implemented.

The job of the integrand function $N(\ell)$ is to provide a refractivity value N at an arbitrary along-path distance ℓ . But N and ℓ are only the input and output — there are other

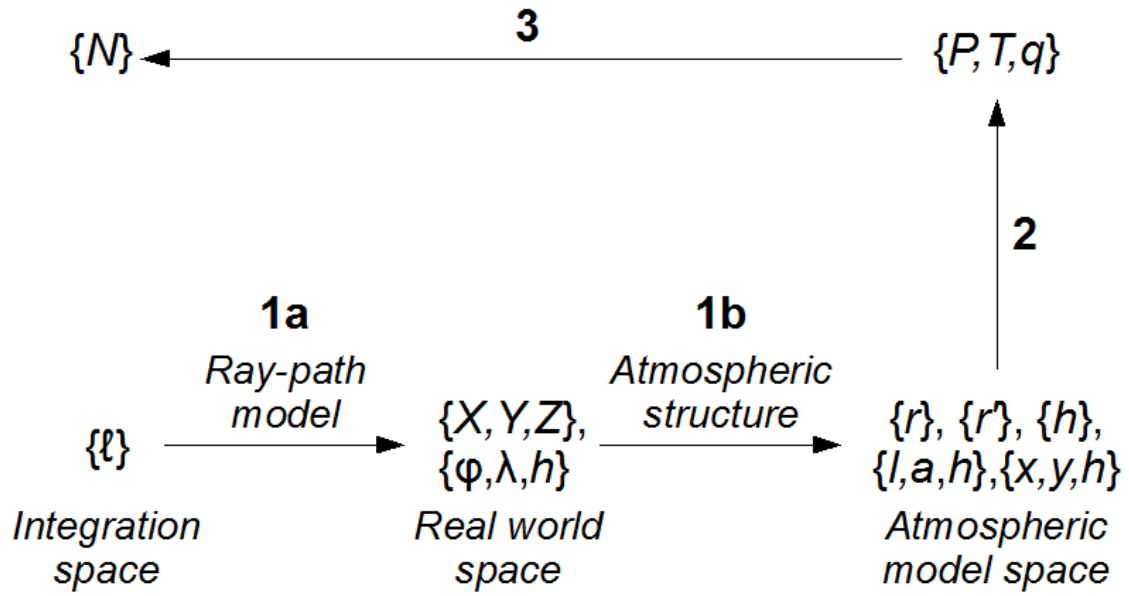


Figure I.2: Diagram of steps in delay solver.

intermediary results involved, as shown in Figure I.2. Let us see how to go from ℓ to N , but backwards (i.e., from output to input).

An individual refractivity value N is calculated from the respective atmospheric parameters pressure P , temperature T , and humidity q , corresponding to a common location. The calculation is done according to the refractivity formulation introduced in section 2.1, eq. (2.3). This step is depicted as arrow number 3 in Figure I.2.

The atmospheric parameters P, T, q at a particular location, in their turn, are provided by the atmospheric model. Please note that we require the atmospheric model to be able to provide atmospheric parameters at any reasonable location. If the atmospheric model is made available as a discretization in profiles, tables, or grids, then to fulfil that requirement we would interpolate among the nodes making up the discretization. If the atmospheric model is made available in analytical form (e.g., as an expansion in volumetric spherical harmonics), then one would simply evaluate those harmonics at the desired location. This step is depicted as arrow number 2 in Figure I.2.

Surely one could try to merge steps 3 and 2, pre-computing refractivity for the whole atmospheric model and then interpolating directly refractivity instead of the underlying atmospheric parameters upon which it depends; that would bring the benefit of a reduced number of interpolations (one N or two N_h, N_{nh} , instead of three P, T, q); on the downside, unless the discretization of the atmospheric model is sufficiently dense, it is more accurate to interpolate separately in the underlying atmospheric parameters.

To interpolate in or evaluate the atmospheric model, we need the location at which the atmospheric parameters are sought. Such a location must be expressed in the same coordinate space in which the atmospheric model itself is provided. It is at this step that the atmospheric structures, introduced in section 3.2, are implemented. Recall that if the atmospheric structure is profile-based, then its coordinate space is simply a vertical coordinate, different for spherical concentric (r), spherical osculating (r'), and ellipsoidal (h) atmospheres; if the atmospheric structure is of the so-called gradient type, then its coordinate space is expressed with a vertical coordinate h as well as with special coordinates l, a describing the loxodrome joining the sought-after location and the vertical profile at a base point; finally, if we make no assumption about the atmospheric structure, a case which we call “3d”, then the location is expressed in whatever coordinate space that atmospheric model is provided (e.g., a vertical coordinate h plus horizontal coordinates x, y in a particular map projection). It is interesting to realize that, in spite of some atmospheric structures being uni-variate (i.e., parametrized in terms of a single variable), all atmospheric models are three-dimensional, in the sense that they are embed in a real-world space. This step is depicted as arrow number 1b in Figure I.2.

The initial step is to convert from the along-path distance ℓ to a position in the real-world, expressed in either geodetic coordinates φ, λ, h or global Cartesian coordinates X, Y, Z . It is at this step that the ray-path models, introduced in section 3.1, are implemented. The zenith and straight-line ray-path models are significantly simpler than the

bent-2d or bent-3d models, because they reveal us the entire ray-path beforehand. In contrast, in the bent models, to know the path, we must proceed step-by-step or iteratively. This step is depicted as arrow number 1a in Figure I.2.

It helps to distinguish among the three different coordinate spaces involved in ray-tracing: the integration space (ℓ), the real-world space ($\{\varphi, \lambda, h\}$ or $\{X, Y, Z\}$), and the atmospheric model space ($\{r\}, \{r'\}, \{h\}, \{l, a, h\}$, or $\{x, y, h\}$). Surely one could try to convert directly from the integration space to the atmospheric space; that seems sensible at least for the popular combination of a zenith ray-path model with a spherical osculating atmosphere structure, for which the direct conversion would be simply, e.g., $r = r_0 + \ell$. In general, though, it helps to convert to the intermediary real-world space first; that allows us to freely combine any ray-path model with any other atmospheric structure. In fact, each ray-path and atmospheric structure implementation need not be aware of the details in any other — all they need to work together is to honour a set of input and output values.

I.3 Bent ray-path

I.3.1 Integration limits and orientation

Whereas in reality the ray travels from the transmitting satellite to the receiver, we adopt a reversed orientation for the ray-path, i.e., from receiver to satellite. That reversal is permitted because the neutral atmosphere is passive to the microwave radiation employed in GPS; otherwise, the atmosphere would possibly change due to the radiation's energy. That reversal is desired mainly for two reasons, which we now discuss.

In Appendix I.1.2 we have seen that it is beneficial to integrate only the portion of the ray-path *not* in outer space. If we set the integration lower limit to the receiver, then it does not matter much where exactly we set the integration upper limit, because a change of up to a few km in the location of the interface between neutral atmosphere and outer space

will not affect the integrated along-path delay at all, as refractivity N is certainly already sufficiently close to zero at such heights. In contrast, if we do not reverse the integration limits, then the value for the integration upper limit becomes important: it must correspond exactly to the receiver position, and that value is very hard to know beforehand, at least on a bent ray-path.

In Appendix I.3.2 we will see that most of the bending happens close to the receiver, near the ground, and that the bent ray-path naturally converges to a straight-line aloft. Therefore it is easier to reach the satellite starting from the receiver than the other way around. Furthermore, the consequence of missing the satellite by 1 km yields a much smaller discrepancy in delay than if we miss the receiver by the same 1 km; that is simply because the satellite lies in outer space (near) vacuum, whereas close to the receiver, near the ground, the gradient of the atmospheric parameters is strongest.

By the reasons stated above, we will set the integration lower limit to zero, corresponding to the receiver position, and the integration upper limit to the approximate distance at which the ray pierces through a nominal interface between neutral atmosphere and outer space, which we define more precisely in the next section.

I.3.2 Notable points and directions

In Figure I.3 we see (i) the origin of the global Cartesian coordinate system (close to the Earth's center, roughly speaking); (ii) the receiver and its position vector, \mathbf{r}_{rec} ; (iii) the satellite and its position vector, \mathbf{r}_{sat} ; (iv) the direction from receiver to satellite, $\Delta\mathbf{r}_{\text{sat}} \equiv \mathbf{r}_{\text{sat}} - \mathbf{r}_{\text{rec}}$. The direction $\Delta\mathbf{r}_{\text{sat}}$ is expressed in global Cartesian coordinates; upon a suitable rotation in the coordinate system, the same direction can be expressed in local Cartesian coordinates (north, east, up), which we can convert from Cartesian to polar coordinates to obtain the elevation angle, azimuth, and distance to the satellite, as seen from the receiver.

The bent ray-path model introduces a few more notable points and directions, shown

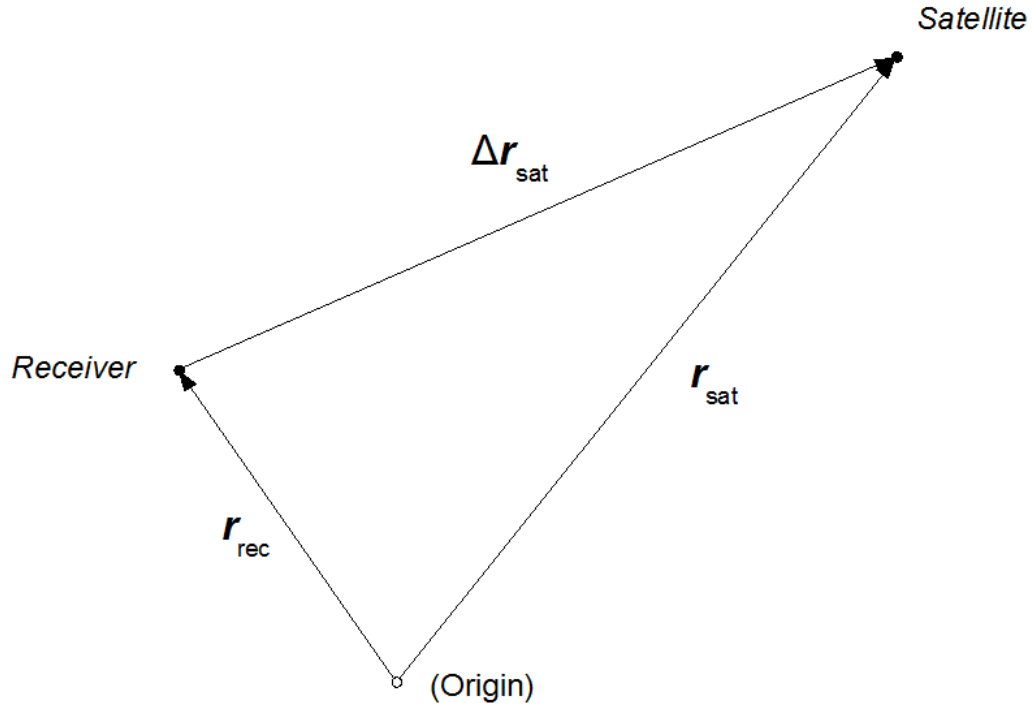


Figure I.3: Receiver–satellite geometry.

respectively in Figures I.4 and I.5. In those figures we see a sample bent ray-path reaching the receiver. Following Appendix I.3.1, we adopt the reversed orientation for the ray-path, i.e., *from receiver to satellite*, instead of the other way around, as it happens in reality. Notice that most of the bending occurs close to the receiver, because near the ground the gradient of the atmospheric parameters is strongest (both the vertical and horizontal components of the gradient), and that the bent ray approaches an asymptote as the ray reaches less dense and smoother air parcels aloft.

It is convenient to define \mathbf{r}_{exit} , known as *pierce point*: it is the point along the bent ray-path at which the ray exits the neutral atmosphere, corresponding to the along-path distance ℓ_{exit} . To obtain ℓ_{exit} , we assume that the neutral atmosphere vanishes at an approximate height H ,² and that the ray-path is approximately a straight-line, and that the

²We follow the recommendation from Mendes [1999, p. 166–167], who found that “The contribution of the [neutral] atmosphere above 75 km can be ignored, even though we kept the 100 km boundary as the upper limit in our ray-trace computations.” As a middle ground, we adopted the value of $H = 80$ km.

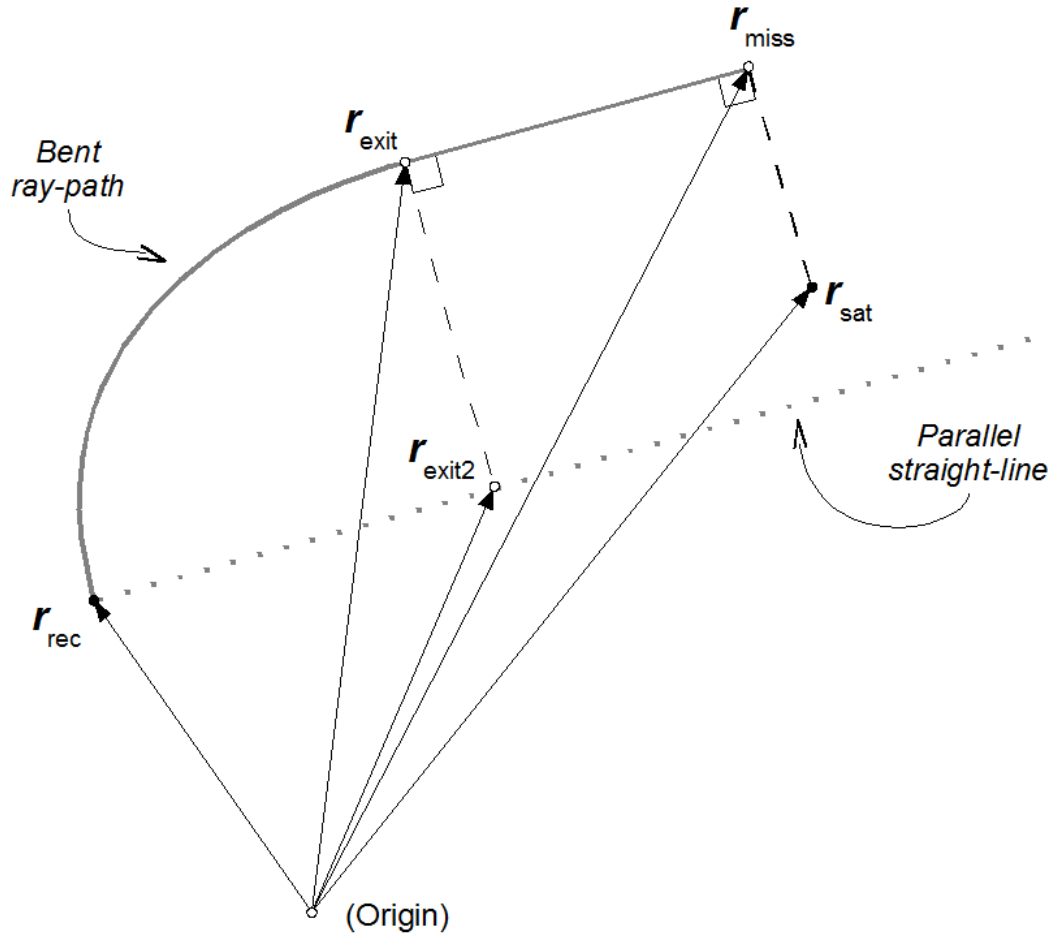


Figure I.4: Notable positions in a bent ray-path.

solid Earth is approximately a sphere with radius R . The assumptions above simplify considerably the calculations below; however gross these assumptions might seem, they are harmless as per discussion in Appendix I.3.1. We calculate ℓ_{exit} applying the law of cosines to the triangle with vertices at the origin of the global Cartesian coordinate system, at the receiver \mathbf{r}_{rec} , and at the (unknown) approximate pierce point, as shown in Figure I.6 and formulated as:

$$(R + H)^2 = R^2 + \ell_{\text{exit}}^2 - 2R\ell_{\text{exit}} \cos \theta, \quad (\text{I.7})$$

where the factor $\cos \theta$ is known from:

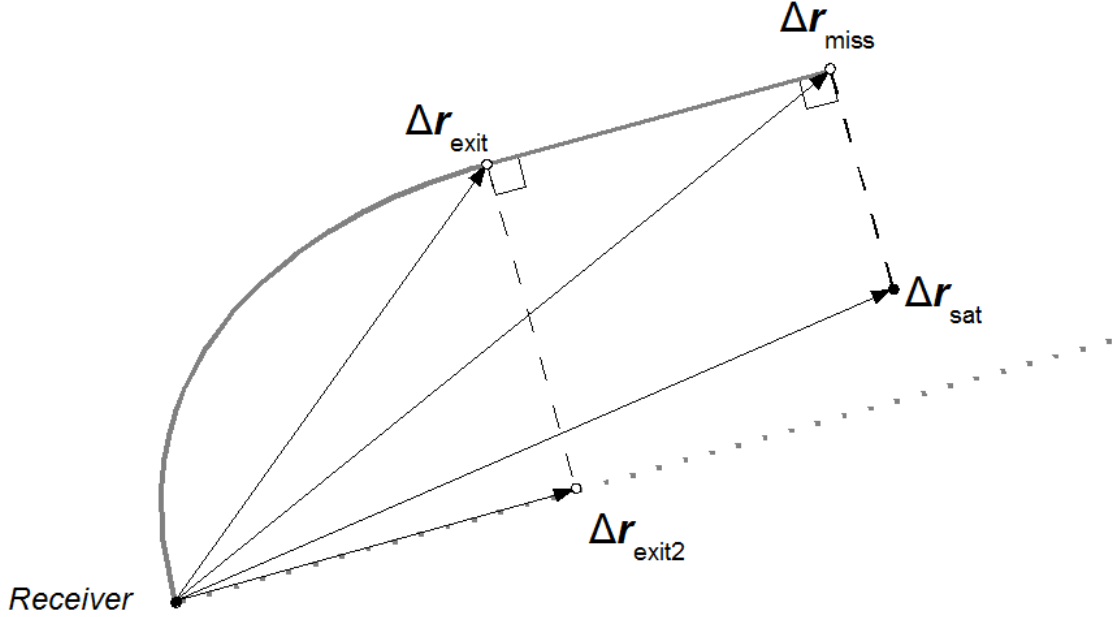


Figure I.5: Notable relative directions in a bent ray-path.

$$\cos \theta = \frac{-\mathbf{r}_{\text{rec}} \cdot \Delta \hat{\mathbf{r}}_{\text{sat}}}{|\mathbf{r}_{\text{rec}}| |\Delta \mathbf{r}_{\text{sat}}|}. \quad (\text{I.8})$$

Finally, we solve the quadratic eq. (I.7) and pick the positive root for ℓ_{exit} . Please note that such an approximate pierce point, given simply by $\mathbf{r}_{\text{rec}} + \ell_{\text{exit}} \Delta \hat{\mathbf{r}}_{\text{sat}}$, is different than the rigorous point along the bent ray-path corresponding to the same distance ℓ_{exit} , $\mathbf{r}_{\text{exit}} = \mathbf{r}_0 + \int_0^{\ell_{\text{exit}}} \hat{\mathbf{s}}(\ell) d\ell$ (see eq. (I.15) on p. 174).

We also define the ray's notable *tangent directions* $\hat{\mathbf{s}}$, shown in Figure I.7. They are distinct from the *relative directions* $\Delta \mathbf{r}$ reckoned from the receiver, shown in Figure I.5. The notable tangent directions are: $\hat{\mathbf{s}}_0$, the initial ray direction, also known as apparent direction because it is the direction at which the satellite would be seen (if it were visible and if the ray did); $\hat{\mathbf{s}}_{\text{exit}}$, the tangent direction at the pierce point; and $\hat{\mathbf{s}}_{\text{miss}}$, the tangent direction at the point of closest approach (point which is defined in the next paragraph below). The angle $\arccos(\hat{\mathbf{s}}_0 \cdot \hat{\mathbf{s}}_{\text{exit}})$ is the so-called *bending angle*.

Notice that the tangent directions $\hat{\mathbf{s}}_{\text{miss}} = \hat{\mathbf{s}}_{\text{exit}}$, i.e., beyond the pierce point \mathbf{r}_{exit} the

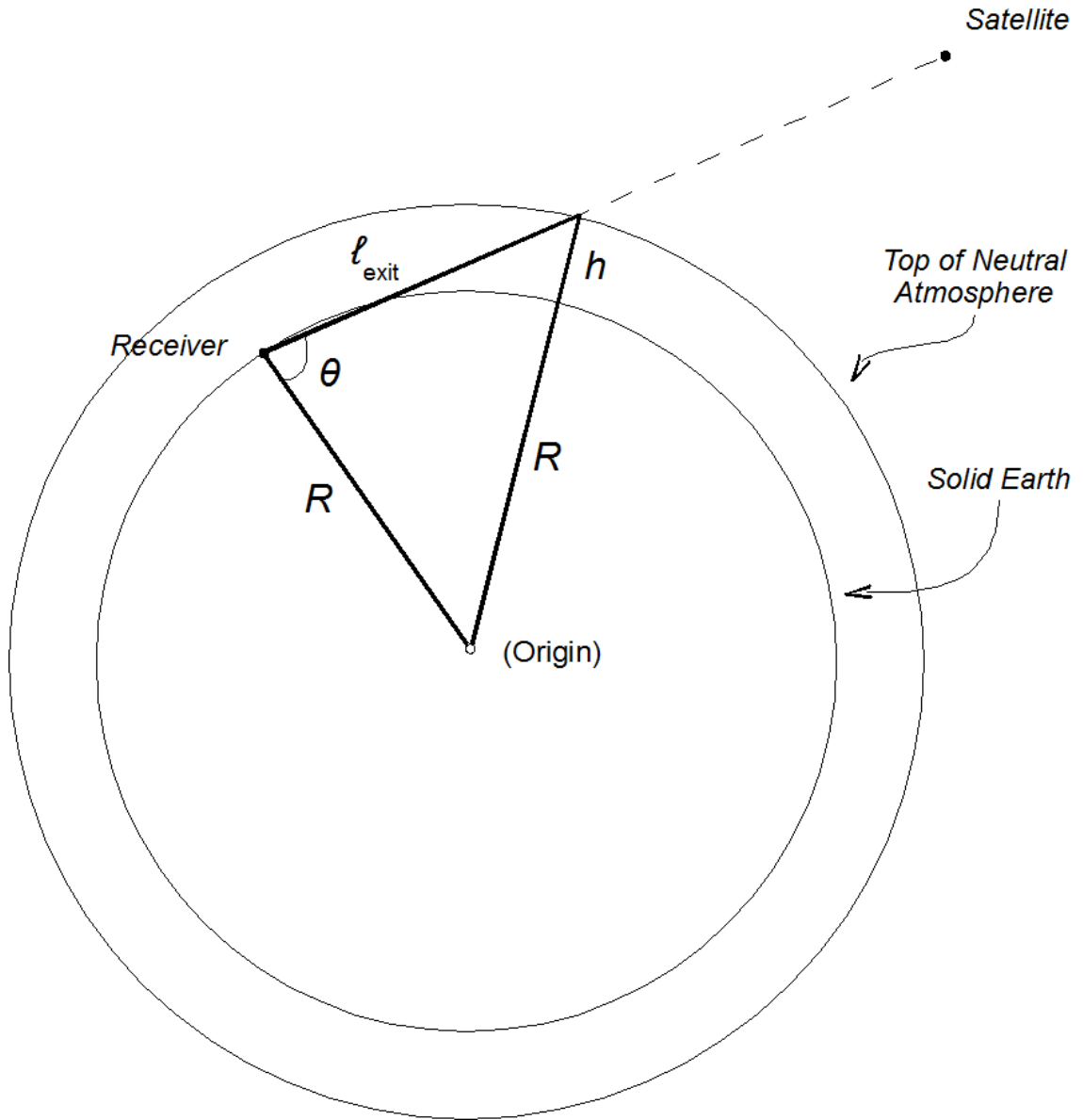


Figure I.6: Definition of maximum along-path distance.

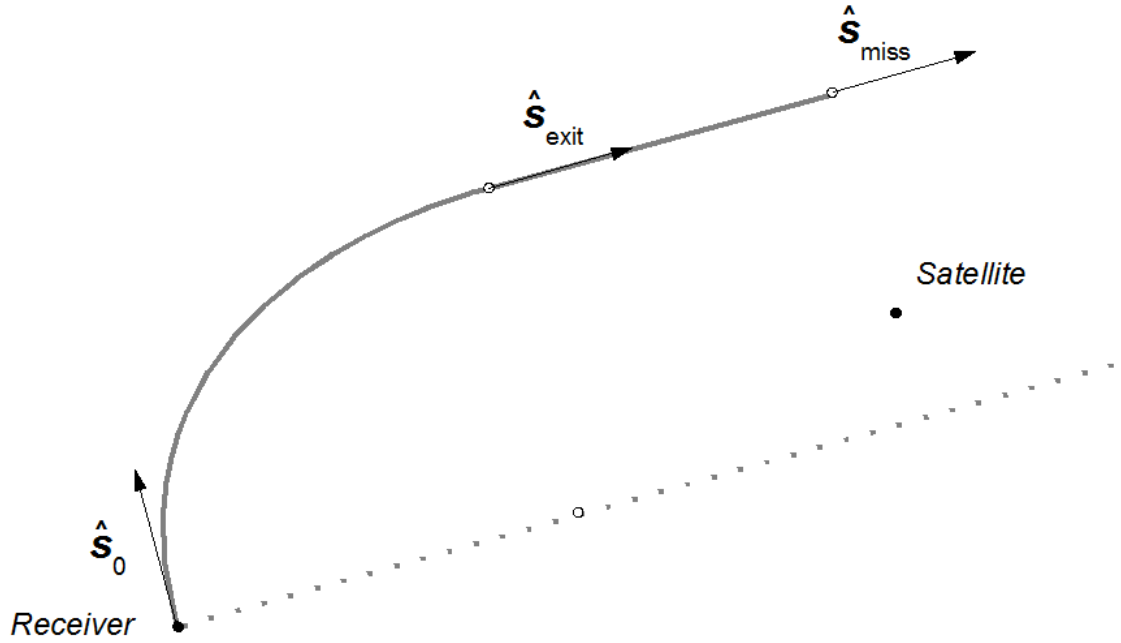


Figure I.7: Notable tangent directions in a bent ray-path.

ray ceases to change direction. In other words: the portion of the bent ray-path beyond the pierce point \mathbf{r}_{exit} — i.e., in outer space — is a straight-line segment. Notice further that the ray does not necessarily reach the satellite. Therefore we define \mathbf{r}_{miss} , the point of closest approach between the bent ray-path and the satellite. The working definition for \mathbf{r}_{miss} is the orthogonal projection of the satellite position \mathbf{r}_{sat} onto the ray-path's outer space straight-line segment, segment which is defined as $\mathbf{r}_{\text{exit}} + d\hat{\mathbf{s}}_{\text{exit}}$ for any distance d . The exact distance d_{miss} corresponding to the position \mathbf{r}_{miss} can be calculated as $d_{\text{miss}} \equiv (\mathbf{r}_{\text{sat}} - \mathbf{r}_{\text{exit}}) \cdot \hat{\mathbf{s}}_{\text{exit}}$, from which \mathbf{r}_{miss} itself is calculated as $\mathbf{r}_{\text{miss}} \equiv \mathbf{r}_{\text{exit}} + d_{\text{miss}}\hat{\mathbf{s}}_{\text{exit}}$.

At last, to help in calculating the geometric delay (Appendix I.5.2), we define $\mathbf{r}_{\text{exit}2}$, the orthogonal projection of the pierce point \mathbf{r}_{exit} onto a straight-line passing through the receiver and parallel to the ray-path's outer space straight-line segment (not necessarily parallel to the receiver–satellite direction, $\Delta\mathbf{r}_{\text{sat}}$), segment which is defined as $\mathbf{r}_{\text{rec}} + d\hat{\mathbf{s}}_{\text{exit}}$ for any distance d and represented as a dotted line in Figure I.5. The exact distance $d_{\text{exit}2}$

corresponding to the position $\mathbf{r}_{\text{exit}2}$ can be calculated as $d_{\text{exit}2} \equiv (\mathbf{r}_{\text{exit}} - \mathbf{r}_{\text{rec}}) \cdot \hat{\mathbf{s}}_{\text{exit}}$, from which $\mathbf{r}_{\text{exit}2}$ itself is calculated as $\mathbf{r}_{\text{exit}2} \equiv \mathbf{r}_{\text{rec}} + d_{\text{exit}2} \hat{\mathbf{s}}_{\text{exit}}$.

In the next section (Appendix I.3.3) we see how the notable points and directions above are employed as boundary conditions for the ray evolution.

I.3.3 Boundary conditions

The differential equations for the bent-3d eq. (3.1) and bent-2d eq. (3.4) ray-path models describe nothing but the *evolution* of the ray. A second, essential, ingredient for the actual bent ray-path is a set of *initial* or, more generally, *boundary conditions* that the ray must satisfy. Those conditions are readily expressed in terms of the notable points and directions defined in Appendix I.3.2. Let us see a few examples of boundary conditions.

In the simplest case we would specify an initial position \mathbf{r}_0 and an initial tangent direction $\hat{\mathbf{s}}_0$. The remaining notable points and directions cannot be specified as conditions — they are determined by the ray’s evolution. Therefore when we set the ray’s initial direction $\hat{\mathbf{s}}_0$ there is no guarantee that the ray will reach the satellite. Furthermore, in the practice of GPS usually we do not know the satellite’s apparent direction, only the satellite’s relative direction $\Delta\hat{\mathbf{r}}_{\text{sat}}$. The initial position would be set equal to the receiver position, $\mathbf{r}_0 = \mathbf{r}_{\text{rec}}$. Although this simplest case of initial conditions might seem useless in the practice of GPS (since there is no guarantee that the ray will reach the satellite), it is useful in solving the more complicated case of boundary conditions.

In a more complicated case we would like to specify an initial position \mathbf{r}_0 , again equal to the receiver position \mathbf{r}_{rec} , and a final position, equal to the satellite position \mathbf{r}_{sat} . A final position is ill-defined, though, because there is nothing stopping the ray from progressing indefinitely, increasingly farther away. It is for this reason that in Appendix I.3.2 we have introduced the point of closest approach, \mathbf{r}_{miss} , which can always be defined precisely. The second condition above gets therefore restated as requiring that the point of closest

approach \mathbf{r}_{miss} be equal to the satellite position \mathbf{r}_{sat} . An equivalent condition is requiring that the point of closest approach's relative direction $\Delta\hat{\mathbf{r}}_{\text{miss}}$ equal the satellite's relative direction $\Delta\hat{\mathbf{r}}_{\text{sat}}$. This boundary condition problem can be solved in terms of the initial value problem above, finding among the numerous solutions for the initial value problem $(\mathbf{r}_0, \hat{\mathbf{s}}_0)$ the one solution that also satisfies the present boundary value problem (\mathbf{r}_0 and \mathbf{r}_{miss} , or \mathbf{r}_0 and $\Delta\hat{\mathbf{r}}_{\text{miss}}$); see Appendix I.5.3.2.

We can simplify the rigorous case above under the assumption of a satellite or transmitting source lying at a practically infinite distance, which is completely appropriate in VLBI and a very good approximation in GPS. In such a case the definition of a point of closest approach \mathbf{r}_{miss} and respective relative direction $\Delta\hat{\mathbf{r}}_{\text{miss}}$ and tangent direction $\hat{\mathbf{s}}_{\text{miss}}$ becomes unnecessary. Therefore, as boundary conditions, in addition to a start position equal to the receiver position, $\mathbf{r}_0 = \mathbf{r}_{\text{rec}}$, it suffices to require that the exit tangent direction $\hat{\mathbf{s}}_{\text{exit}}$ equals the satellite (or transmitting source) relative direction $\Delta\hat{\mathbf{r}}_{\text{sat}}$ — the non-overlapping yet parallel straight-lines (expressed as $\mathbf{r}_{\text{exit}} + d\hat{\mathbf{s}}_{\text{exit}}$ and $\mathbf{r}_{\text{rec}} + d\Delta\hat{\mathbf{r}}_{\text{sat}}$ for any distance d) will eventually meet at an infinite distance, $d = \infty$. Under such a simplification, we may call $\hat{\mathbf{s}}_{\text{exit}} = \Delta\hat{\mathbf{r}}_{\text{sat}}$ the satellite's *geometric direction*, in contrast with the satellite's apparent direction $\hat{\mathbf{s}}_0$ (in the general case, an unqualified “geometric direction” would be ambiguous).³

³The difference between the relative directions to the point of closest approach $\Delta\hat{\mathbf{r}}_{\text{miss}}$ and to the exit point $\Delta\hat{\mathbf{r}}_{\text{exit}}$ — or, equivalently, between the corresponding elevation angles — has been recognized previously by at least the following authors. Marini [1972, p. 225] mentions en passant that “it has been assumed that the satellite height is great enough to permit the upper limit [in the bending integral] to be extended to infinity”. Yan [1996, section 5] gives an expression, function of the distance receiver–satellite, to correct for that difference. Yan and Wang [1999, p. 610] quote some numerical values for that correction and conclude that “This term might be considerable for satellites with heights of several hundred km or less that are to be observed at lower elevation angles”. It seems, therefore, negligible in GPS. Indeed, Ifadis [2000, Tab. 3] agrees that “for radio sources at GPS height or greater the difference [in delay] is less than 1 cm at elevations greater than 3°.”

I.3.4 Ray evolution

Now we finally discuss the evolution of the ray. The numerical details of each initial and boundary value problems are discussed in Appendices I.5.3.1 and I.5.3.2. We recall once again the Eikonal equation eq. (3.1):

$$\frac{d}{d\ell} \left(n \frac{d\mathbf{r}}{d\ell} \right) = \nabla n, \quad (3.1)$$

which can be expanded as:

$$\frac{dn}{d\ell} \frac{d\mathbf{r}}{d\ell} + n \frac{d^2\mathbf{r}}{d\ell^2} = \nabla n. \quad (I.9)$$

Here we define, based on the position vector \mathbf{r} , the *tangent vector* $\hat{\mathbf{s}}$:

$$\hat{\mathbf{s}} \equiv \frac{d\mathbf{r}}{d\ell}, \quad (I.10)$$

which is a unit vector (in other words, $|d\mathbf{r}| = d\ell$), and the *curvature vector* \mathbf{a} (usually denoted \mathbf{K}):

$$\mathbf{a} \equiv \frac{d\hat{\mathbf{s}}}{d\ell} = \frac{d^2\mathbf{r}}{d\ell^2}. \quad (I.11)$$

Isolating the curvature vector \mathbf{a} in the expanded Eikonal eq. (I.9) we obtain:

$$\mathbf{a} = \frac{1}{n} \left(\nabla n - \frac{dn}{d\ell} \hat{\mathbf{s}} \right). \quad (I.12)$$

Equation (I.12) tells us that what makes the ray direction $\hat{\mathbf{s}}$ change ($\mathbf{a} \neq \mathbf{0}$) is the component of the gradient of refractivity ∇n *perpendicular* to the current ray direction [Kursinski et al., 2000, eq. (9)]: notice that $dn/d\ell = \nabla n \cdot \hat{\mathbf{s}}$ is the directional derivative of refractivity along the ray, which is subtracted from the full gradient ∇n , thus leaving only the

component of ∇n perpendicular to $\hat{\mathbf{s}}$:

$$\nabla_{\perp \hat{\mathbf{s}}} n \equiv \nabla n - \nabla n \cdot \hat{\mathbf{s}}, \quad (\text{I.13})$$

with which we can rewrite eq. (I.12) compactly as:

$$\mathbf{a} = \frac{\nabla_{\perp \hat{\mathbf{s}}} n}{n}. \quad (\text{I.14})$$

Given initial conditions $\mathbf{r}_0, \hat{\mathbf{s}}_0$, the ray position corresponding to any along-path distance ℓ is obtained as:

$$\mathbf{r}(\ell) = \mathbf{r}_0 + \int_0^\ell \hat{\mathbf{s}}(\ell) d\ell, \quad (\text{I.15})$$

where $\hat{\mathbf{s}} = \mathbf{s}/|\mathbf{s}|$, and \mathbf{s} is:

$$\mathbf{s}(\ell) = \hat{\mathbf{s}}_0 + \int_0^\ell \mathbf{a}(\ell) d\ell, \quad (\text{I.16})$$

so

$$\mathbf{r}(\ell) = \mathbf{r}_0 + \ell \hat{\mathbf{s}}_0 + \int_0^\ell \int_0^\ell \mathbf{a}(\ell) d\ell d\ell. \quad (\text{I.17})$$

What makes the integrals above specially difficult is the fact that the position \mathbf{r} at a given along-path distance ℓ depends on the curvature \mathbf{a} up to ℓ , but the curvature itself depends on the position \mathbf{r} ; to get past this difficulty we need to proceed step-by-step or iteratively; see details in Appendix I.5.3.

I.3.4.1 Special cases

In a propagating medium of uniform refractivity the gradient is null, $\nabla n = \mathbf{0}$, thus the curvature eq. (I.14) is null, $\mathbf{a} = \mathbf{0}$, and the integrated tangent eq. (I.16) is always equal to the initial direction, $\hat{\mathbf{s}}(\ell) = \hat{\mathbf{s}}_0$, for any ℓ ; therefore, in a medium of uniform refractivity, the ray-path is rigorously a straight-line.

In a spherical osculating or ellipsoidal atmosphere, there is no bending for a ray directed at zenith, because $\nabla_{\perp\hat{s}}n = \nabla_H n \equiv \mathbf{0}$, i.e., the component of the gradient perpendicular to the ray direction (zenith) equals the horizontal component of the gradient, which is postulated zero in such atmospheric structures. In contrast, in a gradient or 3d atmosphere, there is a non-zero horizontal component that in principle may yield some bending; therefore the geometric delay may be non-zero at zenith under such atmospheric structures (and a bent-3d ray-path model, of course). The bending at zenith is usually small, though.

In a spherical atmosphere, the ray-path is rigorously a plane curve; in other words, in a spherical atmosphere, there is no out-of-plane bending. First we have to realize that in a spherical atmosphere the gradient of refractivity points to the center of the sphere:

$$\hat{\nabla}n = \frac{\nabla n}{|\nabla n|} = -\hat{r}' = -\frac{\mathbf{r}'}{|\mathbf{r}'|}, \quad (3.2)$$

To demonstrate that the ray-path is a plane curve, following Born and Wolf [1999, p. 130] we demonstrate that $(\mathbf{r} \times n\hat{s})$, which represents (a multiple of) the normal to plane spanned by \mathbf{r} and \hat{s} , is constant throughout the ray-path:

$$\frac{d}{d\ell}(\mathbf{r} \times n\hat{s}) = \frac{d\mathbf{r}}{d\ell} \times \hat{s} + \mathbf{r} \times \frac{d(n\hat{s})}{d\ell}; \quad (I.18)$$

from the Eikonal eq. (3.1), $d(n\hat{s})/d\ell = \nabla n$, which from eq. (3.2) equals $\nabla n = -|\nabla n|\hat{r}'$; therefore:

$$\frac{d}{d\ell}(\mathbf{r} \times n\hat{s}) = \hat{s} \times \hat{s} + \mathbf{r} \times (-|\nabla n|\hat{r}'), \quad (I.19)$$

which equals zero because the cross-product of a vector with (a multiple of) itself is a zero vector. The derivative of a $(\mathbf{r} \times n\hat{s})$ being zero implies such a quantity being constant, which finishes the demonstration.

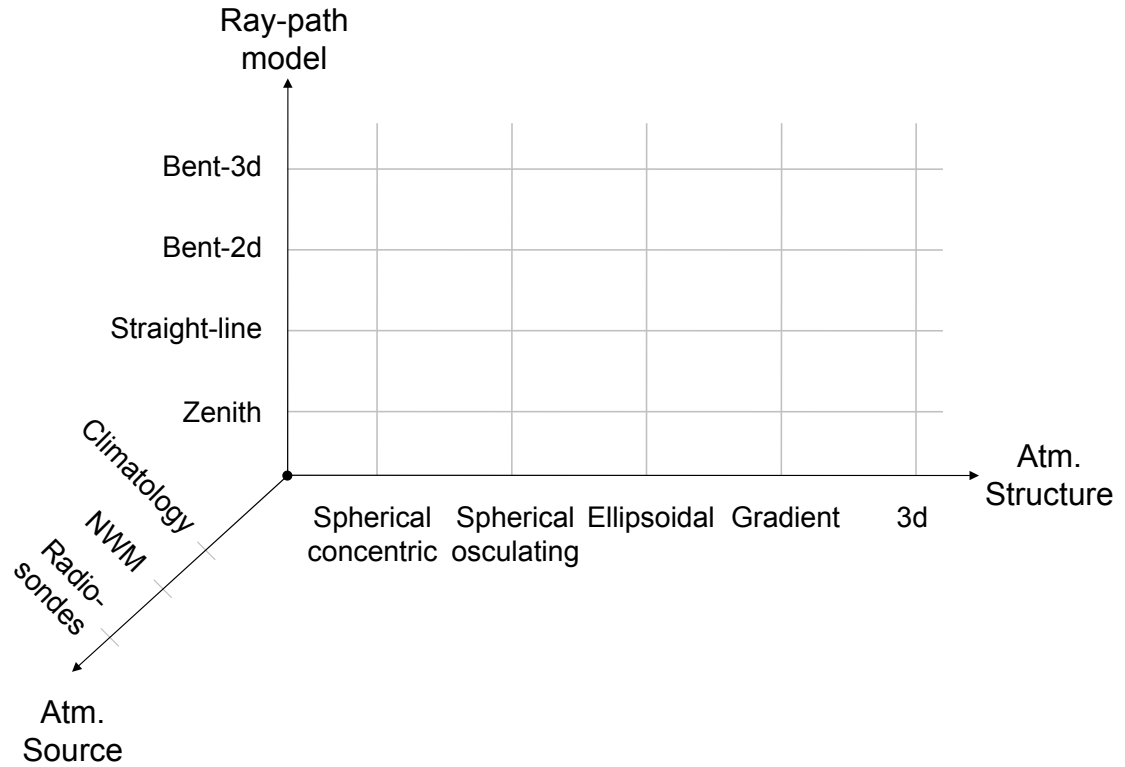


Figure I.8: Diagram of options available in ray-tracing.

I.4 Cross-validation

The diagram of options available in ray-tracing, repeated in Figure I.8 for convenience, shows that there are several different combinations of ray-path and atmospheric structures. In general we expect different combinations to give different results. In the present section, we would like to explore the different combinations for which we expect the *same* results. That redundancy is a terrific opportunity to validate more complicated combinations against simpler ones.

The condition expected to be satisfied in such comparisons is usually equality in the delay given by different combinations. Please note that the comparison conditions must be evaluated considering the precision of the results. For example, if each of two delays was integrated under an absolute tolerance of $\epsilon = 1$ mm, then their equality should be evaluated

numerically as $\text{abs}(d_1 - d_2) < \varepsilon'$, where the tolerance ε' equals $\varepsilon' = \sqrt{2}\varepsilon$, assuming d_1, d_2 uncorrelated.

I.4.1 Cross-validating ray-path models

Table I.1 summarizes the comparisons among different ray-path models; they correspond to horizontal lines in the diagram of options (Figure I.8), starting from the bottom and advancing to the top. Now we take the time to briefly discuss each comparison.

a1 If the ray direction happens to be the zenith direction, then the more general straight-line ray-path model should give the same result as the specialized zenith ray-path model, as the zenith ray-path is a special case of the straight-line ray-path model. That is true for any atmospheric structure and receiver location.

a2 In the zenith direction, the bent-2d ray-path degenerates into a straight-line; as discussed in Appendix I.3.4.1, that is because there is no component of the gradient of refractivity perpendicular to the zenith direction. That is true even when the bent-2d ray-path model is used in conjunction with atmospheric structures which do allow for a non-zero horizontal component in the gradient of refractivity; in this case, the magnitude of the horizontal component will be mis-attributed by the bent-2d ray-path model as an increase in the magnitude of the vertical component of the gradient, since the bent-2d ray-path model assumes that the direction of the full gradient is always completely vertical.

a3 Even though ray bending obviously increases the geometric delay, it also decreases even more the along-path delay; the net result is a decrease in total delay due to bending. Therefore the total delay given by a bent-2d ray-path model should always be smaller than that given by the straight-line ray-path (and equal at zenith).

Table I.1: Comparisons for cross-validating ray-path models.

Number	Ray-path 1	Ray-path 2	Atmospheric structure	Location	Direction	Condition
a1	straight-line	zenith	any	any	zenith	$d_{sl} = d_{zen}$
a2	bent-2d	straight-line	any	any	zenith	$d_{b2d} = d_{sl}$
a3	bent-2d	straight-line	any	any	any	$d_{b2d} \leq d_{sl}$
a4	bent-2d	straight-line	any	any	any	$\text{abs}(d_{b2d} - d_{sl}) 1 / \propto \epsilon$
a5	bent-3d	bent-2d	spherical	any	any	$d_{b3d} = d_{b2d}$
a6	bent-3d	bent-2d	any	any	any	$d_{b3d} \leq d_{b2d}$
a7	bent-3d	bent-3d	any but 3d	any	any	$d_{b3d, \text{with}} = d_{b3d, \text{without}}$
a8	bent-2d	bent-2d	spherical	any	any	no out-of-plane bending

Table I.2: Comparisons for cross-validating atmospheric structures.

Number	Atm. structure 1	Atm. structure 2	Ray-path	Location	Direction	Condition
b1	sph. osculating	sph. concentric	any	poles, equator	any	$d_{osc} = d_{conc}$
b2	sph. osculating	sph. osculating	any	any	any	exhibit azimuthal symmetry
b3	ellipsoidal	sph. osculating	any	any	any	$\text{abs}(d_{ell} - d_{osc}) 1 / \propto \epsilon$
b4	gradient	ellipsoidal	any	any	any	$d_{osc} = d_{conc}$
b5	3d	gradient	any	any	any	$\text{abs}(d_{3d} - d_{grad}) 1 / \propto \epsilon$

a4 The magnitude of the discrepancy in delay between bent-2d and straight-line ray-path models should be the greater the lower the elevation angle is.

a5 Even though the bent-3d ray-path model allows for out-of-plane bending, if such a ray-path model is used in conjunction with an atmospheric structure that postulates zero horizontal gradients, then the results from the more complicated bent-3d should equal those given by the simpler bent-2d ray-path model. That is the case for both spherical atmospheric structures, but not for an ellipsoidal atmosphere (as the up direction does not always coincide with the radial direction), nor in a gradient or 3d atmosphere (of course).

a6 For the same reasons as in comparison a4, the total delay given by a bent-3d ray-path model should be smaller than or equal to the total delay given by a bent-2d ray-path model.

a7 This comparison tells us that we expect the same result when we make a correct assumption about the direction of the gradient of refractivity, compared to obtaining ∇n numerically, with no prior assumption; see Appendix I.5.3.4.

a8 The bent-2d ray-path should yield a plane curve. That property can be checked in the following way. Starting with the position vector \mathbf{r}_i of each point discretizing the bent ray-path:

- obtain the relative directions with respect to the receiver \mathbf{r}_0 :

$$\Delta\mathbf{r}_i \equiv \mathbf{r}_i - \mathbf{r}_0;$$

- obtain the orientation of the initial plane, spanned by the ray initial direction $\hat{\mathbf{s}}_0$, the receiver position \mathbf{r}_0 , and the center \mathbf{r}_c of the particular spherical atmospheric

structure employed:⁴

$$\hat{\boldsymbol{p}}_0 \equiv \hat{\boldsymbol{s}}_0 \times \hat{\boldsymbol{r}}'_0;$$

where $\hat{\boldsymbol{r}}'_0 = \boldsymbol{r}'_0 / |\boldsymbol{r}'_0|$ and $\boldsymbol{r}'_0 = \boldsymbol{r}_0 + \boldsymbol{r}_c$; $\hat{\boldsymbol{p}}_0$ is the normal to the initial plane.

– finally, the out-of-plane component of $\Delta \boldsymbol{r}_i$ is obtained as:

$$\Delta_{\perp} r_i \equiv \Delta \boldsymbol{r}_i \cdot \hat{\boldsymbol{p}}_0,$$

$\Delta_{\perp} r_i$ is a scalar with units of length, which we expect to be zero.

I.4.2 Cross-validating atmospheric structure models

Table I.2 summarizes the comparisons among different atmospheric structures; they correspond to vertical lines in the diagram of options (Figure I.8), starting from the left and advancing to the right.

b1 At the poles and along the equator, the ellipsoidal normal coincides with the radial direction, towards the origin of the coordinate system. Consequently, the osculating sphere coincides with the concentric sphere.

b2 A spherical osculating atmosphere should give the same delay for different azimuths but same elevation angle and receiver position. That is true for any ray-path model.

b3 The spherical osculating atmosphere is a perfect local approximation for the ellipsoidal atmosphere. As long as the ray remains in the vicinity of the base point or point of osculation, results from both atmospheric structures are expected to be the same. In

⁴ $\boldsymbol{r}_c = \{X=0, Y=0, Z=0\}$ for the concentric atmosphere; $\boldsymbol{r}_c \stackrel{XYZ}{\leftarrow}_{\phi\lambda h} (\varphi_0, \lambda_0, h = -R)$ for an osculating atmosphere.

other words, the discrepancy between results from ellipsoidal and spherical osculating atmospheres should be greater the lower the elevation angle is.

b4 This comparison is somewhat artificial: if we force to zero the horizontal gradient employed in a gradient atmosphere, then we expect the more complicated gradient atmosphere to give the same result as the simpler ellipsoidal atmosphere.

b5 Similar rationale as for comparison b3, only different atmospheric structures involved.

I.5 Further numerical aspects

I.5.1 Along-path delay

Whereas the implementation of the integrand $N(\ell)$ involved in the definition of along-path delay d_a (eq. (2.12)):

$$d_a = 10^{-6} \int_{\text{bent ray-path}} N(\ell) d\ell, \quad (2.12)$$

was discussed in Appendix I.2, a pending issue was how best to discretize the curve of refractivity N versus along-path distance ℓ .

Surely the value of the definite integral in eq. (2.12) and the path over which the same integral is evaluated are just two slightly different aspects of the same, more fundamental, differential equation problem. Nevertheless we still find it useful to tackle each of the two problems one at a time. Therefore in the present section we assume that the ray-path is known beforehand (even in the case of a bent ray-path, which we deal with in Appendix I.5.3). The quadrature problem, then, can be stated as follows: given an unknown integrand $N(\ell)$, integration limits ℓ_{\min}, ℓ_{\max} , and a tolerance for the resulting definite inte-

gral, at which locations ℓ_i should we sample the integrand N , so as to fulfill the tolerance?

One could proceed as follows: sample N at arbitrary locations ℓ_i and calculate the resulting definite integral; then densify the sampling and calculate an updated value for the integral; finally, if the absolute discrepancy between the two integral values is smaller than a prescribed tolerance, stop, otherwise, sample yet more.

The question, then, would be where to locate the samples. One simple approach would be sampling N at at regularly spaced ℓ ; then densifying it simply dividing each segment in two. Ideally, though, we would like to avoid two situations: (i) undersampling thus missing relevant regions and (ii) oversampling thus wasting evaluations of the sometimes computationally costly integrand function $N(\ell)$.

The solution to this problem is under the name “adaptive quadrature”, more specifically “*global* adaptive quadrature”. It is able to automatically find the sub-intervals $\ell_i < \ell < \ell_j$ that require denser or less dense samplings vis-à-vis the tolerance set for the integral; furthermore, in contrast with “*local* adaptive quadrature”, the global one is able to choose the next most relevant sub-interval that needs densification amongst *all* the current sub-intervals within the original limits ℓ_{\min}, ℓ_{\max} (for example, near the ground vs. aloft). The best implementation that we could find was that of Espelid [2007], which we have adopted in this work; we refer to that publication for details.

I.5.2 Geometric delay

Recall that the geometric delay d_g is defined (eq. (2.13)):

$$d_g \equiv G - D, \tag{2.13}$$

where G is the geometric ray-path length:

$$G \equiv \int_{\text{bent ray-path}} 1 \, d\ell, \quad (2.8)$$

and D is the geometric distance:

$$D \equiv \int_{\text{straight line}} 1 \, d\ell. \quad (2.7)$$

In Appendix I.1.1 we discussed that it is helpful to split the actual, bent ray-path into two parts, each lying in the atmosphere and in outer space. Therefore we split G as:

$$G = G_{\text{in}} + G_{\text{out}}, \quad (I.20)$$

where each $G_{\text{in}}, G_{\text{out}}$ are defined as (see Figure I.4, repeated here for convenience as Figure I.9):

$$G_{\text{in}} \equiv \int_{\text{bent ray-path}}^{\mathbf{r}_{\text{rec}}}^{\mathbf{r}_{\text{exit}}} 1 \, d\ell, \quad (I.21)$$

$$G_{\text{out}} \equiv \int_{\text{bent ray-path}}^{\mathbf{r}_{\text{exit}}}^{\mathbf{r}_{\text{miss}}} 1 \, d\ell; \quad (I.22)$$

and calculated as:

$$G_{\text{in}} = \ell_{\text{exit}} - \ell_0 = \ell_{\text{exit}}, \quad (I.23)$$

$$G_{\text{out}} = |\mathbf{r}_{\text{miss}} - \mathbf{r}_{\text{exit}}|. \quad (I.24)$$

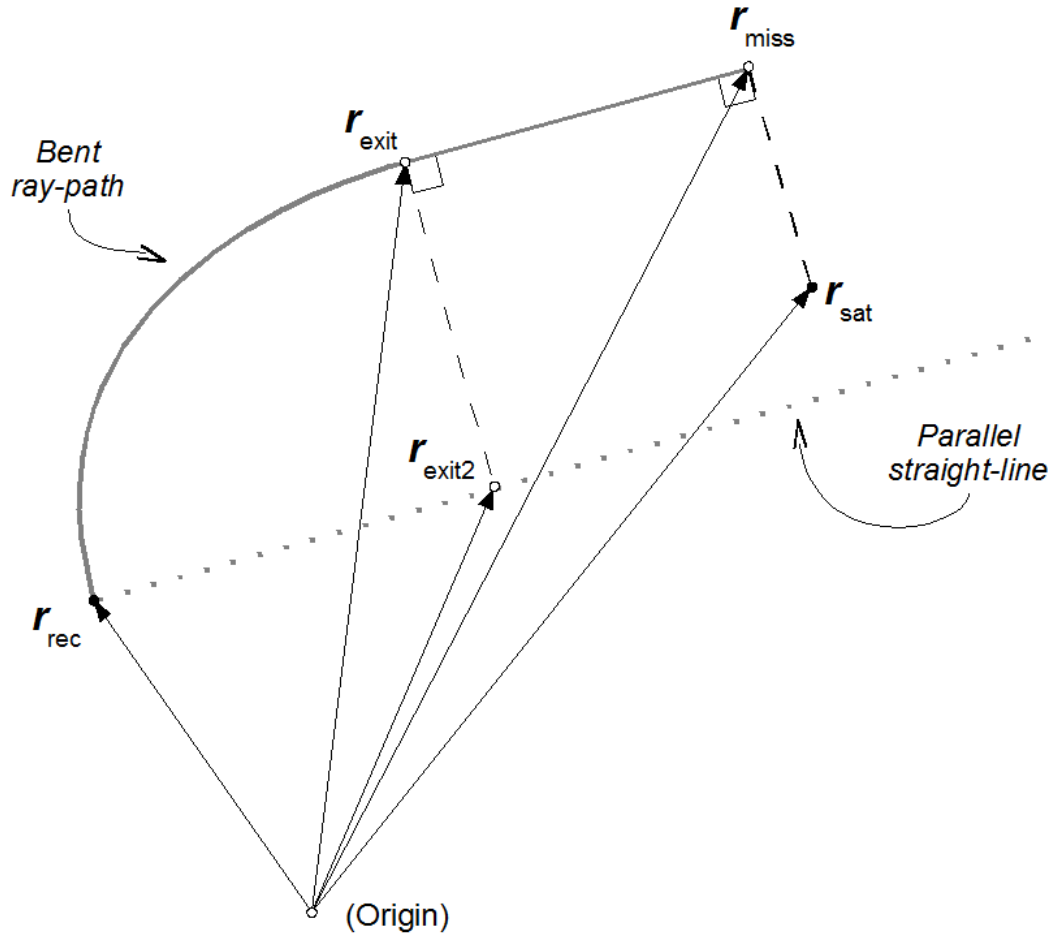


Figure I.9: Notable positions in a bent ray-path.

In contrast, D is generally more easily calculated in its entirety:

$$D = |\mathbf{r}_{miss} - \mathbf{r}_{rec}|. \quad (\text{I.25})$$

Under the assumption of a transmitting source at practically infinite distances, the geometric delay eq. (2.13) can be approximated as:

$$d_g \xrightarrow{D \rightarrow \infty} G_{in} - D_{in}, \quad (\text{I.26})$$

where $D_{\text{in}} = |\mathbf{r}_{\text{exit2}} - \mathbf{r}_{\text{rec}}| = (\mathbf{r}_{\text{exit}} - \mathbf{r}_{\text{rec}}) \cdot \hat{\mathbf{s}}_{\text{exit}}$, as discussed in Appendix I.3.2. Notice that the operands involved in the subtraction in eq. (I.26) have much smaller magnitude than the ones in the original eq. (2.13).

I.5.3 Bent ray-path

I.5.3.1 Initial value problem

Recalling, the initial value problem is that in which we are given an initial position \mathbf{r}_0 , corresponding to the receiver position, and an initial ray tangent direction $\hat{\mathbf{s}}_0$, hopefully corresponding to a satellite's apparent direction, and we are asked for the ray path until it exits the neutral atmosphere. The bent ray-path will be discretized in several straight-line segments; however crude such an approximation might seem, it can be made as accurate as needed simply by decreasing the length of the each segment.

Before we discuss the method that we have employed, it is insightful to quickly review a more basic method, which we call step-by-step. In the step-by-step method, one would obtain refractivity n_0 at the initial location; then, given an arbitrary step length $\Delta\ell$, one would obtain the next ray's position as $\mathbf{r}_1 = \mathbf{r}_0 + \Delta\ell\hat{\mathbf{s}}_0$, at which location the refractivity n_1 would be obtained. One would obtain the gradient of refractivity ∇n_{0-1} between locations \mathbf{r}_0 and \mathbf{r}_1 (more about that in Appendix I.5.3.4), which would be used, together with n , to calculate the (constant) curvature \mathbf{a}_{0-1} along the same interval, which, finally, would be used to update the direction vector, $\mathbf{s}_1 = \hat{\mathbf{s}}_0 + \Delta\ell\mathbf{a}_{0-1}$. The updated direction $\hat{\mathbf{s}}_1$, in its turn, would give the next position $\mathbf{r}_2 = \mathbf{r}_1 + \Delta\ell\hat{\mathbf{s}}_1$, and the cycle would be repeated, until the upper limit ℓ_{max} is reached or exceeded.

A more elaborate method would use information about the curvature \mathbf{a} itself to find the optimal step length $\Delta\ell$; the so-called Runge-Kutta methods are a prime example; see an example of use in ray-tracing in Thessin [2005].

The method employed in this work is based on a perturbation of the straight-line ray-path.⁵ Its rationale is that the straight-line ray-path is a model of sufficient accuracy for most of the sky; therefore it seemed too cautious to proceed step-by-step when in fact the ray-path did not deviate much from a straight-line. The second key aspect of the method is that we alternate between using the original, rigorous, atmospheric model and a surrogate, simplified, source, as given by a linear probe through the original atmospheric model. That is beneficial because the interpolation in the original atmospheric model can be very costly (computationally), a cost which we avoid using the linear surrogate; for example, if the original model follows a 3d atmospheric structure, then it requires 3d interpolations, more expensive than the linear interpolations required in the surrogate.

The procedure is as follows:

0. obtain the total delay d (along-path d_a + geometric d_g) according to a straight-line ray-path model. The use of a global adaptive quadrature routine will automatically sample more densely where the refractivity field is less smooth. Collect and save the (slant) profile of index of refraction n versus along-path distance ℓ , as a discretization in terms of pairs $\{\ell_i, n_i\}$; keep as well the straight-line ray-path itself, discretized as $\{\mathbf{r}_i\}$.
1. update the total delay, according to a bent ray-path model, using the previous results as a surrogate for the rigorous atmospheric model:

⁵We first came in contact with this idea in the article of Stam and Langu  nou [1996]; a detailed description can be found in the comprehensive book of Kravtsov [2005, section 2.9]. The form presented here is less general but hopefully more straightforward.

(a) keep the latest previously determined delay, ray-path, and index of refraction:

$$\begin{aligned} d^{\text{prev}} &\leftarrow d, \\ \{\mathbf{r}_i^{\text{prev}}\} &\leftarrow \{\mathbf{r}_i\}, \\ \{n_i^{\text{prev}}\} &\leftarrow \{n_i\}; \end{aligned}$$

(b) obtain the gradient of refractivity ∇n_i^{prev} at each of the previously determined positions $\mathbf{r}_i^{\text{prev}}$, as detailed in Appendix I.5.3.4;

(c) obtain a new ray-path $\{\mathbf{r}_i\}$, based on ∇n_i^{prev} , n_i^{prev} , and ℓ_i , besides the initial conditions \mathbf{r}_0 and $\hat{\mathbf{s}}_0$, employing eq. (I.17):

$$\mathbf{r}(\ell) = \mathbf{r}_0 + \ell \hat{\mathbf{s}}_0 + \int_0^\ell \int_0^\ell \mathbf{a}(\ell) d\ell d\ell; \quad (\text{I.17})$$

(notice that the previous ray-path $\mathbf{r}_i^{\text{prev}}$ is *not* used directly to obtain the new one, \mathbf{r}_i);

(d) obtain a new index of refraction n_i , using the previous results as a first-order probe into the rigorous atmospheric model:

$$n_i = n_i^{\text{prev}} + \nabla n_i^{\text{prev}} \cdot (\mathbf{r}_i - \mathbf{r}_i^{\text{prev}});$$

(e) obtain a new along-path d_a and geometric delays d_g , thus new total delay d , according to Appendices I.5.1 and I.5.2;

(f) compare the current total delay d with the previous determination d^{prev} : if $|d - d^{\text{prev}}| < \varepsilon$, for a user-prescribed tolerance ε (in meters), then stop;

2. update the total delay, according to a bent ray-path, using the rigorous atmospheric model:

- (a) keep the latest previously determined delay $d^{\text{prev}} \leftarrow d$ and ray-path $\{\mathbf{r}_i^{\text{prev}}\} \leftarrow \{\mathbf{r}_i\}$;
 - (b) obtain a new index of refraction n_i , interpolating in the original atmospheric model, at each $\mathbf{r}_i^{\text{prev}}$;
 - (c) obtain a new along-path delay d_a , thus new total delay d (the geometric delay d_g remains the same, because the ray-path $\{\mathbf{r}_i\}$ remains the same at this step);
 - (d) compare the current total delay d with the previous determination d^{prev} : if $|d - d^{\text{prev}}| < \varepsilon$, for a user-prescribed tolerance ε (in meters), then stop.
3. return to step 1.

I.5.3.2 Boundary value problem

As discussed in Appendix I.3.3, given initial conditions, the boundary conditions cannot be specified — they are a consequence of the atmospheric conditions along the ray-path. In other words, given an initial position \mathbf{r}_0 and an initial direction $\hat{\mathbf{s}}_0$, both the relative direction to the point of closest approach $\Delta\hat{\mathbf{r}}_{\text{miss}}$ and the tangent direction at the pierce point $\hat{\mathbf{s}}_{\text{exit}}$ cannot be specified. The way to honour such boundary conditions depends on whether we wish to obtain the delay in a single satellite direction or in multiple satellite directions, as reckoned from the same receiver location.

To honour the boundary conditions at a single direction, we employ the so-called *shooting method*:

1. solve the initial value problem (Appendix I.5.3.1);
2. calculate the discrepancy between actual and desired values for the boundary conditions;
3. if the discrepancy is smaller than tolerance, stop;

4. correct the initial condition;
5. return to step 1.

Without loss of generality, let us be more specific, considering only the elevation angle ε of the directions involved:

1. solve the initial value problem for an approximate initial or apparent elevation angle ε_0 ;
2. calculate the discrepancy between the elevation angle of the point of closest approach $\varepsilon_{\text{miss}}$ and the satellite's elevation angle ε_{sat} : $\delta\varepsilon = \varepsilon_{\text{miss}} - \varepsilon_{\text{sat}}$; (under the assumption of a transmitting source at practically infinite distances, we would calculate instead the discrepancy between the elevation angle corresponding to the exit tangent direction $\varepsilon_{\text{exit}}$ and the satellite's elevation angle ε_{sat} : $\delta\varepsilon = \varepsilon_{\text{exit}} - \varepsilon_{\text{sat}}$);
3. if the discrepancy is smaller than the tolerance ($\delta\varepsilon < \varepsilon_\varepsilon$), then stop;
4. correct the approximate initial elevation angle: $\varepsilon_0 \leftarrow \varepsilon_0 - \delta\varepsilon \partial\varepsilon_0 / \partial\varepsilon$, where $\partial\varepsilon_0 / \partial\varepsilon$ converts a change in the final elevation angle ε to a change in the initial elevation angle ε_0 ;
5. return to step 1.

In contrast, to honour the boundary conditions at multiple directions, we would solve the initial value problem for several different initial or apparent directions, collecting both the delay and the final direction resulting from each ray-trace. Then, the delay at any unsampled (final) direction would be obtained interpolating among the samples. Better yet, one would fit a model to the samples (e.g., Marini's [1972] for the variation with elevation angle and Davis et al.'s [1993] for the variation with azimuth — see section 2.6.3), and then obtain the delay at any unsampled direction evaluating the fitted model, followed by an interpolation in the model residual delay (rather than interpolating in the delay itself).

I.5.3.3 Tolerance for boundary conditions

We need not require that the boundary conditions are met exactly — it suffices that the discrepancy between the approximate and desired solutions be negligible. Therefore we set a *tolerance* for the boundary conditions. What would be the tolerance for the point of closest approach, or for the exit tangent direction? To find that tolerance, we convert the discrepancies in the boundary conditions to the corresponding discrepancies in the resulting delay. The latter is more meaningful than the former; e.g., 1 mm is a reasonable tolerance for the delay, given the precision of GPS measurements, but would 1 km be a reasonable tolerance for the point of closest approach?

The desired conversion is easiest when the boundary conditions are specified in terms of directions ($\Delta\hat{\mathbf{r}}$ or $\hat{\mathbf{s}}$) instead of positions (\mathbf{r}), and these directions are expressed in local polar coordinates (elevation angle ε and azimuth α) instead of global Cartesian coordinates (X, Y, Z) or local Cartesian coordinates (north x , east y , up z). Without loss of generality, let us consider only the elevation angle ε . Given the discrepancy between the exit elevation angle $\varepsilon_{\text{exit}}$ and the satellite's elevation angle ε_{sat} :⁶

$$\delta\varepsilon = \varepsilon_{\text{exit}} - \varepsilon_{\text{sat}}, \quad (\text{I.27})$$

the corresponding discrepancy in delay is obtained as:

$$\delta d = \frac{\partial d}{\partial \varepsilon} \delta\varepsilon. \quad (\text{I.28})$$

The question, then, is how to obtain the rate of change $\partial d / \partial \varepsilon$.

⁶The elevation angles $\varepsilon_{\text{exit}}$ and ε_{sat} are to be reckoned from the same ellipsoidal normal passing through \mathbf{r}_{rec} . In other words, even though the directions $\hat{\mathbf{s}}_{\text{exit}}$ and $\Delta\hat{\mathbf{r}}_{\text{sat}}$ are shown in Figures I.7 and I.5 (on p. 170 and p. 168) at different positions (\mathbf{r}_{exit} and \mathbf{r}_{sat} , respectively), in actuality such direction vectors have no base position defined. Upon conversion from global to local coordinates, both direction vectors must be converted to the same local coordinate system, defined by the same ellipsoidal normal, if their discrepancy is to have any meaning.

An approximate rate of change can be obtained from two samples of delay at slightly different elevation angles:

$$\frac{\partial d}{\partial \varepsilon} \approx \frac{d_1 - d_2}{\varepsilon_1 - \varepsilon_2}. \quad (\text{I.29})$$

But that would require at least two ray-traces before we are able to decide if the approximation is good enough. We may employ, instead, a nominal rate of change:

$$\frac{\partial d}{\partial \varepsilon} \approx -d^z \cos \varepsilon / \sin^2 \varepsilon, \quad (\text{I.30})$$

where the units of ε are radians (otherwise, $\frac{\partial d}{\partial \varepsilon} \leftarrow \frac{\partial d}{\partial \varepsilon} \frac{\pi}{180^\circ}$); eq. (I.30) follows from a nominal slant delay, $d = d^z / \sin \varepsilon$. However crude the nominal rate above might seem, we found it of sufficient accuracy because (i) it tends to super-estimate the converted error δd , which is less harmful than sub-estimating it, and (ii) the raw discrepancy $\delta \varepsilon$ is always calculated rigorously, regardless of the rate of change $\partial d / \partial \varepsilon$ (nominal or actual) employed. If we end up having to iterate more than two times, we may replace the nominal rate of change eq. (I.30) with the actual rate of change eq. (I.29).

1.5.3.4 Gradients and partial and directional derivatives

Here we discuss how to obtain the gradient of refractivity ∇n .

We can always obtain ∇n numerically, i.e., numerically evaluating the partial derivatives involved in its definition:

$$\nabla n = \frac{\partial n}{\partial X} \hat{\mathbf{i}} + \frac{\partial n}{\partial Y} \hat{\mathbf{j}} + \frac{\partial n}{\partial Z} \hat{\mathbf{k}}, \quad (\text{I.31})$$

e.g.:

$$\frac{\partial n(\mathbf{r})}{\partial X} \approx \frac{n(\mathbf{r} + \delta \hat{\mathbf{i}}) - n(\mathbf{r} - \delta \hat{\mathbf{i}})}{2\delta}. \quad (\text{I.32})$$

That is very costly, because the numerical derivative along each direction requires two in-

terpolations, in a total of six interpolations for each point at which we require the gradient ∇n .

The general case above is valid in any atmospheric structure. It is the only solution available in a 3d atmospheric structure, for which we make no assumption about the gradient ∇n . But we can do better in other atmospheric structures, especially when we have prior knowledge about the direction $\hat{\nabla} n$ of the gradient ∇n . Given the assumed gradient direction $\hat{\nabla} n$ we can obtain the gradient magnitude $|\nabla n|$ from the directional derivative $dn/d\ell$ as:

$$\nabla n \cdot \hat{s} = dn/d\ell, \quad (\text{I.33})$$

$$|\nabla n| \hat{\nabla} n \cdot \hat{s} = dn/d\ell, \quad (\text{I.34})$$

$$|\nabla n| = \frac{dn/d\ell}{\hat{\nabla} n \cdot \hat{s}}. \quad (\text{I.35})$$

In fact, when the direction of the gradient of refractivity is known beforehand (i.e., for atmospheric structures other than 3d), a bent-3d ray-path model can be processed with no greater computational cost (in terms of number of interpolations) than a bent-2d ray-path model.

In a spherical atmosphere, the gradient of refractivity points to the center of the (osculating or concentric) sphere:⁷

$$\hat{\nabla} n = -\hat{r}'. \quad (\text{3.2})$$

In an ellipsoidal atmosphere, the direction of the gradient of refractivity is also known:

$$\hat{\nabla} n = -\hat{k}, \quad (\text{I.36})$$

⁷We implemented the bent-2d ray-path model fixing the direction of the gradient of refractivity according to eq. (3.2) and integrating the 3d Eikonal eq. (3.1); that is equivalent to evaluating the 2d differential eq. (3.4) (closely related to Bouguer's formula eq. (3.3) and to Snell's law eq. (3.5)); in fact, eq. (3.4) is a simplification of eq. (3.1) exactly when eq. (3.2) holds, as we demonstrated in Appendix I.3.4.1.

where $\hat{\mathbf{k}}$ is the local up direction, along the ellipsoidal normal.

The directional derivative $dn/d\ell$ can be obtained numerically from the refractivity samples already needed to compute the definite integral involved in the definition of the along-path delay (Appendix I.5.1). I.e., given pairs $\{\ell_i, n_i\}$, we can obtain $dn/d\ell$ as:

$$\frac{dn}{d\ell_i} \approx \frac{n_{i+1} - n_{i-1}}{\ell_{i+1} - \ell_{i-1}}. \quad (\text{I.37})$$

With that we reduce the number of numerical derivatives from six to only one and we need not to interpolate refractivity values at points other than $\{\mathbf{r}_i\}$.

As for the gradient atmosphere, we do have some prior knowledge about it (section 3.2.4). That can be shown expressing the gradient of refractivity in terms of local Cartesian coordinates (x, y, z) :

$$\nabla n = \nabla^{XYZ} n = \mathbf{J} \nabla^{xyz} n, \quad (\text{I.38})$$

(where \mathbf{J} is a suitable rotation matrix — see Appendix V), which, in its turn, is expressed in terms of the directional derivatives in the local Cartesian coordinates:

$$\nabla^{xyz} n = \frac{\partial n}{\partial x} \hat{\mathbf{i}} + \frac{\partial n}{\partial y} \hat{\mathbf{j}} + \frac{\partial n}{\partial z} \hat{\mathbf{k}} \quad (\text{I.39})$$

As per definition of gradient atmosphere, the derivatives in local coordinates are function of height only. So, in principle, such a prior knowledge about the gradient of refractivity is an advantage over the general case explained at the beginning of the present section. In practice, though, the fact that we have defined the gradient atmosphere in terms of atmospheric parameters (pressure, temperature, humidity) instead of refractivity itself, makes the advantage above less attractive, in terms of reduction in the number of interpolations.

Appendix II

Geopotential height

In the interest of keeping the discussion about atmospheric structures (section 3.2) to the point, we have intentionally skipped a necessary topic, which we now cover. In atmospheric models, the vertical coordinate is *not* a geometric quantity such as radius or height, identified in section 3.2. For atmospheric scientists, it seems more natural to take pressure as the independent vertical coordinate over which all other atmospheric parameters vary. It is so much so that, e.g., typically a NWM is provided in layers of constant pressure (the so-called isobars), and in the assimilation of radiosonde observations, pressure is treated as a vertical coordinate, not as an observation itself. Fortunately there is a parameter bridging the gap between pressure and geometric heights: the so-called geopotential. Geopotential or, more specifically, its derived geopotential height, is a very convenient bridge, because it can be calculated independently based upon only atmospheric or only geodetic quantities.

To define geopotential height we start defining *increments in geopotential height*, denoted dZ :

$$dZ \equiv g \, dz / g_c, \tag{II.1}$$

where dz is an increment in geometric height (e.g., as would be measured by a ruler), g

is the actual gravity (function of position, mainly height), and g_c is a nominal, constant, value of gravity. If we integrate the increments in geopotential height dZ we obtain *relative geopotential height*, denoted ΔZ :

$$\Delta Z \equiv \int_{Z_{\text{ref}}}^Z dZ. \quad (\text{II.2})$$

To obtain *absolute geopotential heights*, denoted Z , we need to specify as well a *datum*:

$$Z \equiv Z_{\text{ref}} + \int_{h_{\text{ref}}}^h dZ = Z_{\text{ref}} + \Delta Z. \quad (\text{II.3})$$

The datum comprises two parts: (i) the location of the reference surface from which relative heights ΔZ are reckoned (expressed in a precise, unambiguous way, such as the ellipsoidal height of that surface, h_{ref}), and (ii) the postulated value of absolute geopotential height at that reference surface, Z_{ref} (which can be quite arbitrary and need not be necessarily physically consistent with the definition of increments in geopotential height dZ).

What makes geopotential height so convenient is that it can be obtained independently with only atmospheric quantities or only geodetic quantities. That is extremely convenient in the practice of ray-tracing, in which we are given an atmospheric model including geopotential heights, and we wish to obtain the atmospheric parameters at an arbitrary position inside the atmospheric model, given a position in the real world. Consequently, as part of the delay solver (Appendix I.2), there is a hidden step (let us call it 1c in the diagram of steps in the delay solver, Figure I.2, p. 162), in which we perform one further coordinate conversion, before interpolating the atmospheric parameters.

What is the order of magnitude of possible errors in the relative heights ΔZ and in the datum $Z_{\text{ref}}, h_{\text{ref}}$? Figure II.1 shows a typical profile of (hydrostatic) refractivity versus geopotential height, as well as profiles representing the two types of errors (greatly exag-

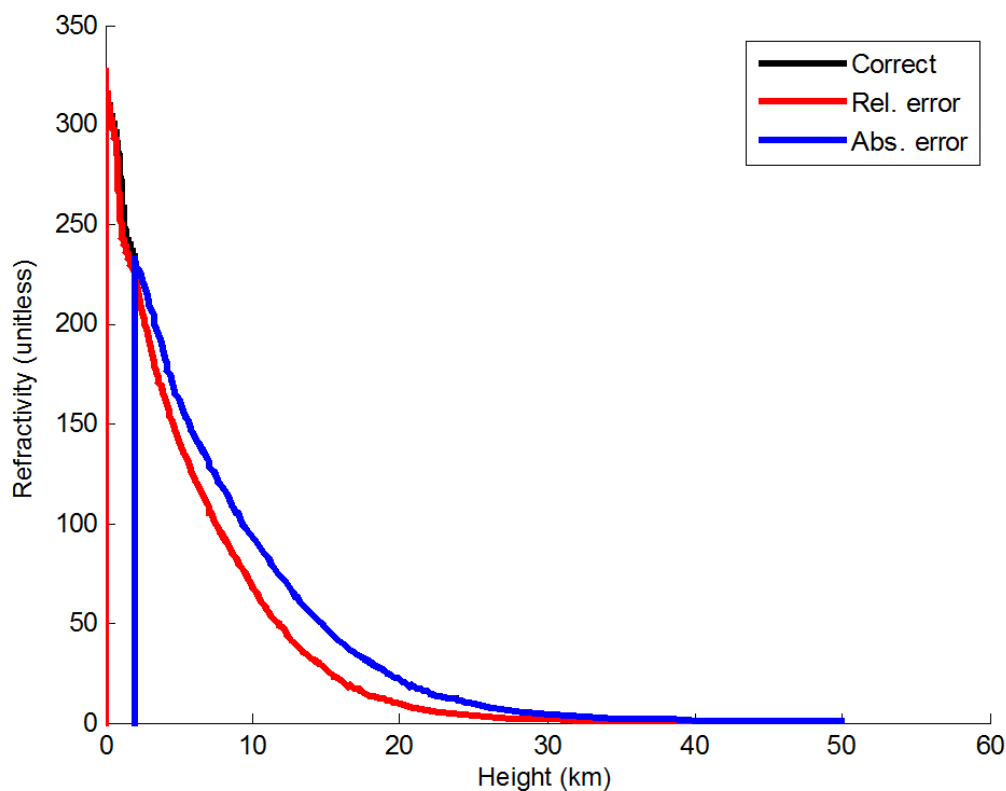


Figure II.1: Distortion in a refractivity profile due to errors in the datum and in relative geopotential heights.

gerated), as compared to the true profile. An error in relative height ΔZ implies that the true profile is going to be stretched; an error in the datum $Z_{\text{ref}}, h_{\text{ref}}$ implies that the true profile is going to be shifted.

What is the impact of those two types of errors in the delay? In other words, how much area is lost or gained under the curve of refractivity versus along-path distance due to the stretching and shift in the refractivity profile illustrated in Figure II.1? Without loss of generality, let us consider the impact in zenith hydrostatic delay. Errors in relative height ΔZ have little impact (sub-mm), because the product of refractivity and relative height error is always small. More specifically, Figure II.2 shows that near the ground (on the left side of the figure), refractivity is largest but the relative height error is small; aloft

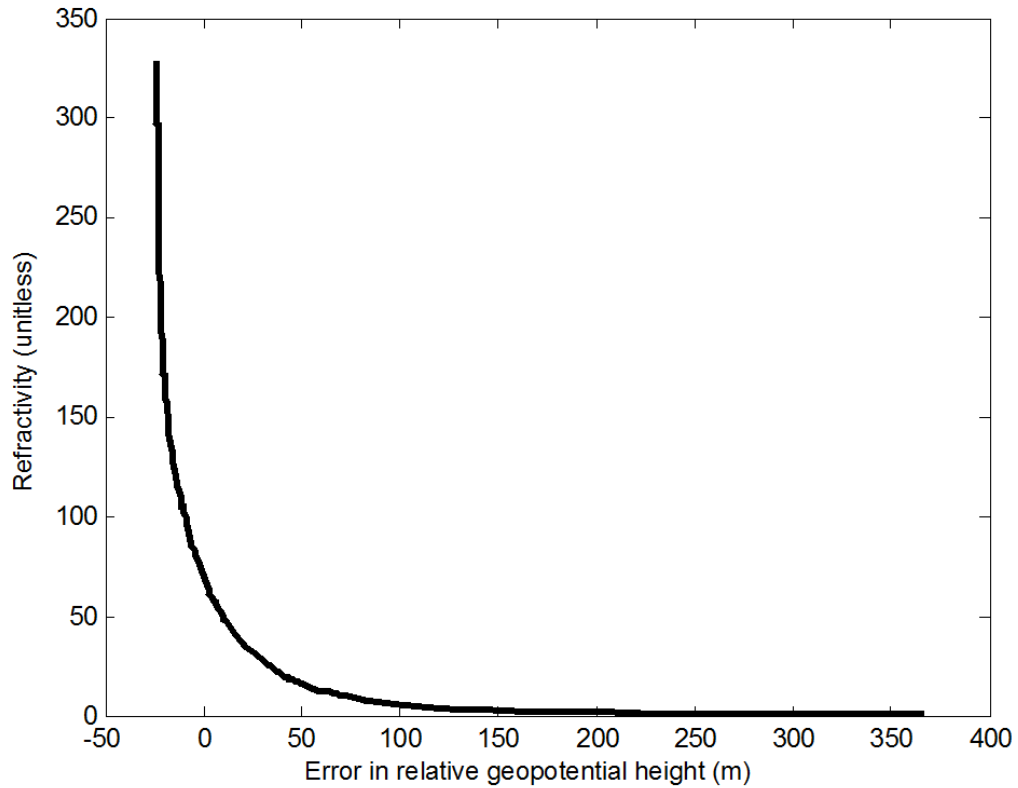


Figure II.2: Refractivity versus error in relative geopotential height in a vertical profile.

(on the right side of the figure), relative height errors are larger (the exact value shown in the figure is hypothetical), but refractivity is already very small anyway, following an exponential decrease. Errors in the datum $Z_{\text{ref}}, h_{\text{ref}}$, in contrast, have the potential to cause much greater damage, roughly 3.5 mm in zenith delay per 10m of error in the reference height, for a receiver near the ground. Figure II.3 shows a nominal rate of change in the zenith hydrostatic delay with respect to height, $\partial d^z / \partial Z$, along a profile; near the ground the zenith delay varies much more rapidly than aloft. An error in the datum corresponds to misplacing the receiver inside the atmospheric model, putting it a little higher or lower than it should be. Such a positioning error will lead to an error in the interpolated atmospheric parameters, especially pressure, because it decays exponentially with height.

In Appendices II.1 and II.2 we discuss relative geopotential heights from the perspec-

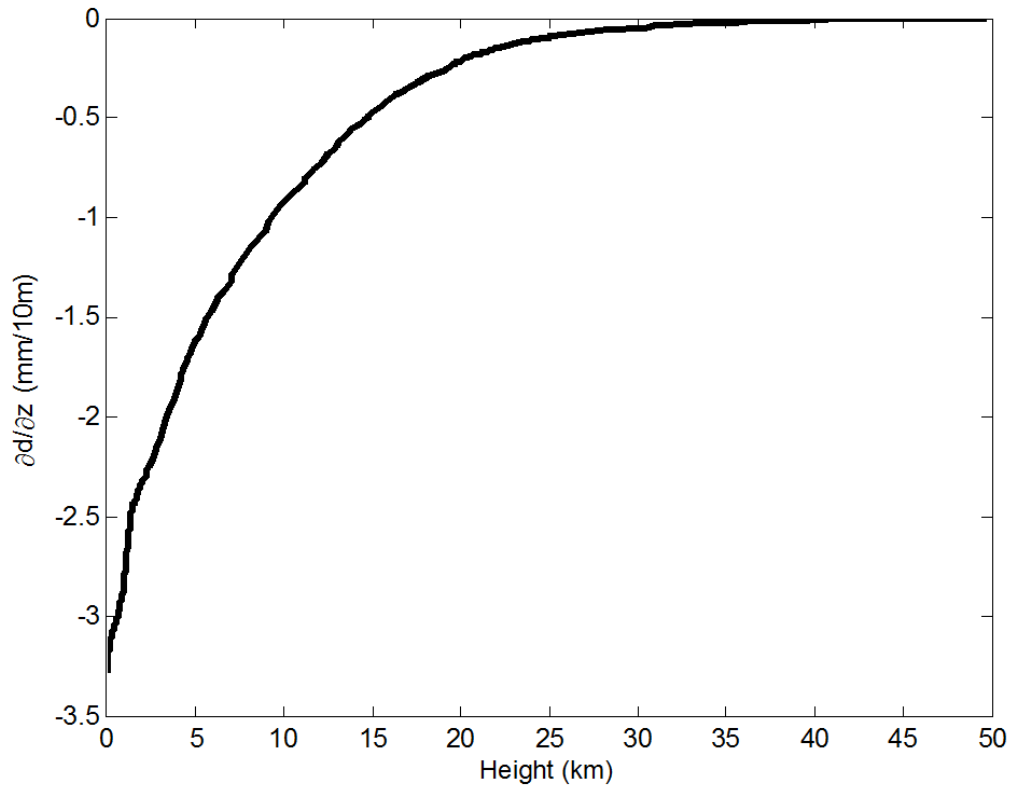


Figure II.3: Nominal rate of change of zenith hydrostatic delay with respect to height, along a vertical profile.

tive of atmospheric sciences and geodesy, respectively. Then, in Appendix II.3, we discuss the datum $Z_{\text{ref}}, h_{\text{ref}}$ necessary to define absolute geopotential heights Z . In Appendix II.3 we also give practical details about how the conversion to geopotential heights was performed in this work.^{1,2}

¹Please notice that in this work we have no need for a conversion *from* geopotential heights; following the diagram of steps in the delay solver (Figure I.2, p. 162), we obtain the geopotential height Z only to obtain the atmospheric parameters P, T, q only to obtain the refractivity N , which is finally kept — the remaining, intermediary, quantities are usually discarded.

²A few words about previous related work. COESA [1962, section I.2.3–I.2.4] contains an excellent, self-contained discussion about geopotential heights (which is much too shortened in the newer COESA [1976]). List [1951, section IV] describes a particularly successful geopotential model, due to Lambert [1949], that allows one to retain a simple inverse-square relation to account for both the height-decreasing gravitation and the height-increasing centrifugal potential, through the introduction of a fictitious radius. Ge [2006, Appendix A] compares several simplified formulations for the geopotential. Vedel [2000] was, to the best of our knowledge, the first to recognize the need for a datum to define unambiguous absolute geopotential heights, which we seek to clarify and generalize in here.

II.1 Increments in geopotential height in atmospheric sciences

In atmospheric sciences, the equation:

$$g dz = -\frac{dP}{\rho} \quad (\text{II.4})$$

models the so-called *hydrostatic equilibrium*: a parcel of air is held static by the balance of its downward weight and the upward pressure of the underlying atmosphere (neglecting the presence of other vertical forces acting upon the parcel, as might happen during, e.g., a hurricane); dP is an increment in pressure, ρ is density of the mixed, humid air, $dz \neq dZ$ is an increment in geometric height, and g is actual gravity.

We may expand the density ρ based on the *gas state equation* for mixed humid air:

$$P = \rho R_d T_v, \quad (\text{II.5})$$

where R_d is the gas constant for dry gases, and T_v is virtual temperature.³ Replacing the density ρ (eq. (II.5)) in the hydrostatic equilibrium equation (eq. (II.4)) we obtain:

$$g dz = -\frac{dP}{P/(R_d T_v)}; \quad (\text{II.6})$$

now making use of the well-known definition of natural logarithm from calculus, $d \ln P \equiv dP/P$, we obtain:

$$g dz = -d \ln(P) R_d T_v. \quad (\text{II.7})$$

³“... rather than use a gas constant for moist air, the exact value of which would depend on the amount of water vapor in the air (which varies considerably), it is convenient to retain the gas constant for dry air and use a fictitious temperature (called virtual temperature) in the ideal gas equation.” [Wallace and Hobbs, 2006, p. 66].

The definition of geopotential height is particularly useful for obtaining the pressure P above or below a given point where pressure P_0 is known, because it allows us to simplify (under certain assumptions) an integral to a closed-form expression. Integrating an increment of natural logarithm of pressure $d \ln P$:

$$\int_{\ln P_0}^{\ln P} d \ln P = \ln P - \ln P_0 = \ln \frac{P}{P_0} \quad (\text{II.8})$$

and also:

$$\int_{\ln P_0}^{\ln P} d \ln P = \int_{z_0}^z -g / (R_d T_v) dz, \quad (\text{II.9})$$

Combining the two equations above we obtain:

$$P = P_0 \exp \left(- \int_{z_0}^z g / (R_d T_v) dz \right), \quad (\text{II.10})$$

where the actual gravity g varies with geometric height z . From the definition of an increment in geopotential height (eq. (II.1)), we can replace dz by $g_c dZ/g$:

$$P = P_0 \exp \left(-g_c \int_{Z_{\text{ref}}}^Z (R_d T_v)^{-1} dZ \right). \quad (\text{II.11})$$

where the constant g_c can be taken out of the integral, a fact from which derives all the convenience in the definition of increments in geopotential heights (eq. (II.1)).

The last equation can be simplified to a closed-form expression based upon assumptions for the variation of virtual temperature (which itself is function of actual temperature and humidity) over geopotential height (not over geometric height), $\beta \equiv dT_v/dZ$, the so-called (*virtual temperature*) *lapse-rate*; if we assume $T_v = \text{const.}$, then:

$$P = P_0 \exp \left(-g_c (Z - Z_{\text{ref}}) / (R_d T_v) \right); \quad (\text{II.12})$$

if $\beta = \text{const.} \neq 0$, then:

$$P = P_0 \left(\frac{T_{v,0} + \beta(Z - Z_{\text{ref}})}{T_{v,0}} \right) \wedge \left(-\frac{g_c}{R_d \beta} \right), \quad (\text{II.13})$$

where the symbol \wedge denotes the power operation.

Such simplifications would not be possible had we not got rid of the height-dependent actual gravity g . The best that we could achieve without the introduction of increments in geopotential height dZ , based upon assumptions for the variation of virtual temperature over geometric height (not geopotential height), $\beta' \equiv dT_v/dz \neq \beta = dT_v/dZ$, would be the following: if we assume $T_v = \text{const.}$, then:

$$P = P_0 \exp(-g_m(z - z_{\text{ref}})/(R_d T_v)); \quad (\text{II.14})$$

if $\beta' = \text{const.} \neq 0$, then:

$$P = P_0 \left(\frac{T_{v,0} + \beta'(z - z_{\text{ref}})}{T_{v,0}} \right) \wedge \left(-\frac{g_m}{R_d \beta'} \right); \quad (\text{II.15})$$

where the integral persists in the newly-defined g_m :

$$g_m \equiv \frac{\int g d \ln T_v}{\int d \ln T_v} = \frac{\int g/T_v dT_v}{\int 1/T_v dT_v}, \quad (\text{II.16})$$

which can be interpreted as a “mean gravity weighted by the reciprocal of virtual temperature”. However similar eqs. (II.14) and (II.15) are to eqs. (II.12) and (II.13), the ones employing the constant g_c remain more convenient than ones employing the integrated g_m .

II.2 Increments in geopotential height in geodesy

Let us start by defining *geopotential*, denoted W : it is the Earth's gravity potential at a particular point on or above the Earth's surface. Gravity includes both gravitation, whose potential is denoted V , due to mass attraction, pointing inward,⁴ as well as the centrifugal force, whose potential is denoted Φ , due to the Earth's rotation, pointing away from the rotation axis.^{5,6} It has units of (specific) work (m^2/s^2). *Geopotential difference*, known as *geopotential number* in geodesy, is simply the difference in geopotential between two positions:

$$\Delta W \equiv \int_{\mathbf{r}_1}^{\mathbf{r}_2} dW = W(\mathbf{r}_2) - W(\mathbf{r}_1) = W_2 - W_1, \quad (\text{II.17})$$

where the integral is path-independent.

The link between increments in geopotential height and increments in geopotential comes from the very definition of the gravity vector \mathbf{g} as the gradient of the gravity potential:

$$\mathbf{g} \equiv \nabla W. \quad (\text{II.18})$$

which we can express in scalar terms as:

$$g \, dz = dW. \quad (\text{II.19})$$

where g is the magnitude of the vector \mathbf{g} , and the increments in geometric height dz are aligned with the plumb-line, which is the curve whose tangent is $\hat{\mathbf{g}}$. Consequently, incre-

⁴One might pose the question of whether we should consider the mass of the atmosphere itself, in addition to the mass of the solid Earth and oceans, when calculating gravitation; we ignore that because the resulting discrepancy is negligible for our purposes here.

⁵The addition of the centrifugal potential for calculations of gravity affecting a parcel of air is due to the observation that the atmosphere is co-rotating with the solid Earth. If that was not true, then we would be experiencing a strong, constant, east–west wind, strongest along the equator and null at the poles. (Which is not to say that the solid Earth is continuously forcing the atmosphere; they happen to be co-rotating because of the way that the atmosphere was formed.)

⁶Here we neglect tidal variations in gravity.

ments in geometric height dz corresponds to increments in the quantity known as *orthometric height* in geodesy, denoted H , and thence also to increments in ellipsoidal heights h , because of their well-known relationship, $h = H + N$, where N is geoidal undulation or separation.^{7,8,9} Notice that the correspondence is valid only for increments ($dz = dH = dh$) and relative heights ($\Delta z = \Delta H = \Delta h$), *not* absolute heights (z, H, h); in other words, an unqualified “[absolute] geometric height” z is ambiguous in geodesy, because it could refer to either or none of orthometric heights H or ellipsoidal heights h .

Relative geopotential heights ΔZ can be expressed simply as the geopotential difference ΔW scaled by a nominal, constant, value of gravity, so as to have a result in units of length (m):

$$\Delta Z = \Delta W / g_c. \quad (\text{II.20})$$

In geodesy such a quantity is called *dynamic height*. In spite of its units, geopotential height represents only approximately a geometric length, as would be measured by a ruler. It is common to define new units, the so-called “geopotential metre”, to emphasize that distinction.¹⁰

In geodesy, models for the gravity potential W , at various accuracy levels, are commonplace. We review a few in the next section.

⁷The geoid is the geopotential surface coinciding with mean-sea-level, roughly speaking.

⁸Geoidal undulation N which is calculated as described in Lemoine et al. [1998, section 11] or, more easily, interpolated in pre-calculated grids available at <http://earth-info.nga.mil/GandG/wgs84/gravitymod/egm96/binary/binarygeoid.html>.

⁹Strictly speaking, the relationship between h and H through N is only approximately — not exactly — equal because, whereas h increases along the ellipsoidal normal, H increases along the plumb-line; the ellipsoidal normal remains a satisfactory model for the actual plumb-line in current geodetic practice, though.

¹⁰We would argue that such practice should be deprecated, in favour of encouraging the proper qualification of ΔZ as scaled geopotential — still more of a gravity potential than a length.

II.3 Conversion to geopotential height

In the present sub-section we are interested in obtaining the absolute geopotential height Z , as employed in an atmospheric model, corresponding to a position in the real-world, expressed in either global Cartesian coordinates X, Y, Z or geodetic coordinates φ, λ, h . In the introduction above, we saw that absolute geopotential heights Z have two parts: a datum $Z_{\text{ref}}, h_{\text{ref}}$ and a relative height ΔZ above or below a reference surface (eq. (II.3)):

$$Z \equiv Z_{\text{ref}} + \int_{h_{\text{ref}}}^h dZ. \quad (\text{II.3})$$

Before we tackle each part, let us state upfront the conclusions from the experiments described below: (i) with regard to the datum, relative heights ΔZ are to be reckoned from the geoid (i.e., $h_{\text{ref}} = N$, where N is the geoidal undulation) and absolute geopotential height is to be zero on the geoid (i.e., $Z_{\text{ref}} = 0$); (ii) with regard to the relative height ΔZ , the discrepancy among different geopotential models is quite negligible for our purposes here.

Starting with the calculation of relative geopotential heights ΔZ , from Appendix II.2 we learned that it may be expressed in terms of geopotential W (eq. (II.20)). The nominal, constant, gravity value is $g_c \equiv 9.80665 \text{ m/s}^2$, “a value decided upon by the WMO [World Meteorological Organization] and used by all meteorological offices [world-wide]” [Vedel, 2000, p. 3]. So the problem now becomes how to calculate the geopotential W . We have compared the following geopotential models:

- normal gravity, in three implementations: closed formula in ellipsoidal coordinates (which are different from geodetic coordinates); infinite series in spherical coordinates; and first two terms of the series in spherical coordinates [Torge, 2001];
- EGM96 expansion of the actual gravitation potential V in spherical harmonics, up

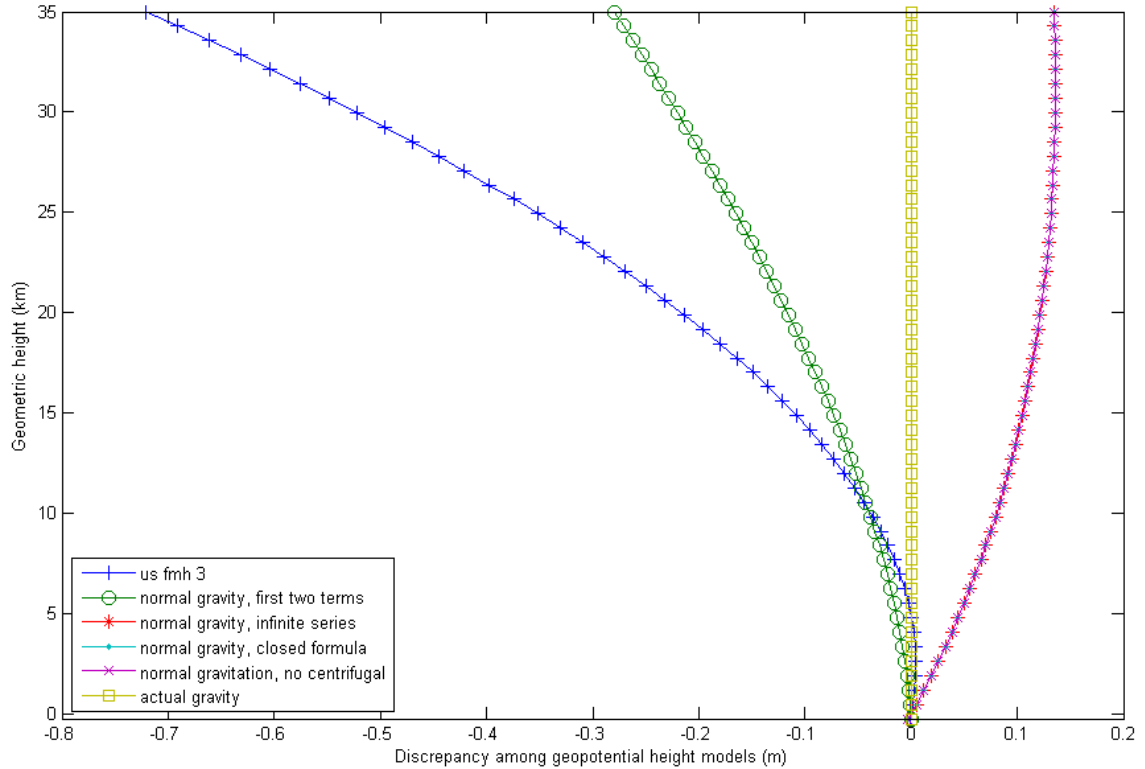


Figure II.4: Discrepancy in geopotential height due to different gravity formulations.

to degree and order 360 [Lemoine et al., 1998], plus normal centrifugal potential Φ (model which we denote “actual gravity” in Figure II.5);

- formula as given by OFCM [2007, Appendix D] (which gives directly geopotential height, not the geopotential itself).

Whereas the discrepancy among those different gravity models in terms of geopotential height reaches almost a meter at 35km of height (Figure II.4), the corresponding discrepancy in zenith hydrostatic delay (Figure II.5) reaches only μm -level (10^{-6} m) — *quite* negligible.

Regarding the datum $Z_{\text{ref}}, h_{\text{ref}}$, it is important to stress that here we are *not* trying to find an optimal datum, only trying to adopt a datum consistent with the one employed for the generation of the atmospheric model we were using. In that regard, matters are a little

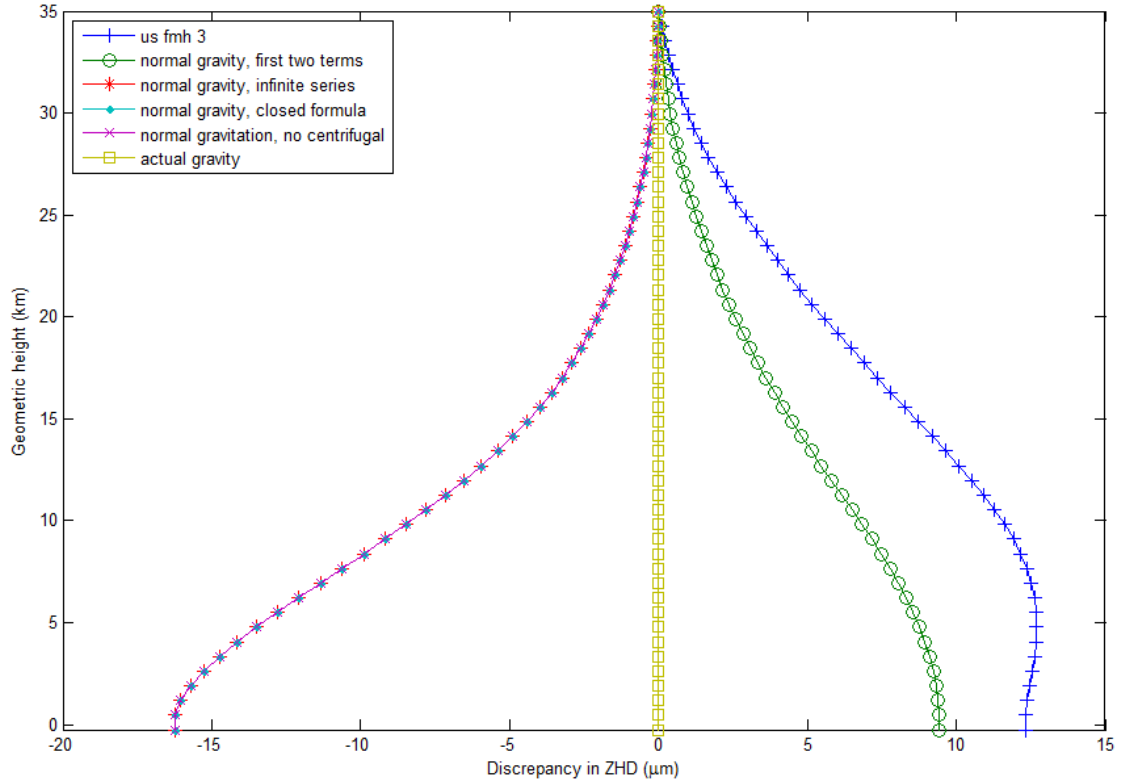


Figure II.5: Discrepancy in zenith hydrostatic delay due to different gravity formulations.

complicated, because it seems that each atmospheric model employs an ad hoc, arbitrary, reference. For example, Vedel [2000] reports that the Danish NWM adopted as reference surface the ground, as represented by a digital elevation model (DEM), presumably expressed in orthometric heights, thus $h_{\text{ref}} = H_{\text{ground}} + N$, and that the postulated reference absolute geopotential height is calculated from the DEM as $Z_{\text{ref}} = H_{\text{ground}} g'_c / g_c$, where g_c is as before, and $g'_c = 9.81 \text{ m/s}^2$ (as an approximation to the actual gravity near the ground).

We compared the following options for the reference $Z_{\text{ref}}, h_{\text{ref}}$:

1. the ellipsoid as reference surface, $h_{\text{ref}} = 0$, and zero absolute geopotential height at such a reference surface, $Z_{\text{ref}} = 0$;
2. the geoid as reference surface, $h_{\text{ref}} = N$, and absolute geopotential height equal to the orthometric height at such a surface, $Z_{\text{ref}} = H_{\text{geoid}}$, where H_{geoid} is zero by definition

of orthometric height, thus $Z_{\text{ref}} = 0$.

The comparison was done as follows: we obtained the ellipsoidal height as well as measured pressure (at a single, arbitrary epoch) in each of the IGS stations inside the horizontal extents of the NWM that we were using (17 stations in total). Then we (i) interpolated the atmospheric model's geopotential height corresponding to the measured pressure (in ray-tracing we do the inverse interpolation) and independently (ii) converted the ellipsoidal height to geopotential height according to each of the two options listed above. Finally, we checked the discrepancy between (i) and (ii). Table II.1 shows the root-mean-square error of the discrepancies in geopotential height. Notice that the option in which the ellipsoid is taken as reference surface yields much worse RMS than the option in which the geoid is taken instead. Therefore we adopted the latter option.

Table II.1: RMS of discrepancy in geopotential height due to different references.

Option 1	Option 2
35.0cm	12.5cm

As one last remark, notice that the conversion to geopotential heights is a function not only of ellipsoidal height h , but also of latitude φ and, to a minor extent, longitude λ , depending on the adopted gravity model. Therefore the conversion to geopotential heights has the potential to create some spurious azimuthal asymmetry in the delay, undesired in all atmospheric structures but the 3d one. To understand the source of such an azimuthal asymmetry, it might be easier to think in terms of the simpler straight-line ray-path: unless some care is exercised, points at the same along-path distance ℓ and same elevation angle ε but different azimuth α (thus geometrically symmetrical around the base point) will be attributed different geopotential heights, which in its turn will lead to different values for the interpolated atmospheric parameters. To avoid that, we adopted the following procedure: given the position of the point of interest in the real world space, we take

only its vertical coordinate (h, r, r') and append it to the base point's (i.e., the receiver's) horizontal geodetic coordinates φ_0, λ_0 , to only then input those three coordinate into the routine of conversion to geopotential heights.

Appendix III

Pressure interpolation

In developing the ray-tracer employed in this work, one of the validations that we performed was comparing the zenith hydrostatic delay, as integrated numerically, with that given by Saastamoinen's [1972] model, which requires as atmospheric parameter input only the surface (station) pressure. In early developments we found a very systematic discrepancy (meaning a discrepancy with very small standard deviation), amounting to 1 cm, which we eventually tracked down to an inappropriate interpolation algorithm for pressure. The present chapter is a quick note documenting that issue and its solution.

In Appendix II.1 we discussed how pressure decays exponentially with height, a decay which is modulated by the virtual temperature lapse rate β . Therefore, given a profile of pressure versus height discretized at several nodes, if one is to apply a linear interpolation algorithm, it is more correct to do so in the logarithm of pressure rather than directly in pressure itself (see below). The exact logarithm base (e.g., base 10, natural logarithms, etc.) is not as important. Of course, if the original profile is already provided at a sufficiently dense discretization (i.e., at sufficiently small node spacing), then the interpolator does not matter much.

The consequence of a sub-optimal linear interpolator, as compared to a better log-linear

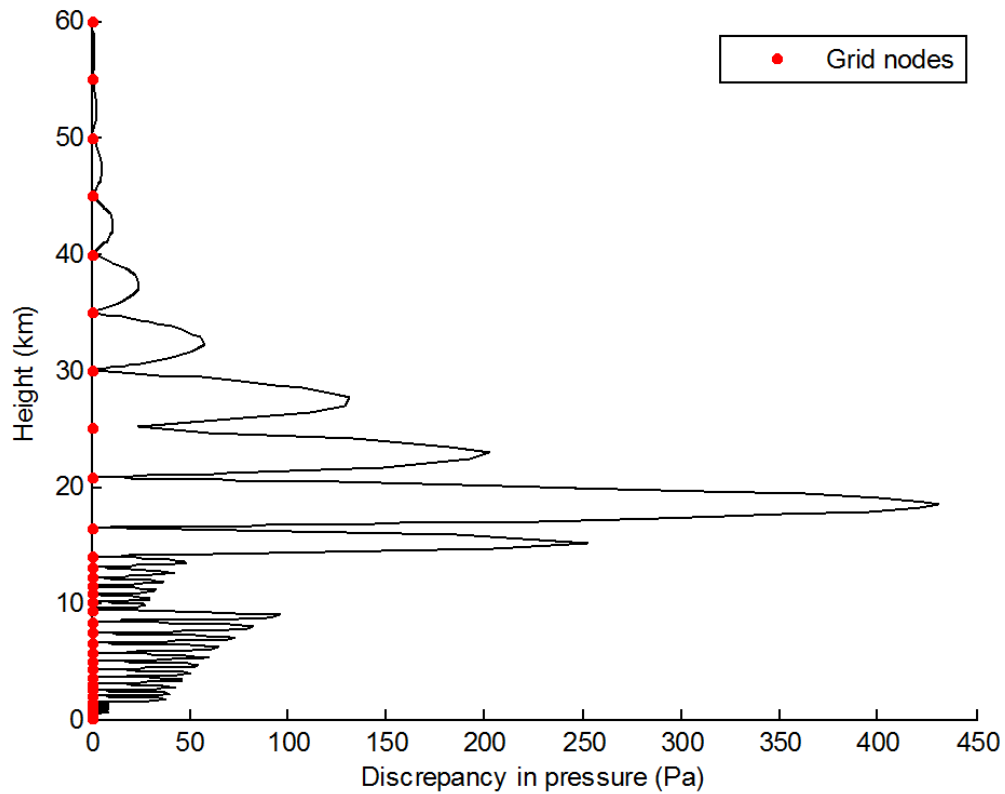


Figure III.1: Consequence of linear versus log-linear interpolation of pressure.

interpolator, is illustrated in Figures III.1 and III.2. Notice how the linear interpolator systematically over-estimates pressure at all heights but those at which a node exists; that over-estimation, in its turn, creates an excess area under the curve of refractivity versus along-path distance, amounting to the 1 cm bias between numerically integrated and Saastamoinen's zenith hydrostatic delay, that we reported before. Applying the more correct log-linear interpolator, the above-mentioned bias is reduced by one order of magnitude.

In fact, such a bias can be decreased even more if pressure is obtained integrating it hydrostatically from the nearest node discretizing the original profile (through the evaluation of eq. (II.11) or its special cases, eqs. (II.12) and (II.13)). That bias reduction is a consequence of the fact that the assumption of hydrostatic equilibrium is built-in in both

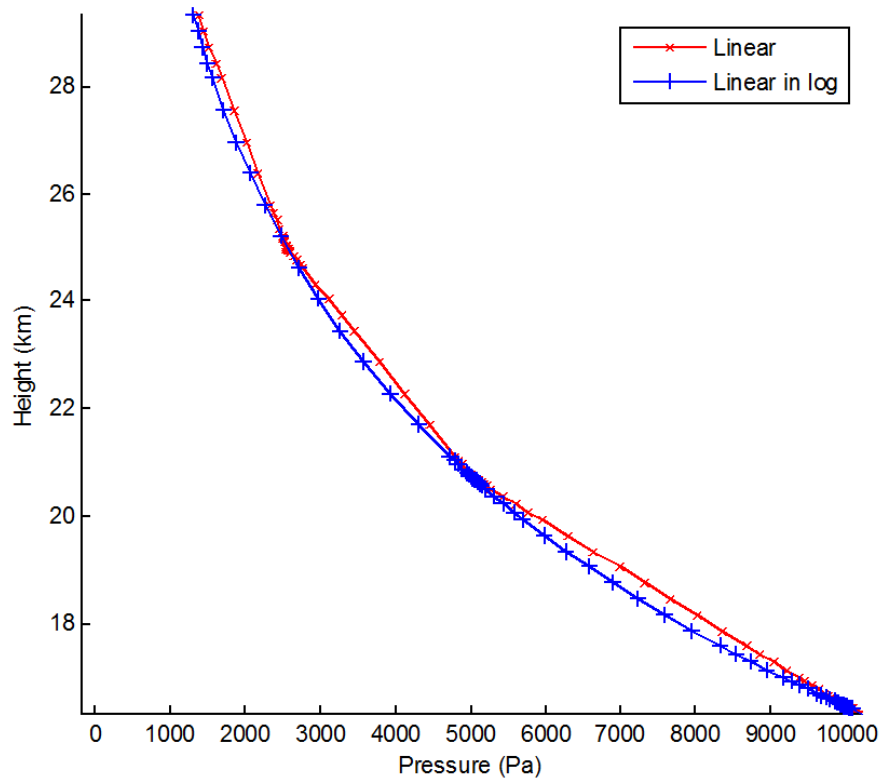


Figure III.2: Consequence of linear versus log-linear interpolation of pressure (zoom).

eq. (II.11) and Saastamoinen's [1972] model.¹

¹Obtaining pressure as per eq. (II.11) usually would not imply any additional assumption, since hydrostatic equilibrium is already built-in in most atmospheric sources — see footnote on p. 24.

Appendix IV

Humidity conversion

There are many ways to express humidity values. For example, the refractivity formulation eq. (2.3) or eq. (2.28) asks for partial pressure of water vapor; radiosondes usually measure relative humidity [Dabberdt et al., 2002] but are required to report dew point temperature, following international convention [Brettle and Galvin, 2003]; in the Canadian regional NWM [Côté et al., 1998] as well as in the CIRA86aQ_UoG climatology [Kirchengast et al., 1999], specific humidity is preferred. The present section aims at serving as a handy primer on converting among different yet equivalent forms of expressing humidity.¹

Tables IV.1 and IV.2 summarize the expressions to convert from any humidity variable to any other (in a total of 6^2 combinations), using partial pressure of water vapor P_w as an intermediary variable; pressure P and temperature T are assumed available; details can be found in the derivations that follow in the rest of the present section.

Let us start considering a volume V of humid air at temperature T and total pressure P . Its constituents are mass m_d of dry air and mass m_w of water vapor. Its total mass m is therefore:

$$m = m_d + m_w. \tag{IV.1}$$

¹Useful general references are Guyot [1998]; Wallace and Hobbs [2006]; Glickman [2000]; Curry [2002a,b].

Table IV.1: Conversion *to* partial pressure of water vapor.

Humidity variable	Input	Output	Expression
Partial pressure of water vapor	P, T, P_w	P_w	$P_w = P_w$
Specific humidity	P, T, q	P_w	$P_w = qP / (\epsilon + (1 - \epsilon)q)$
Mixing ratio	P, T, r	P_w	$P_w = rP / (r + \epsilon)$
Relative humidity	P, T, RH	P_w	$P_w = (RH/100)P_w^s, P_w^s = P_w^s(T)$
Dew point temperature	P, T, T_d	P_w	$P_w = P_w^s(T = T_d)$
Virtual temperature	P, T, T_v	P_w	$P_w = P(1 - T/T_v) / (1 - \epsilon)$

Table IV.2: Conversion *from* partial pressure of water vapor.

Humidity variable	Input	Output	Expression
Partial pressure of water vapor	P, T, P_w	P_w	$P_w = P_w$
Specific humidity	P, T, P_w	q	$q = P_w \epsilon / (P_w \epsilon + P_d), P_d = P - P_w$
Mixing ratio	P, T, P_w	r	$r = (P_w / P_d) \epsilon, P_d = P - P_w$
Relative humidity	P, T, P_w	RH	$RH = 100 P_w / P_w^s, P_w^s = P_w^s(T)$
Dew point temperature	P, T, P_w	T_d	$T_d = (P_w^s)^{-1}(P_w^s = P_w)$
Virtual temperature	P, T, P_w	T_v	$T_v = T / (1 - (P_w / P)(1 - \epsilon))$

Specific humidity q is defined as:

$$q \equiv \frac{m_w}{m}. \quad (IV.2)$$

Mixing ratio r is defined as

$$r \equiv \frac{m_w}{m_d}. \quad (IV.3)$$

The two are obviously closely related:

$$q = \frac{r}{r + 1}, \quad r = \frac{1}{1 - q}.$$

To have a sense for the magnitude of such variables, we quote Guyot [1998], who states that “In the atmosphere, the magnitude of r is a few grams per kilogram in the middle latitudes, but in the tropics it can reach approximately 20 g/kg.”

The relationship between *partial pressure of water vapor* P_w and either q or r requires a bit of theory. The partial pressure of a particular gas in a mixture of gases is the pressure

that such a gas would exert if it were to occupy alone the volume occupied by the whole mixture, at the same temperature. Assuming that the constituent gases do not interact chemically, *Dalton's law of partial pressures* states that

$$P = P_d + P_w. \quad (\text{IV.4})$$

Each gas constituent is assumed to obey the *state equation of ideal gases*,

$$P_d = \rho_d T R_d, \quad (\text{IV.5})$$

$$P_w = \rho_w T R_w, \quad (\text{IV.6})$$

where the same temperature T is shared by the two gases, following the assumption that they are in equilibrium. ρ_d and ρ_w are the *gases densities* (ρ_w is sometimes called *absolute humidity* or *volumetric humidity*),

$$\rho_d = m_d/V, \quad (\text{IV.7})$$

$$\rho_w = m_w/V; \quad (\text{IV.8})$$

R_d, R_w are, respectively, the specific *gas constants* for dry gases and water vapor:

$$R_d \equiv R^*/M_d,$$

$$R_w \equiv R^*/M_w,$$

defined in terms of the *universal gas constant*

$$R^* \equiv 8.316963 \times 10^3 \text{ J/kmol}\times\text{K},$$

and the *molar masses* of dry air and water vapor,

$$M_d \equiv 28.9644 \text{ kg/kmol},$$

$$M_w \equiv 18.0152 \text{ kg/kmol}.$$

(The numerical values are taken from Glickman [2000].) With that, the mixing ratio r can be rewritten as

$$\begin{aligned}
 r &= \frac{m_w}{m_d} \\
 &= \frac{\rho_w V}{\rho_d V} \\
 &= \frac{\rho_w}{\rho_d} \\
 &= \frac{P_w / (TR_w)}{P_d / (TR_d)} \\
 &= \frac{P_w / R_w}{P_d / R_d} \\
 &= \frac{P_w M_w / R^*}{P_d M_d / R^*} \\
 &= \frac{P_w M_w}{P_d M_d} \\
 &= \frac{P_w}{P_d} \varepsilon,
 \end{aligned} \tag{IV.9}$$

where

$$\varepsilon \equiv \frac{M_w}{M_d}. \tag{IV.10}$$

Similarly, specific humidity q can be expressed as

$$\begin{aligned}
 q &= \frac{m_w}{m_w + m_d} \\
 &= \frac{\rho_w V}{\rho_w V + \rho_d V} \\
 &= \frac{\rho_w}{\rho_w + \rho_d} \\
 &= \frac{P_w / (TR_w)}{P_w / (TR_w) + P_d / (TR_d)} \\
 &= \frac{P_w / R_w}{P_w / R_w + P_d / R_d} \\
 &= \frac{P_w M_w / R^*}{P_w M_w / R^* + P_d M_d / R^*} \\
 &= \frac{P_w M_w}{P_w M_w + P_d M_d} \\
 &= \frac{P_w M_w}{P_w M_w + P_d M_d} \frac{1/M_d}{1/M_d} \\
 &= \frac{P_w \varepsilon}{P_w \varepsilon + P_d}.
 \end{aligned} \tag{IV.11}$$

The inverse relationships require some care in their derivation: we shall never try to simplify expressions dividing them by q or P_w , because those quantities might assume the zero value:

$$\begin{aligned}
 r &= \frac{P_w}{P_d} \varepsilon \\
 r &= \frac{P_w}{P - P_w} \varepsilon \\
 r - r P_w - P_w \varepsilon &= 0 \\
 -P_w (r + \varepsilon) &= -r P \\
 P_w &= \frac{r P}{r + \varepsilon}
 \end{aligned}$$

and

$$q = \frac{P_w \varepsilon}{P_w \varepsilon + P_d}$$

$$q = \frac{P_w \varepsilon}{P_w \varepsilon + P - P_w}$$

$$qP_w \varepsilon + qP - qP_w - P_w \varepsilon = 0$$

$$P_w(q\varepsilon - q - \varepsilon) = -qP$$

$$P_w = \frac{-qP}{q\varepsilon - q - \varepsilon}$$

$$P_w = \frac{qP}{\varepsilon + q - q\varepsilon}$$

$$P_w = \frac{qP}{\varepsilon + (1 - \varepsilon)q}.$$

In the remaining conversions below, we will need the so-called *saturation water vapor pressure* P_w^s . It is important to realize that P_w^s itself is *not* a way of expressing humidity values, just an auxiliary quantity involved in the conversion. It is defined as the partial pressure at which water vapor is in equilibrium with a flat surface of liquid water, at a given temperature (equilibrium meaning that evaporation and condensation are both occurring — at the same rate, though). There are numerous formulas in use for P_w^s . Murphy and Koop [2005] provide a recent thoroughly review and comparison of both old and new formulations against experimental data, from which we quote:

All of the commonly used parametrizations for the vapor pressure of supercooled water are extrapolations that were not originally intended for use below the freezing point. In addition, the World Meteorological Organization definition of the vapor pressure of supercooled water contains an easily overlooked typographical error. Recent data on the molar heat capacity of supercooled water are used [by us] to derive its vapor pressure.

We therefore adopt their formulation:

$$\begin{aligned}
\ln(P_w^s) \approx & + 54.842763 \\
& - 6763.22/T \\
& - 4.210\ln(T) \\
& + 0.000367T \\
& + \tanh(0.0415(T - 218.8)) \\
& \times (53.878 - 1331.22/T - 9.44523\ln(T) + 0.014025T),
\end{aligned} \tag{IV.12}$$

valid for $123\text{ K} < T < 332\text{ K}$. Below we denote $P_w^s = P_w^s(T)$ the function that returns the value of P_w^s at a particular value of T . We also need the inverse function, which we denote $T = (P_w^s)^{-1}(P_w^s)$, returning the value of T at a particular value of P_w^s .²

With P_w^s at hand, we define *relative humidity* RH as the ratio

$$RH \equiv 100 \frac{P_w}{P_w^s}, \tag{IV.13}$$

and *dew point temperature* T_d as the temperature to which the air must be lowered, at constant pressure, for the water vapor to saturate:

$$T_d \mid P_w = P_w^s(T = T_d). \tag{IV.14}$$

To have a sense for such variables, we quote Wallace and Hobbs [2006, p. 83]:

²Equation (IV.12) cannot be inverted analytically, so we resort to the Newton–Raphson numerical method; a good approximate solution upon which to start and improve iteratively is given by Bolton [1980, eq. (10)]:

$$\begin{aligned}
x &= \ln((P_w^s/1000)/0.6112)/17.67, \\
T &= 273.15 + 243.5x/(1-x).
\end{aligned}$$

At the earth's surface, the pressure typically varies by only a few percent from place to place and from time to time. Therefore, the dew point [temperature] is a good indicator of the moisture content of the air. In warm, humid weather the dew point is also a convenient indicator of the level of human discomfort. (...) In contrast to the dew point, relative humidity depends as much upon the temperature of the air as upon its moisture content.

Finally, *virtual temperature* T_v is defined such that

$$T_v | P = \rho T_v R_d, \quad (\text{IV.15})$$

where $P = P_d + P_w$ and $\rho = \rho_d + \rho_w$ refer to the moist air as a whole, even though eq. (IV.15) employs the constant of dry gases R_d , as explained by Wallace and Hobbs [2006, p. 66]:

Moist air has a lower apparent molecular weight than dry air. Therefore, the gas constant for 1 kg of moist air is larger than that for 1 kg of dry air. However, rather than use a gas constant for moist air, the exact value of which would depend on the amount of water vapor in the air (which varies considerably), it is more convenient to retain the gas constant for dry air and use a fictitious temperature (called the *virtual temperature*) in the ideal gas equation.

To express T_v in terms of P_w , first we re-write ρ as

$$\begin{aligned} \rho &= P_d / (TR_d) + P_w / (TR_w) \\ &= P_d M_d / (TR^*) + P_w M_w / (TR^*) \\ &= \frac{P_d M_d + P_w M_w}{TR^*} \\ &= \frac{P_d M_d + P_w M_w}{TR^*} \frac{M_d P}{M_d P} \\ &= \frac{PM_d}{TR^*} (1 - (P_w/P)(1 - \epsilon)), \end{aligned} \quad (\text{IV.16})$$

then we isolate T_v and substitute ρ , obtaining

$$\begin{aligned} T_v &= P/(\rho R_d) \\ &= P/(\rho R^*/M_d) \\ &= PM_d/(\rho R^*) \\ &= PM_d(R^* \frac{PM_d}{TR^*} (1 - (P_w/P)(1 - \varepsilon))) \\ &= \frac{T}{1 - (P_w/P)(1 - \varepsilon)}. \end{aligned} \tag{IV.17}$$

Appendix V

Loxodrome in a gradient atmospheric structure

In section 3.2.4 we met the following integral eq. (3.12):

$$f_{\nabla_H}(\varphi, \lambda, h) \equiv \int_{r_{0,h}}^r \nabla_{H^0}(\mathbf{r}) \cdot d\mathbf{r}, \quad (3.12)$$

According to the *gradient theorem*, also known as the fundamental theorem of calculus for line integrals, we have:

$$\Phi(\mathbf{r}_2) - \Phi(\mathbf{r}_1) = \int_L \nabla\Phi(\mathbf{r}) \cdot d\mathbf{r}; \quad (V.1)$$

in words: a line integral through a gradient field is path-independent (i.e., depends only on the path end-points, \mathbf{r}_1 and \mathbf{r}_2).¹ Therefore, we are free to choose at will the integration path in eq. (3.12). There is one path, though — called the *loxodrome* — over which the

¹That is true for any gradient field, i.e., for any vector field that is the gradient of a scalar field.

integral eq. (3.12) rigorously simplifies to the following closed-form expression eq. (3.13):

$$f_{\nabla_H}(\varphi, \lambda, h) = l \left(\sin a \left. \frac{\partial v_0}{\partial x} \right|_h + \cos a \left. \frac{\partial v_0}{\partial y} \right|_h \right), \quad (3.13)$$

Now we demonstrate how to proceed to achieve the simplification above.

First recall that when we write \mathbf{r} we imply that the position vector is expressed in global Cartesian coordinates, $\mathbf{r} = X\hat{\mathbf{i}} + Y\hat{\mathbf{j}} + Z\hat{\mathbf{k}}$. When we need the same vector expressed in local Cartesian coordinates, then we denote it as $\mathbf{r}^{xyz} = x\hat{\mathbf{i}} + y\hat{\mathbf{j}} + z\hat{\mathbf{k}}$, instead. The transformation from \mathbf{r}^{xyz} to \mathbf{r} involves a translation and a rotation; the Jacobian of that transformation relates increments in x, y, z to increments in X, Y, Z :

$$\mathbf{J}(\mathbf{r}) = \left. \frac{\partial(X, Y, Z)}{\partial(x, y, z)} \right|_{\mathbf{r}};$$

\mathbf{J} is a 3-by-3 matrix containing the partial derivatives of each global Cartesian coordinate X, Y, Z with respect to each other local Cartesian coordinate x, y, z . Furthermore, \mathbf{J} is a rotation matrix,² therefore it is an orthogonal matrix, whose inverse equals its transpose, $\mathbf{J}^{-1} = \mathbf{J}^T$.

Local Cartesian coordinates is a very convenient coordinate basis for use with a gradient atmosphere, because we postulate for such an atmosphere the horizontal gradient vector ∇_{Hv_0} to be constant (for different positions \mathbf{r} yet same height h) under such a basis:

$$\nabla_H^{xyz} v_0(\mathbf{r}) = \nabla_H^{xyz} v_0(h) = \left. \frac{\partial v_0}{\partial x} \right|_h \hat{\mathbf{i}} + \left. \frac{\partial v_0}{\partial y} \right|_h \hat{\mathbf{j}} + 0\hat{\mathbf{k}}.$$

In contrast, the same horizontal gradient is *not* constant when expressed in a global Carte-

²The translation contained in the transformation itself is absent in the Jacobian, since the derivative of a constant term is zero.

sian coordinate basis:

$$\nabla_{\mathbf{H}} v_0(\mathbf{r}) = \mathbf{J}(\mathbf{r}) \nabla_{\mathbf{H}}^{xyz} v_0(h) = \frac{\partial v_0}{\partial X} \Big|_{\mathbf{r}} \hat{\mathbf{i}} + \frac{\partial v_0}{\partial Y} \Big|_{\mathbf{r}} \hat{\mathbf{j}} + \frac{\partial v_0}{\partial Z} \Big|_{\mathbf{r}} \hat{\mathbf{k}}.$$

That non-constancy is caused by the position-dependent Jacobian $\mathbf{J}(\mathbf{r})$. The transformation of the gradient vector between global and local Cartesian coordinate bases is allowed because the gradient is invariant under orthogonal transformations.

The vector increment $d\mathbf{r}$ can also be expressed in the same way:

$$d\mathbf{r} = \mathbf{J}(\mathbf{r}) d\mathbf{r}^{xyz}.$$

Now let us rewrite eq. (3.12) expressing the vectors in local Cartesian coordinates:

$$f_{\nabla_{\mathbf{H}}}(\varphi, \lambda, h) \equiv \int_{\mathbf{r}_{0,h}}^{\mathbf{r}} (\mathbf{J}(\mathbf{r}) \nabla_{\mathbf{H}}^{xyz} v_0(h)) \cdot (\mathbf{J}(\mathbf{r}) d\mathbf{r}^{xyz}); \quad (\text{V.2})$$

expanding the spatial vectors as linear algebra vectors, we get:

$$f_{\nabla_{\mathbf{H}}}(\varphi, \lambda, h) \equiv \int_{\mathbf{r}_{0,h}}^{\mathbf{r}} \left[\frac{\partial v_0}{\partial x} \Big|_h, \frac{\partial v_0}{\partial y} \Big|_h, 0 \right] \mathbf{J}^T \mathbf{J} \begin{bmatrix} dx \\ dy \\ dz \end{bmatrix}, \quad (\text{V.3})$$

where the transpose on the first operand is required to guarantee dimensional consistency, when the spatial dot-product is expressed as a linear algebra matrix/vector product.

With the form above (eq. (V.3)), we can perform the simplification to achieve the desired closed-form expression eq. (3.13) in four steps. First, the Jacobian \mathbf{J} being orthogonal, the product $\mathbf{J}^T \mathbf{J}$ cancels out to an identity matrix:

$$f_{\nabla_{\mathbf{H}}}(\varphi, \lambda, h) = \int_{\mathbf{r}_{0,h}}^{\mathbf{r}} \left(\frac{\partial v_0}{\partial x} \Big|_h \hat{\mathbf{i}} + \frac{\partial v_0}{\partial y} \Big|_h \hat{\mathbf{j}} + 0 \hat{\mathbf{k}} \right) \cdot (dx \hat{\mathbf{i}} + dy \hat{\mathbf{j}} + dz \hat{\mathbf{k}}). \quad (\text{V.4})$$

Second, the partial derivatives with respect to local Cartesian coordinates being constant for the same height h , they can be taken out of the integral if we evaluate the integral along a path of constant height:

$$f_{\nabla_H}(\varphi, \lambda, h) = \left(\frac{\partial v_0}{\partial x} \Big|_h \hat{\mathbf{i}} + \frac{\partial v_0}{\partial y} \Big|_h \hat{\mathbf{j}} + 0\hat{\mathbf{k}} \right) \cdot \int_{r_{0,h}}^r \left(dx\hat{\mathbf{i}} + dy\hat{\mathbf{j}} + dz\hat{\mathbf{k}} \right). \quad (\text{V.5})$$

Third, if we evaluate the integral along a path of constant azimuth a — the so-called *loxodrome* — (in addition to constant height h), then the increment:

$$dx\hat{\mathbf{i}} + dy\hat{\mathbf{j}} + dz\hat{\mathbf{k}} = \left(\sin(a)\hat{\mathbf{i}} + \cos(a)\hat{\mathbf{j}} + 0\hat{\mathbf{k}} \right) dl,$$

becomes constant, which allows us to take the vector factor out the integral

$$f_{\nabla_H}(\varphi, \lambda, h) = \left(\frac{\partial v_0}{\partial x} \Big|_h \hat{\mathbf{i}} + \frac{\partial v_0}{\partial y} \Big|_h \hat{\mathbf{j}} + 0\hat{\mathbf{k}} \right) \cdot \left(\sin(a)\hat{\mathbf{i}} + \cos(a)\hat{\mathbf{j}} + 0\hat{\mathbf{k}} \right) \int_{r_{0,h}}^r dl. \quad (\text{V.6})$$

$h=\text{const.}$
 $a=\text{const.}$

Fourth, replacing the remaining integral, $\int dl$, for the loxodrome length, l , we obtain, finally, the desired closed-form expression eq. (3.13):

$$f_{\nabla_H}(\varphi, \lambda, h) = l \left(\sin a \frac{\partial v_0}{\partial x} \Big|_h + \cos a \frac{\partial v_0}{\partial y} \Big|_h \right). \quad (3.13)$$

And that concludes the derivation.

Apart from the derivation above, the loxodrome azimuth a and length l can be calculated as per Alexander [2004].³ Were the Earth plane, the azimuth and length would be simply the straight-line azimuth and distance, and the simplification of the integral would be trivial. But recall that the ray at a low elevation angle travels several hundred kilome-

³In Alexander's [2004] expressions, the height must be added to the spherical radius, to obtain the loxodrome at the desired height.

tres before exiting the neutral atmosphere; at distances of this extent, the Earth's curvature plays a role.

Appendix VI

Refractivity coefficients

There is some controversy around the refractivity coefficients for radio waves in the neutral atmosphere. See Thessin [2005, sec. 2.3, p. 38–42.], Rüeger [2002], Mendes [1999, p. 59–60], and Bevis et al. [1994] for discussions. Especially Rüeger [2002] provides a recent, thorough, review.

Mendes [1999, sec. 5.1.3, p. 163–164] reports negligible, sub-mm, discrepancies in zenith hydrostatic delay due to four different determinations of those coefficients. In addition to the determinations compared by Mendes [1999], we wanted to compare two determinations, Bevis et al. [1994] and Rüeger [2002], that recognized the controversy and aimed at providing *best average* coefficients, an alternative to best available values that “provides a certain robustness against unmodeled systematic errors and increases the reliability of the values” [Rüeger, 2002, p. 6]. The two other determinations (included in Mendes [1999]) that we compared were IUGG and Thayer [1974]. IUGG is a slight simplification of Essen and Froome [1951]; it was endorsed by a resolution of the International Union of Geodesy and Geophysics (IUGG) in 1963, a resolution which is still officially in effect.¹ Thayer’s [1974] is disputed by both Rüeger [2002] and Bevis et al. [1994], on the

¹The 1963 resolution is in effect regarding radio refractivity coefficients; in contrast, regarding optical refractivity coefficients, it was recently superseded by a newer resolution.

grounds that it is biased compared to other determinations and unrealistically precise. In fact, its claimed (optimistic) precision is probably the reason why Thayer's [1974] is still widely in use, contrary to advice from both Rüeger [2002] and Bevis et al. [1994]. All refractivity values are reported in Table VI.1.

We performed the comparison for a single profile, at a single epoch; that was sufficient for our objective here, namely, to check the order of magnitude of the discrepancies involved. We compared zenith delays, slant delays, and slant factors. The discrepancies, taken with respect to IUGG (chosen as the reference because it is the one determination still officially endorsed by the IUGG), are shown in Tables VI.3, VI.4, and VI.5 for zenith delays, slant delays, and slant factors, respectively. Slant quantities correspond to a direction at 3° elevation angle. Discrepancies in slant factors are reported multiplied by IUGG's zenith delay, for a more intuitive interpretation (i.e., in metre rather than unitless). IUGG's zenith and slant delays themselves are shown in Table VI.2. The delays were integrated setting a tolerance of 0.1 mm for the ray-tracer.

Our results are in agreement with Mendes [1999, Table 5.2, p. 164]: sub-mm discrepancy in zenith hydrostatic delay between IUGG and Thayer [1974]. The second thing to notice in Tables VI.3 to VI.5 is that the results based on Rüeger [2002] stand out from the others; in fact, it is the only determination in the comparison which shows non-negligible discrepancies with respect to IUGG, both in zenith and slant delays. The very thorough review presented in Rüeger [2002] — including results of comparisons between the relatively simple closed-form expression usually employed (eqs. (2.3) and (2.28)) and more rigorous, non-analytical formulations — gives credit to his determination (more correctly, to his recommendations, based on determinations made by other authors). In other words, the non-negligible discrepancies found with respect to IUGG cast doubt over the IUGG resolution still officially in effect.

Finally, the discrepancy most relevant for the present work is that in terms of slant

factors (Table VI.5): it is very negligible, but not quite zero. That is as expected: on one hand, we do expect slant factors to be less affected than slant delays, because factors are a ratio of two delays, both affected by the same refractivity coefficients; on the other hand, we do not expect factors to be completely unaffected, because even though refractivity N (eq. (2.3)) is linear in the coefficients k_1, k_2, k_3 , the slant delay d is slightly non-linear in refractivity N (because of ray bending).

In this work, we employed Thayer’s [1974] determination — we were unaware of Rüeger’s [2002] and Bevis et al.’s [1994] when we started. Luckily, results for slant factors, which constitute the bulk of this work, are unaffected.

Table VI.1: Refractivity coefficient values.

Coefficient	$k_1 (\times 10^{-2} \text{K/Pa})$	$k_2 (\times 10^{-2} \text{K/Pa})$	$k_3 (\times 10^{+3} \text{K}^2/\text{Pa})$
Rüeger [2002]	77.6890	71.2952	3.75463
Bevis et al. [1994]	77.60	70.4	3.739
Thayer [1974]	77.60	64.8	3.776
IUGG	77.624	64.700	3.71897

Table VI.2: Zenith and slant delays based on refractivity coefficients from IUGG.

Delay component	Zenith delay (m)	Slant delay (m)
Total	2.3562	34.6093
Geometric	0.0000	0.4920
Hydrostatic	2.3406	33.8675
Non-hydrostatic	0.0154	0.2498

Table VI.3: Discrepancy (with respect to IUGG) in zenith delay (in m) due to different determinations of the refractivity coefficients.

Component	Rüeger [2002]	Bevis et al. [1994]	Thayer [1974]
Total	0.0022	-0.0006	-0.0005
Geometric	0	0	0.0000
Hydrostatic	0.0020	-0.0007	-0.0007
Non-hydrostatic	0.0002	0.0001	0.0002

Table VI.4: Discrepancy (with respect to IUGG) in slant delay (in m) due to different determinations of the refractivity coefficients.

Component	Rüeger [2002]	Bevis et al. [1994]	Thayer [1974]
Total	0.0319	-0.0081	-0.0067
Geometric	0.0009	-0.0002	-0.0001
Hydrostatic	0.0274	-0.0103	-0.0104
Non-hydrostatic	0.0035	0.0023	0.0038

Table VI.5: Discrepancy (with respect to IUGG) in slant factor (multiplied by a nominal zenith delay, thus in m) due to different determinations of the refractivity coefficients.

Component	Rüeger [2002]	Bevis et al. [1994]	Thayer [1974]
Total	-0.0001	0.0004	0.0005
Geometric+Hydrostatic	-0.0004	0.0002	0.0002
Non-hydrostatic	-0.0000	0.0000	0.0000

Appendix VII

Related work in secondary areas

In section 1.4 we stated that ray-tracing has connections to many different subject areas.¹ We further acknowledged that among those areas, we see the present work falling under the umbrella of geodesy, more specifically along the thread of developments in mapping functions for radio space geodetic applications. Whereas in section 2.8 we delved into that thread, in the present section we would like to quickly skim through the secondary connections. Secondary areas are relevant because they might offer innovative solutions to the shared problem of ray-tracing — after all, sometimes a piece of knowledge that is well-known in one area might be considered an innovation in a different area.

In *astronomy*, more specifically in *astrometry* [Kovalevsky and Seidelmann, 2004], exactly the same phenomena of interest in this work has been studied under the name *astronomical refraction* for millennia, literally [Lehn and van der Werf, 2005]. One main difference between the two application areas is that, while we are interested in the refraction delay, they are interested in the refraction angle. See, for example, Yatsenko [1995]; Wittmann [1997]; Auer and Standish [2000]; Young [2004]. There is a beneficial overlap between astrometry and geodesy; for example, VLBI serves equally well the two areas

¹See Kravtsov [2005] for a general treatment.

[Sovers et al., 1998]. That overlap in matters of refraction started with the birth of space geodesy and remains to the present day.

In *surveying* there is a related topic under the name of *terrestrial refraction*, sometimes also called *geodetic refraction* [Brunner, 1984; Ingensand, 2002]. Surveying itself has a wide overlap with geodesy, under terrestrial geodesy (in contrast with space geodesy), as well as with metrology, under applied or industrial metrology. Refraction is studied in surveying to correct distances and angles measured with, e.g., electronic distance meters (EDM), theodolites, leveling instruments, etc. In that case, receiver and transmitter are usually co-located and one employs additionally a reflective target; in terrestrial refraction they are all within the Earth's atmosphere.

In *radio science*, a remote sensing technique known as *radio occultation* (RO) has been used since the early 1970's for measuring the physical properties of planetary atmospheres. In the last decade we have seen its application to the Earth's atmosphere, employing GPS signals (thus *GPS radio occultation* — Kursinski et al. [1997]). Ray-tracing is embedded in the so-called *data assimilation operators* (see Syndergaard et al. [2006] and references therein) that provide the model counterparts to the RO measurements, allowing their exploitation in weather and climate forecasting.

In *computer science*, more specifically *computer graphics*, there is a rich technique under the same name [Rademacher, 1997; Pharr and Humphreys, 2004]. The application intended is, in a nutshell, the computerized synthesis of images depicting virtual scenes. It differs from our usage primarily because it seeks to quantify radiance instead of phase. It differs also because usually it allows for refraction only at the surface of objects or at the interface of media, each of which is assumed to have uniform index of refraction. There are exceptions, though, under the key-word *ray-tracing in inhomogeneous media*; see Gutierrez et al. [2006]; Seron et al. [2005].

In *optics*, ray-tracing is a century-old technique employed in the design of systems of

lenses, for instruments such as telescopes and microscopes, not necessarily restricted to the visual band [Spencer and Murty, 1962]. Like in computer graphics, usually the index of refraction is not allowed to vary gradually. The exceptions fall under the key-words *gradient-index optics* [Marchand, 1978]. Also periodicals dedicated to *applied optics* often serve as vehicles for publications related to our subject area, e.g., van der Werf [2003]; Hase and Höpfner [1999]; Nener et al. [2003].

In *seismology*, geophysicists use ray-tracing to aid in earthquake location, measuring plate tectonic processes, etc. [Rawlinson et al., 2008; Červený, 2001]. Their application is more complicated than ours because the Earth's interior can be much more heterogeneous a medium than the Earth's atmosphere. It is further complicated by the fact that the transmitter is contiguous to the media, instead of being far away, as in the case of GPS satellites or VLBI sources. On the other hand, in their case at least the receiver and transmitter are at rest with respect to the propagation medium, in contrast to our case.

Vita

Full Name: Felipe Geremia Nievinski

Place and Date of Birth: Porto Alegre, Rio Grande do Sul, Brazil; December 27, 1980.

Permanent Address:

Rua Gonçalves Dias, 478/206
Porto Alegre, RS, Brazil
90130-060

Schools Attended:

Escola Técnica da Universidade Federal do Rio Grande do Sul (1996–1997)
Secondary and technical school (Computer systems).

Escola Técnica da Universidade Federal do Rio Grande do Sul (1998–1999)
Technical school (Network systems).

Universities Attended:

Universidade Federal do Rio Grande do Sul (1999–2004)
Undergraduate degree
(Bachelor in Geomatics Engineering, “Engenharia Cartográfica”).

Publications:

Nievinski, F., K. Cove, M. Santos, D. Wells, and R. Kingdon (2005). “Range-extended GPS kinematic positioning using numerical weather prediction model.” *Proceedings of the ION 61st Annual Meeting — ION AM 2005*, The Institute of Navigation, Cambridge, Mass., June 27–29, pp. 902–911. <http://gge.unb.ca/Research/GRL/PrincessOfAcadia/Publications/E2-2-Nievinski.pdf>.

Santos, M., F. Nievinski, K. Cove, R. Kingdon, and D. Wells (2005). “Range-extended post-processing kinematic (PPK) in a marine environment.” *Proceedings of the 18th International Technical Meeting of the Satellite Division of the Institute of Navigation — ION GNSS 2005*, The Institute of Navigation, Long Beach, Calif., September 13–16, pp. 805–809, (best presentation award). <http://gge.unb.ca/Research/GRL/PrincessOfAcadia/Publications/E2-2-Nievinski.pdf>.

- Nievinski, F. G. (2006). "Numerical weather models for tropospheric mitigation in marine kinematic GPS: A daylong analysis." *Proceedings of the 19th International Technical Meeting of the Satellite Division of the Institute of Navigation — ION GNSS 2006*, The Institute of Navigation, Fort Worth, Tex., September 26–29, pp. 2017–2026, (student paper award). <http://gge.unb.ca/Research/GRL/PrincessOfAcadia/Publications/NievinskiIONGNSS2006.pdf>.
- Nievinski, F. G., and M. C. Santos (2007). "An analysis of errors introduced by the use of transformation grids." In *Dynamic Planet — Monitoring and Understanding a Dynamic Planet with Geodetic and Oceanographic Tools, IAG Symposium, Cairns, Australia, 22–26 August, 2005*, eds. P. Tregoning, and C. Rizos, Vol. 130 of *International Association of Geodesy Symposia*, Springer, Berlin, ISBN 978-3-540-49349-5, pp. 677–684, doi:10.1007/978-3-540-49350-1_98.
- Sükeová, L., M. C. Santos, R. B. Langley, R. F. Leandro, O. Nnani, and F. Nievinski (2007). "GPS L2C signal quality analysis." *Proceedings of the 63rd Annual Meeting*, The Institute of Navigation, Cambridge, Mass., April 23–25, pp. 232–241. <http://gauss.gge.unb.ca/papers.pdf/ionam2007.sukeova.pdf>.
- Chaves, J. C., M. C. Santos, F. G. Nievinski, and M. R. Craymer (2008). "Distortions sources in the Brazilian geodetic network." *Boletim de Ciências Geodésicas*, Vol. 14, No. 2, pp. 186–199, (in Portuguese), <http://ojs.c3s1.ufpr.br/ojs2/index.php/bcg/article/view/11815/8326>.
- Oliveira, L. C., M. C. Santos, F. G. Nievinski, R. F. Leandro, S. M. A. Costa, M. F. Santos, J. Magna, M. Galo, P. O. Camargo, J. G. Monico, C. U. Silva, and T. B. Maia (2008). "Searching for the optimal relationships between SIRGAS2000, South American Datum of 1969 and Córrego Alegre in Brazil." In *Observing our Changing Earth — Proceedings of the 2007 IAG General Assembly, Perugia, Italy, July 2–13*, ed. M. G. Sideris, Vol. 133 of *International Association of Geodesy Symposia*, Springer, Berlin, ISBN 978-3-540-85425-8, pp. 71–79, doi:10.1007/978-3-540-85426-5_9.

Presentations:

- Nievinski, F. (2005). "Further developments in range-extended GPS kinematic positioning using numerical weather prediction models." 2005 Semi-Annual Graduate Seminar, University of New Brunswick, Department of Geodesy and Geomatics Engineering, Fredericton, N.B., December 8. (Oral presentation).
- Nievinski, F. G., and M. C. Santos (2006). "Further developments in range-extended GPS kinematic positioning using a numerical weather prediction model." 2006 Joint Assembly, American Geophysical Union, Geochemical Society, The Microbeam Analysis Society, Mineralogical Society of America, Society of Exploration Geophysicists, Unión Geofísica Mexicana, Baltimore, Md., May 23–26, *Eos Trans. AGU Jt. Assem. Suppl.*, Abstract G41A-02. (Poster presentation).

- Nievinski, F. G. (2007). "Functional and stochastic modeling of distortion in reference frames." 2007 Semi-Annual Graduate Seminar, University of New Brunswick, Department of Geodesy and Geomatics Engineering, Fredericton, N.B., November 9. (Oral presentation).
- Nievinski, F. G., and M. C. Santos (2007a). "Discrepancy in neutral atmospheric delay due to different models for ray path and atmosphere in raytracing." Fall Meeting 2007, American Geophysical Union, San Francisco, Calif., December 10–14, Eos Trans. AGU Fall Meet. Suppl., Abstract G43A-0915. (Oral presentation).
- Nievinski, F. G., and M. C. Santos (2007b). "Accuracy in zenith tropospheric delay from the Canadian regional numerical weather model." 2007 Joint Congress, Canadian Meteorological and Oceanographical Society, Canadian Geophysical Union, and American Meteorological Society, Saint John's, N.L., May 28–June 1. (Oral presentation).
- Nievinski, F. G., and M. C. Santos (2007c). "Accuracy in zenith tropospheric delay from the Canadian regional numerical weather model." Earth: Our Changing Earth — IUGG XXIV General Assembly, International Union of Geodesy and Geophysics, Perugia, Italy, July 2–13. (Poster presentation).
- Santos, M. C., and F. G. Nievinski (2007). "A discussion on height systems used in geodesy and meteorology." 2007 Joint Congress, Canadian Meteorological and Oceanographical Society, Canadian Geophysical Union, and American Meteorological Society, Saint John's, N.L., May 28–June 1. (Oral presentation).
- Nievinski, F. G., L. Urquhart, and M. C. Santos (2008). "Impact of mapping functions based on spherical, ellipsoidal, gradient, and 3d atmospheric structures on GPS Precise Point Positioning." Fall Meeting 2008, American Geophysical Union, San Francisco, Calif., December 15–19, Eos Trans. AGU Fall Meet. Suppl., Abstract G41A-0611. (Poster presentation).