Atmospheric Propagation Effects

on

Radio Interferometry

by

James Louis Davis

B.S., Michigan State University
(1981)

Submitted in Partial Fulfillment
of the Requirements for the
Degree of

Doctor of Philosophy

at the

Massachusetts Institute of Technology
April, 1986

Signature of Author: _____

Department of Earth, Atmospheric, and Planetary Sciences
7 April, 1986

Certified by: _____

Irwin I. Shapiro, Thesis Supervisor

Accepted by: _____

Chairman, Department Committee on Graduate Studies

Atmospheric Propagation Effects

on Radio Interferometry

by

James Louis Davis

Submitted to the Department of Earth, Planetary, and Atmospheric Sciences of the Massachusetts Institute of Technology on April 7, 1986 in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Geophysics.

Abstract

The technique of very-long-baseline interferometry (VLBI) offers geodesy the potential of estimating distances on the earth of several thousand kilometers with uncertainties of a few centimeters or less. Since the completion of the Mark III VLBI system—which combines group-delay measurements at two widely separated frequency bands to estimate ionospheric refraction—the source of error that limits this accuracy has been thought to be refraction by the neutral atmosphere.

The largest component of the radio refractive index of air is due to the dry atmosphere. The integrated effect of this component can be estimated quite accurately from the surface pressure for a signal arriving at the site from the zenith direction. For a signal arriving from other directions, a model for the atmosphere must be combined with surface meteorological measurements to estimate the propagation delay. The accuracy of this estimation is limited by the accuracy of the atmospheric model. The remaining component of the radio refractive index of air is due to water vapor. Water vapor in the lower troposphere is unmixed, and its effect on the group delay can vary from 0–20% of the effect of the dry atmosphere. The "wet delay" is very difficult to model using measurements of surface meteorological parameters.

In this thesis, we attempt to develop an improved model for the elevation-angle dependence of the dry propagation delay. We first show that systematic, elevation-angle dependent errors in estimates of baseline length using the Marini mapping function are of order 5 cm for 8000-km baselines. We hypothesize, based on analysis of the effects of mapping-function errors, that these elevation-dependent systematic errors

are caused by errors in the Marini mapping function. We derive an improved mapping function from the analysis of ray-trace studies utilizing a parametrized model atmosphere. The propagation delay predicted by this new mapping function differs from ray-trace results by less than 5 mm at all elevation angles above 5° elevation. We present the initial testing of this mapping function using VLBI data, and show that estimates of baseline length, which had shown systematic behavior, show no perceptible systematic behavior with the use of the new mapping function.

In order to test further this new mapping function, we undertook a series of special VLBI experiments. These experiments include a large number of observations from low elevation angles. Thus far, we have analyzed eight of these "low-elevation" experiments, and we present their results. The estimates of the Mojave–Haystack baseline length from these experiments possess a weighted-root-mean square repeatability of ~4 mm when the data are analyzed using the new mapping function, whereas the repeatability is ~10 mm if analyzed using the Marini mapping function. Furthermore, these latter estimates show evidence of temperature–dependence. These tests are consistent with—but not conclusive proof for—the conclusion that the new mapping function has adequately modeled the seasonal behavior of the propagation delay.

We also discuss the technique of water-vapor radiometry. The derivation of the dual-frequency algorithm employing opacities as the "observables" appears for the first time. We point out several important features of this algorithm, and discuss possible sources of error. We also compare estimates of the wet propagation delay obtained from WVR data located at the Mojave site during the low-elevation experiments and processed with the dual-frequency algorithm to estimates of that delay obtained from VLBI data processed with a Kalman filter. The short term behavior of these two series of estimates is shown to agree quite well at the level of less than 1 cm, but there are biases between the two series with apparently seasonal behavior with an amplitude of nearly 2 cm. We hypothesize that this seasonal behavior can be attributed to the use of a constant value for the dual-frequency weighting function; recent studies have indicated that this value has a seasonal variation.

Thesis Supervisor: Dr. Irwin I. Shapiro
Paine Professor of Practical Astronomy, Harvard University
Professor of Physics, Harvard University
Director, Harvard–Smithsonian Center for Astrophysics

Acknowledgements

The completion of this thesis would not have been possible without the support of many people. I would at the outset like to thank my advisor, Irwin Shapiro, who by guidance and example has taught me something of the method of scientific inquiry, for which I will always be grateful.

A special thanks goes to Tom Herring, who through daily discussions has contributed immensely to this thesis and to my education; he has spent a great deal of his time guiding me through the intricacies of data analysis.

I would like to thank Brian Corey and George Resch for their time spent with me discussing water-vapor radiometers and the fundamentals of radiometry. Brian Corey also provided commentary and suggestions for the improvement of Chapter 2, and I am especially grateful for his continuing interest in water-vapor radiometers despite many frustrating experiences.

A source of information and enthusiasm has been Gunnar Elgered. I would especially like to thank him and his wife Vanda for their hospitality. I would also like to acknowledge the contributions of the entire WVR group at Onsala, Sweden, supervised by Bernt Rönnäng. These important contributions can be seen at various points throughout this thesis.

The "low-elevation" experiments would not have been possible without the help from the people at Interferometrics, Inc., including Nancy Vandenberg, Dave Schaffer,

and Steve Janoskie. I would also like to acknowledge Mark Hayes, Ed Himwich, and Goran Lundqvist for their contributions to the study of water-vapor radiometers.

Others to whom I would like to express my gratitude for a number of diverse reasons include John Ball, Tom Clark, Ioannis Ifadis, Mike Janssen, Jan Johansson, Chopo Ma, Jim Moran, Doug Robertson, Alan Rogers, Shoshana Rosenthal, Jim Ryan, Jeanne Sauber, John Webber, Alan Whitney, and Al Wu.

Several people from my "past" deserve my thanks for having an especially great influence. They include Donald Collins, Bill Harlow, and Jerry Nolen. I would also like to express my gratitude to my parents for their encouragement and guidance.

Finally, I would like to dedicate this thesis to Julie Kovacs: friend, fellow graduate student, scientist, and wife.

This work was supported by the National Aeronautics and Space Administration, grants NAG5-538, NAS5-27571 and NGT-22-009-904, the National Science Foundation, grants EAR83-06380 and EAR79-20253, and the Air Force Geophysics Laboratory, contracts F19628-83-K-0031 and F19628-82-K-0037.

Table of Contents

Introduction

The phenomena of absorption, scattering, and refraction are all macroscopic and familiar effects of the terrestrial atmosphere on visible light. The blueness of the sky, the redness of the setting sun, the haziness of the moon, and the blurriness of images above a hot road are everyday spectacles. The atmosphere affects the propagation of electromagnetic radiation at all frequencies, of course. In analyzing data obtained from radio interferometric observations, the effects of the atmosphere at radio wavelengths are most obviously of prime importance. In order to use these data to estimate parameters of geodetic, geophysical, or astronomical importance, the effects of the atmosphere must be understood and incorporated into the theoretical models that relate the data to the parameters of interest. If one cannot obtain some "outside" estimate of the atmospheric effects, then theoretical expressions for the effects must be developed and parametrized in terms of quantities that can be estimated along with the interesting parameters.

For the analyst of modern very-long-baseline interferometry (VLBI) data composed of group or phase delays, and phase–delay rates, the most significant atmospheric effect is the refraction of radio waves. Atmospheric absorption and scattering will attenuate the original radio signal, which will decrease the precision of the observable, but atmospheric refraction induces an effective delay which, as mentioned above, must be measured independently or modeled with parameters estimated using the VLBI data. The total effect is quite large: the delay induced by the atmosphere for a radio signal arriving from the zenith direction is ~2.4 m for a site at sea level. For signals

arriving in directions of lower elevation angles, the effect increases, and is about 10 m at 10° elevation, for example.

The simplest way to deal with the atmospheric delay would be to leave it out of the theoretical models; however, previous researchers have come to the conclusion that the resulting errors in estimated parameters of importance are intolerable. Much effort has therefore been spent to develop mathematical models for this delay. These efforts have been aided by the fact that the atmosphere of the earth is very well behaved in terms of its vertical structure, being very nearly in a condition of hydrostatic equilibrium, and having a constant mean molecular weight up to about 80 km, due to vertical mixing. These facts allow the prediction of the electrical path in the vertical direction, and hence the propagation delay in that direction, based solely on the surface pressure, regardless of actual temperature structure. (The propagation delay in this direction is know as the "zenith delay".) Complications arise, however, when one tries to predict the propagation delay for radiation arriving from other directions; the propagation delay then *does* depend on the temperature structure. Furthermore, the presence of water vapor, the local density of which can vary greatly both spatially and temporally, can contribute significantly to the propagation delay in all directions; this variability makes this contribution all but totally unpredictable using only theoretical models based on local meteorological conditions at the receiving location.

From the beginning of VLBI, it was realized that in order to take full advantage of the data obtained by this technique, the effect of tropospheric water vapor on the propagation delay would have to be overcome. (The problem of predicting the propagation delay in off-zenith directions was, at least for the time being, adequately

solved by existing models designed primarily for use in satellite tracking.) A promising instrument known as a water-vapor radiometer (WVR) was studied by several groups, and their results indicated that the WVR could be used to estimate the contribution of water vapor to the propagation delay (the "wet delay") with an accuracy of about 5 mm. These WVR's basically used the same design developed by Dicke in the 1940's to measure the strength of the 22 GHz line of water vapor, except that a second frequency was used to subtract the effects of liquid water (and, of course, the WVR's possessed updated electronics).

In the early 1980's, several VLBI sites as well as mobile VLBI systems started to become equipped with WVR's. In the United States, these WVR's were built at the Jet Propulsion Laboratory under the direction of George Resch. Two of these WVR's were mounted on mobile VLBI systems, while the fixed stations at Haystack/Westford, Owens Valley, Ft. Davis, and Mojave each received one. An independent effort by Gunnar Elgered and Bernt Rönnäng also provided the Onsala Space Observatory in Sweden with a WVR.

The next step was to prove the necessity and the sufficiency of the WVR as a calibration tool in VLBI data analysis. Although it seems that in light of the previous discussion concerning the variability of water vapor the necessity of the WVR might seem to be an obvious premise, rather than something which needs proving, up to that time there had not been any direct evidence in the VLBI data of the effects of water vapor. Thus, the actual size of the errors in the geodetic parameters induced by errors in the wet delay were unknown. Furthermore, it was not known whether the theory behind WVR's is accurate enough to be of use, or which of the infinity

of inversions relating atmospheric brightness temperature to wet path delay is the "best." Of course, if the existing WVR's were able to prove their own necessity, then that would go a long way towards proving their sufficiency.

The strategies for testing the WVR were at that time limited in number. Basically, they all were based on the idea of "trying out the WVR data on the VLBI data to see if anything gets better." The "anything" could be the RMS residual delay, or the repeatability of independent estimates of a particular baseline length or station position, or anything else that might be affected by errors in calibrating the wet delay.

The first hurdle to be overcome was obtaining WVR data simultaneously with a VLBI observing session. The WVR's at the U.S. sites all experienced severe mechanical difficulties such that they were inoperable for a large fraction of the time. Furthermore, the early data taken with these WVR's were extremely noisy, which was later shown to be due to imprecise determinations of the instrumental gains. As a result of these problems, no real conclusions could be reached as to the necessity for, or the sufficiency of, the WVR.

One of the primary subjects of this thesis is the subsequent research devoted to obtaining precise estimates of the wet delay from these WVR's, and employing these estimates to evaluate the concept of water-vapor radiometry. This research is discussed in Chapter 2. Chapter 1 reviews the effects of the atmosphere in VLBI, discusses propagation in the atmosphere, and describes how models may be developed for the predictions of these effects. In Chapter 3 and Appendix A, we discuss the development of a new model for the elevation-mapping of the dry part of the atmosphere, which, as will be described, was necessitated by the research on WVR's. In Chapter 4, we present the results from a series of VLBI experiments intended to study the elevation-angle behavior of the propagation delay and our ability to model this delay.

It should be noted that as this thesis is being written, there is much happening in the area of WVR's. The WVR's originally built by JPL have been rebuilt and are reappearing as instruments like the originals, but more precise and hopefully more durable. Furthermore, a different design is being implemented for a new series of WVR's. The discussions in Chapter 2 of accuracy and precision of WVR's will not apply to the WVR's with this newer design, and because these radiometers have three frequency channels, the discussions in Chapter 4 of the accuracy of the dual-frequency algorithm may also not apply.

Chapter 1

The Atmospheric Propagation Delay

Introduction

In this chapter, we will discuss the manner in which the atmosphere affects the VLBI data, and how this effect may be taken into account in the model. We will begin in Section 1.1 by writing down in its most general form the equation for propagation delay due to the atmosphere. In Section 1.2, we will discuss several specific properties of the refraction of radio waves in the atmosphere, including determination of the ray-path, refraction in a spherically symmetric atmosphere, the structure of the refractivity in the atmosphere, and dispersion. Prior to presenting formulas for the propagation delay, we will in Section 1.3 define some of the terminology often used, and then in Section 1.4 we will discuss the treatment of the propagation delay in VLBI data analysis, presenting the common formulas for the propagation delay. In Section 1.5 we will discuss the effects of errors in the formulas for the propagation delay, and develop simple models for these errors. In Section 1.6, we will discuss the options for treatment of the propagation delay in light of the results of Section 1.5.

1.1 The effect of propagation media on the group delay

1.1.1 The equation for the atmospheric propagation delay

In terms of the VLBI group-delay observable, which is discussed in detail in, *e.g.*, Herring [1983], the propagation delay is defined to be the difference between the

observed value for the group delay, and the value which would have been observed had there been no propagation medium. The sign of the propagation delay is such that if the propagation delay is added to the value for the *in vacuo* group delay, one obtains the value for the observed group delay.

Another quantity which the same term "propagation delay" describes is the difference between the travel time of a signal traveling from the radio source to a *single* site, and the *in vacuo* travel time. In this case the "propagation delay" defined in the previous paragraph is the difference between the "propagation delay" defined in this paragraph for the two different sites, with the sign of the propagation delay being determined by the definition given in the previous paragraph.

The two major components of the propagation medium that affect the group delay are the earth's neutral atmosphere (or, in this thesis, just "atmosphere") and ionosphere. The effects of the ionosphere will not be considered here, since they can be removed approximately by utilizing "dual-frequency" observations. (See Herring [1983] for a detailed discussion of ionospheric effects.) The (single antenna) atmospheric propagation delay τ_a can be written in terms of electrical path lengths, and in a system of units in which the velocity of light is unity, as

$$\tau_a = \left[\int_{atm} ds\, n(s) - \int_{vac} ds \right] \Bigg|_{t_g} \tag{1.1.1}$$

Here t_g is the epoch to which the group delay is referred. The refractive index $n(s)$ is parametrized by the length s along the traveled path. The path-label *atm* indicates that the integration is performed along the actual path of the ray through the atmosphere, while *vac* indicates that the integration is performed along the path the

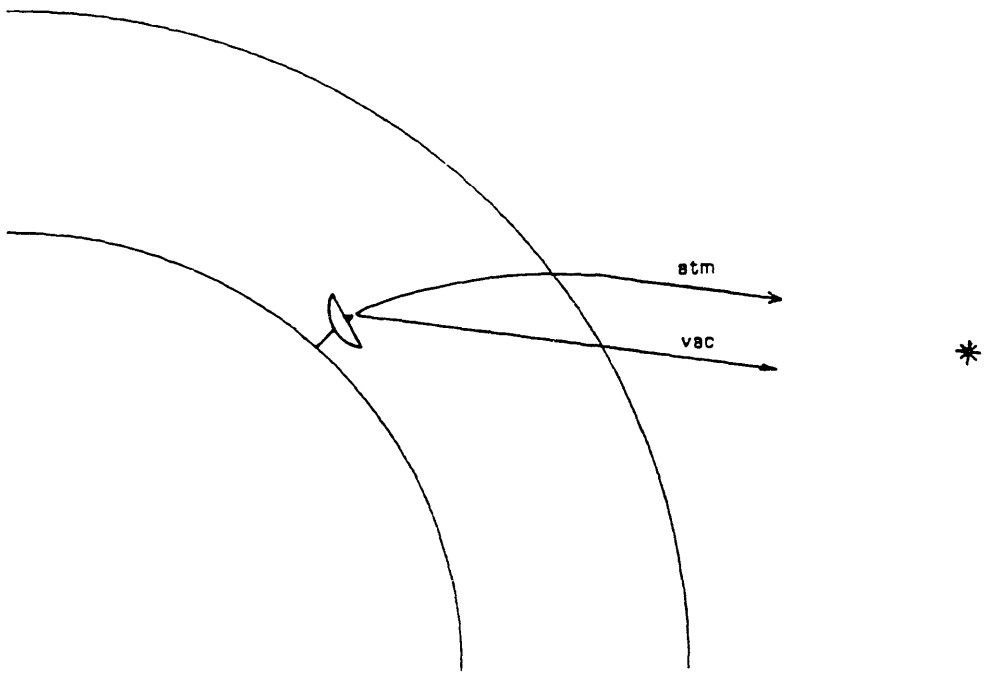


Figure 1.1.1. Illustration of the atmospheric propagation delay. The path labeled *atm* represents the path of the signal through the atmosphere. The path labeled *vac* is the path the signal would have taken were the atmosphere replaced by vacuum. This latter path is shown as a straight line connecting the receiving antenna and the source (see text).

ray would take were the atmosphere replaced by vacuum. These paths are illustrated in Figure 1.1.1. The *in vacuo* path to the radio source is shown as a straight line, while the path through the atmosphere is shown curved. (In reality, the *in vacuo* path, viewed from the surface of the earth, is not a "straight line." The model for the group delay should account for the fact that the measurements are made in a frame of reference that is not inertial, and for the fact that the space-time geometry is not Euclidean. However, since the contribution from these two effects is nearly equal for the refracted and *in vacuo* paths, the effects cancel when forming the difference of the time of propagation along these two paths.) Although (1.1.1) tells us how to calculate the delay given the path indicated by Figure 1.1.1, it does not tell us how to determine that path. The determination of the ray path will be discussed in Section 1.2.

Of what magnitude is the atmospheric propagation delay defined in (1.1.1)? For a site located near sea level at mid-latitudes, the propagation delay for a signal arriving from the zenith direction ranges typically from about 2.2–2.5 m (see below for a discussion of the units of delay), and increases with decreasing elevation angle approximately as the cosecant of the elevation angle. For sites located at a greater elevation (above sea level), the propagation delay is smaller due to the smaller air mass above the site. For example, for a site located 1 km above sea level, a typical zenith propagation delay would be about 1.9 m. For tropical sites, the propagation delay is usually greater because the air above these sites contains a large amount of water vapor. A typical zenith propagation delay above Singapore, for example, might be 2.6 m [Elgered *et al.*, 1985]. Later in this chapter we will present formulas for the propagation delay.

Because the actual path through the atmosphere is curved, the contribution of the first term on the right-hand side of (1.1.1) can be thought of as having two parts. The first—and by far the largest—contribution is due to the increased refractive index in the atmosphere, which "slows" the incoming signal traveling along the path. The second contribution is the deviation of the ray path from a straight line. These facts prompt us to write (1.1.1) as

$$\tau_a = \int_{atm} ds\ (n(s) - 1) + \left[\int_{atm} ds - \int_{vac} ds\right] \tag{1.1.2}$$

The first term on the right-hand side of (1.1.2) represents the delay along the path of the signal. The second term represents the geometrical difference between the two paths. The utility of (1.1.2) will become evident later in this chapter. In (1.1.2) we have omitted stating explicitly the epoch to which the propagation delay is referred, but now and hereafter we will implicitly assume that this epoch is the epoch to which the group delay is referred.

The previous discussion defined the atmospheric delay in terms of the VLBI group delay. This is only one of the VLBI "observables," and the most important today in geodetic work using long (> 100 km) baselines. Since the atmosphere is approximately non-dispersive below about 30 GHz (see Section 1.2.v), the atmosphere affects the phase-delay observable in approximately the same manner as the group delay, and (1.1.2) can be used to express this effect. In order to derive an expression for the effect of the atmosphere on the phase-delay rate, we need to differentiate the

time–dependent form of the propagation delay with respect to time, at the epoch to which the group delay is referred:

$$\dot{\tau}_a = \frac{d}{dt}\left[\int_{atm} ds\, n(s) - \int_{vac} ds\right]\Bigg|_{t_g} \tag{1.1.3}$$

Here $\dot{\tau}_a$ is the symbol for the atmosphere–delay rate, not the time–derivative of the propagation delay τ_a, since τ_a is referred to t_g and is not time–dependent. Henceforth in this thesis, unless explicitly stated, we will deal only with the effects of the atmosphere on the group delay and omit discussion of the phase-delay rate.

1.1.ii A note on units and terminology

The delay equation (1.1.2) is written in a system of units in which the speed of light is unity. Thus, we use the symbol "τ," which is ordinarily reserved for quantities of time. At other times in this thesis, the symbol "L" is used to represent delay. Units of length or time are used interchangeably in this thesis to express values for the delay (with the rate of exchange given by 1 nanosec $\simeq$ 30 cm). Whenever the units for an equation expressing delay are not specified, then either units of length or time may be used.

The exact end-point for the termination of the path integrals in (1.1.2) has not been specified. This topic is discussed in Appendix A, which details the development of a new formula for the propagation delay. In the body of this thesis, we will refer this end-point vaguely as the "site" or "base of the vertical column," depending on the context. The use of "site" refers to the fact that the end-point is somewhere near the location of the radio-antenna forming one element of an interferometer (or of several

interferometers). The use of "base of the vertical column" refers to the fact that the antenna is located within a few tens of meters of the ground. The site may or may not be near sea level.

Another term frequently used in this thesis is "ray." Hecht and Zajac [1979] define ray as "...a line drawn in space corresponding to the direction in flow of...energy." Thus for our purposes, the ray is unambiguously defined, both for the group delay and, in the absence of dispersion, the phase delay.

Finally, although the antennas used for VLBI are passive receivers, we will sometimes speak of a hypothetical ray as originating *at* the site. This terminology is convenient because the real signal travels a great distance before it encounters the atmosphere, and reversing the direction of propagation allows us to ignore propagation along that distance, which does not contribute to (1.1.2). This convenience is allowed because (1.1.2) is symmetric with respect to direction—ds is an element of path *length.*

1.2 The refraction of radio waves in the atmosphere

In this section we will discuss several aspects of the delay equation. First, we will discuss how one determines the path through the atmosphere, and introduce the so-called ray-trace equations. We will then discuss theoretical expressions for and experimental measurements of the refractive index for moist air.

1.2.i The ray-trace equations

A ray of electromagnetic radiation traveling through the atmosphere will obey Fermat's Principle, which states that the path will be such that the optical path length is an extremum. We therefore require that

$$\delta \int_{atm} ds\, n(s) = 0 \tag{1.2.1}$$

where again $n(s)$ is the refractive index parametrized by the distance s along the ray path. Knowledge of the functional form of $n(\mathbf{x})$, where $\mathbf{x}$ is the three-dimensional position vector, will allow (1.2.1) to be solved by choosing a suitable Lagrangian and solving the Euler-Lagrange equations. One may also solve (1.2.1) by the technique of ray tracing. We will discuss the ray-trace technique in Chapter 3. For now, we will concern ourselves with solving (1.2.1) for an atmosphere which is spherically symmetric about the center of the earth. In other words, we assume that the earth is a sphere, and that the properties of the atmosphere surrounding the sphere depend only on the radial distance. The explicit solution will lead to the familiar Snell's Law for spherical refraction, and to the "ray-trace" equations for a spherically symmetric atmosphere. (Although there is a very simple method for solving this particular problem based on Snell's Law for planar refraction, we will elicit the answer the "hard way" for illustrative purposes.)

We first write down the path-length differential ds in (1.2.1) in terms of the spherical coordinates r and ϕ defined in Figure 1.2.1:

$$ds = \sqrt{(dr)^2 + r^2(d\phi)^2} = dr\sqrt{1 + r^2(\phi')^2} \tag{1.2.2}$$

where

$$\phi' \equiv \frac{d\phi}{dr} \tag{1.2.3}$$

The third spherical coordinate does not appear in (1.2.2) because the assumption of spherical symmetry assures that the ray will always remain in the plane containing the origin of the ray, the vector parallel to the path of the ray at this point and the

origin of the coordinate system, located at the center of our spherical earth (see Figure 1.2.1).

By substituting the expression for ds from (1.2.2) into (1.2.1), we can see that an appropriate Lagrangian $\mathcal{L}$ is

$$\mathcal{L}(\phi, \phi'; r) = n(r)\sqrt{1 + r^2(\phi')^2} \tag{1.2.4}$$

The Euler-Lagrange equations then reduce to a single equation with the independent variable being the radial coordinate r

$$\frac{d}{dr}\frac{\partial \mathcal{L}}{\partial \phi'} - \frac{\partial \mathcal{L}}{\partial \phi} = 0 \tag{1.2.5}$$

Using $\mathcal{L}$ from (1.2.4) and noting that $\mathcal{L}$ does not depend on ϕ yields

$$\frac{d}{dr}\left[n(r)\frac{r^2\phi'}{\sqrt{1 + r^2(\phi')^2}}\right] = 0 \tag{1.2.6}$$

Thus the term in brackets is a constant. In order to determine this constant, let us first determine an expression for ϕ'. Figure 1.2.1 also shows the geometry which results from the ray traveling from its position at (r, ϕ) to $(r + \Delta r, \phi + \Delta\phi)$. A triangle is formed whose sides are of length r,$r + \Delta r$, and Δs, and whose corresponding opposite angles measure $\theta - \Delta\phi$, $\pi - \theta$, and $\Delta\phi$, where θ is the angle between the radial vector and the tangent to the ray at the point (r, ϕ). From the law of sines we have

$$\frac{\sin(\theta - \Delta\phi)}{r} = \frac{\sin\theta}{r + \Delta r} \tag{1.2.7}$$

Both sides of (1.2.7) can be expanded to first order in Δr and $\Delta\phi$ to yield

$$\frac{1}{r}(\sin\theta - \Delta\phi\cos\theta) \simeq \frac{1}{r}\left(1 - \frac{\Delta r}{r}\right)\sin\theta \tag{1.2.8}$$

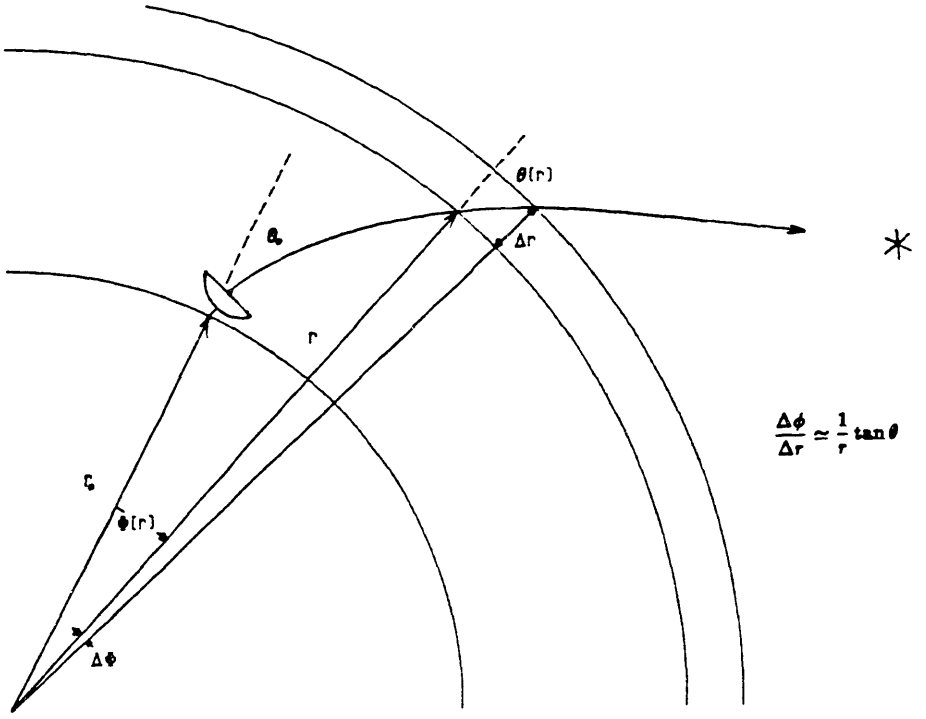


Figure 1.2.1. Geometry for the solution of the Euler-Lagrange equation for a spherically symmetric atmosphere.

Solving for $\Delta\phi$ gives

$$\Delta\phi \simeq \frac{\Delta r}{r}\tan\theta \tag{1.2.9}$$

By taking the limit as $\Delta r \to 0$, we obtain the expression for ϕ':

$$\phi' \equiv \lim_{\Delta r \to 0}\frac{\Delta\phi}{\Delta r} = \frac{1}{r}\tan\theta \tag{1.2.10}$$

Solving (1.2.10) for $\sin\theta$ gives us something which looks very much like the constant term in (1.2.6):

$$\frac{r\phi'(r)}{\sqrt{1+r^2(\phi')^2}} = \sin\theta \tag{1.2.11}$$

Substituting $\sin\theta$, then, into the constant term in (1.2.6) yields

$$n(r)r\sin\theta(r) = constant = n(r_o)r_o\sin\theta(r_o) \tag{1.2.12}$$

where r_o is a reference radius. The equation (1.2.12) is Snell's Law for spherical refraction. Using this equation we can write integral expressions for the electrical path length L_e

$$L_e = \int_{r_o}^{\infty} d\xi \frac{n^2\xi}{\sqrt{n^2\xi^2 - n_o^2 r_o^2\sin^2\theta_o}} \tag{1.2.13}$$

and the position angle $\phi(r)$

$$\phi(r) = n_o r_o \sin\theta_o \int_{r_o}^{\infty} d\xi \frac{1}{\xi\sqrt{n^2\xi^2 - n_o^2 r_o^2\sin^2\theta_o}} \tag{1.2.14}$$

where in (1.2.13) and (1.2.14) $n_o = n(r_o)$, $\theta_o = \theta(r_o)$, and we have assumed that the integration begins at r_o. The equations (1.2.13) and (1.2.14) are the ray-trace equations for a spherically symmetric atmosphere.

The delay equation (1.1.2) for a spherically symmetric atmosphere can also be written in integral form, using (1.2.13), as

$$\begin{aligned} \tau_a = & \int_{r_\circ}^{\infty} d\xi \frac{n\xi(n-1)}{\sqrt{n^2\xi^2 - n_\circ^2 r_\circ^2 \sin^2\theta_\circ}} \\ & + \int_{r_\circ}^{\infty} d\xi \left[\left(1 - \frac{n_\circ^2 r_\circ^2}{n^2\xi^2}\sin^2\theta_\circ\right)^{-\frac{1}{2}} - \left(1 - \frac{r_\circ^2}{\xi^2}\sin^2\theta_f\right)^{-\frac{1}{2}} \right] \end{aligned} \tag{1.2.15}$$

The final integral on the right–hand side of (1.2.15) is the difference in ray paths; θ_f is the "true" zenith angle. This integral converges because $n \to 1$ as $\xi \to \infty$.

The formulas for the propagation delay which are used in VLBI data analysis are based on approximations to (1.2.15). The accuracies of these approximations depend critically on the behavior of the refractive index in the atmosphere. Therefore, before presenting the formulas for the propagation delay, we will discuss the properties of the radio refractive index of moist air. The formulas based on (1.2.15) will then be presented in Section 1.4.

1.2.ii The refractivity of air

Both the general delay equation (1.1.2) and the delay equation for a spherically symmetric atmosphere (1.2.15) contain the term $n - 1$. For radio wavelengths, the refractive index n of the atmosphere differs from unity by less than one part per thousand. Because $n - 1$ is so small, it is convenient to introduce the refractivity N defined by

$$N = 10^6(n-1) \tag{1.2.16}$$

An important feature of the refractivity is that it is a quantity which for linear media depends linearly on the density of the dielectric medium, all other conditions being constant. (A linear medium is a medium for which the electric susceptibility is independent of the applied electric field. The following arguments also hold for media for which the response is for practical purposes linear due to the weakness of the applied electric field.) To see why this is so, recall that the response of a linear dielectric medium to an applied electric field $\mathbf{E}$ can be written in terms of the polarization $\mathbf{P}$ per unit volume [Jackson, 1975] as

$$\mathbf{P} = \chi \mathbf{E} \tag{1.2.17}$$

where χ is the electric susceptibility of the medium, which may vary in space. (χ is unitless and equals zero for vacuum.) Clearly χ depends linearly on the density of the medium, for if the applied electric field remains constant and the density doubles, say, then surely the dipole moment per unit volume must double. Recall that the susceptibility is related to the dielectric constant ϵ by

$$\epsilon = 1 + 4\pi\chi \tag{1.2.18}$$

where $\epsilon = 1$ for vacuum. The dielectric constant is also related to the refractive index by the Maxwell equations for propagation in a macroscopic medium [Jackson, 1975], which yield the relation

$$n^2 = \epsilon\mu \tag{1.2.19}$$

where μ is the permeability of free space ($\mu = 1$ for vacuum). For a medium, such as the atmosphere, with $\mu \simeq 1$ and $n \simeq 1$, (1.2.18) and (1.2.19) yield

$$\frac{1}{\mu}n^2 - 1 \simeq 2(n-1) \simeq 4\pi\chi \tag{1.2.20}$$

Thus the refractivity N is proportional to the electric susceptibility, and therefore also to the density. An approximate expression relating N to the density and temperature of the medium is given by Debye [1929]

$$N = \left(A + \frac{B}{T}\right)\rho \tag{1.2.21}$$

where A and B are constants which vary for different molecular species, and T is the absolute temperature. The first term on the right–hand side of (1.2.21) arises from the induced polarization (also known as the displacement polarization) of the molecules, and the second term arises from the orientation effect of the applied electric field on the permanent electric dipole of the molecules. If the molecule has no permanent electric dipole moment, $B = 0$. The derivation of (1.2.21) can be done rather simply, using a semi-classical scheme, or it can be done quantum-mechanically, which derivation is beyond the scope of this thesis. However, by far the most important feature of (1.2.21) is the dependence of N on the density. Because of this relationship, the refractivity of a mixture of q species is

$$N = \sum_{i=1}^{q}\left(A_i + \frac{B_i}{T}\right)\rho_i \tag{1.2.22}$$

where A_i and B_i are the refractivity constants for the i^{th} constituent, and ρ_i is its constituent density.

Although the atmosphere is a mixture of many gasses, it is usual to consider it as being made up of but two: the "dry" gasses, and water vapor. In the following, we show why this is possible. We will then present experimental determinations for the refractivity constants for moist air.

1.2.iii The dry atmosphere

The atmosphere of the earth is a mixture of gases and aerosols which maintain a proximity to the earth due to gravitational attraction. The major constituents of the dry atmosphere are shown in Table 1.2.1. Of considerable importance in this work is the relative number densities, or fractional volumes of the constituents of dry air. The fractional volume f_i of the i^{th} constituent in a mixture of q species is simply the ratio of the number (or number density) η_i of molecules of that species to the entire number (density) of molecules of all species:

$$f_i = \frac{\eta_i}{\sum_{k=1}^{q} \eta_k} \tag{1.2.23}$$

For an atmosphere in hydrostatic equilibrium, we would expect the heavier species to "sink" to the bottom, and if so the fractional volumes of the various atmospheric constituents would be a function of height; this will occur only if diffusion is a major source of vertical transport. Below about 90 km, however, mixing processes dominate over diffusion, so that the mixing ratios of the dry constituents can be treated as being constant below that height [Colegrove *et al.*, 1965]. (At this height the pressure is about 2×10^{-3} mbar, or about 2×10^{-6} of surface pressure, which for our purposes is to say that the atmosphere is nonexistent there.) This mixing also maintains homogeneity

over the surface of the earth. The size of the departures from homogeneity can be seen from the third column of Table 1.2.1. The entries in this column are a measure of the variability of the respective dry constituents, taken from Glueckauf [1951], expressed in terms of the fractional volume. As one can see, the variabilities of the fractional volumes are very small.

That the fractional volumes of the constituents of dry air are nearly constant is very useful. It allows us to treat the dry atmosphere as a single gas with molar mass M_d given by

$$M_d = \sum_{i=1}^{8} f_i M_i \tag{1.2.24}$$

where the sum is carried out over the eight constituents of dry air listed in Table 1.2.1 and the M_i are the molar masses, also listed in Table 1.2.1. A "sigma" for M_d, based on the variabilities given in Table 1.2.1, can be calculated approximately by ignoring the correlations between the variations in the fractional volumes. This approximation gives

$$\sigma_{M_d}^2 = \sum_{i=1}^{8} M_i^2 \sigma_i^2 \tag{1.2.25}$$

where the σ_i are the (unitless) variabilities listed in Table 1.2.1. From (1.2.24) and (1.2.25) and Table 1.2.1 we can easily compute that $M_d = 28.9644 \pm 0.0014$ kg kmol^{-1}.

In the following, we will use the fact that the fractional volumes of dry air are constant, and combine all the constituents of dry air together in creating a formula for the refractivity of moist air. We will then present experimental determinations for the values of the refractivity constants.

Table 1.2.1
Primary Constituents of Dry Air and Their Variability

Constituent	Molar Weight kg $kmol^{-1}$	Fractional Volume (Unitless)	Standard Deviation for Fractional Volume
N_2	28.0134	0.78084	0.00004
O_2	31.9988	0.209476	0.00002
Ar	39.948	0.00934	0.00001
CO_2	44.00995	0.000314	0.000010
Ne	20.183	0.00001818	0.0000004
He	4.0026	0.00000524	0.00000004
Kr	83.30	0.00000114	0.0000001
Xe	131.30	0.000000087	0.000000001

1.2.iv The refractivity of moist air

In the previous section, we saw that it is possible to treat the fractional volumes of dry air as being constant. We can now write down the expression for the refractivity of dry air. Let us write the refractivity N_d of dry air, from (1.2.22) as

$$N_d = \sum_{i=1}^{8} A_i \rho_i \tag{1.2.26}$$

We have set $B_i = 0$ for all the dry constituents since none of these constituents possesses a permanent electric dipole moment. Treating the mixture of dry gasses as a single species, we should also be able to write N_d as

$$N_d = A_d \rho_d \tag{1.2.27}$$

where ρ_d is the density of dry air (*i.e.*, the sum of the constituent densities of dry air). The constant A_d is, from (1.2.26) and (1.2.27) defined by

$$A_d = \frac{\sum_{i=1}^{8} A_i \rho_i}{\rho_d} = \sum_{i=1}^{8} A_i \frac{\rho_i}{\rho_d} = \sum_{i=1}^{8} A_i f_i \frac{M_i}{M_d} \tag{1.2.28}$$

Since M_d and the f_i are, as we have seen, constant, (1.2.28) shows that A_d is constant and that the dry constituents combined can be treated as a single gas in the refractivity formula. We may therefore write the refractivity of *moist* air as

$$N = k_1 R_d \rho_d + k_2 R_v \rho_v + k_3 R_v \frac{\rho_v}{T} \tag{1.2.29}$$

Here ρ_d is the density of the dry air mixture and ρ_v is the density of water vapor. The various A's and B's have all been expressed as k_1, k_2, and k_3. The specific gas constants R_d (for dry air) and R_v (for water vapor) need not appear explicitly in

(1.2.29), since they can be absorbed into any constant; we have included them however in keeping with the conventional values and dimensions in which k_1, k_2, and k_3 are presented. Again, note that no permanent-dipole term appears for dry air.

The values for k_1, k_2, and k_3 in (1.2.29) must be determined experimentally. (Although in principle expressions for these constant could be determined from theory, in practice the theoretical expressions themselves would contain parameters which are unknown and would have to be determined from experiment.) The method used to determine the constants is to measure the resonant frequency of a cavity into which a known amount of gas has been introduced. This frequency is compared to the resonant frequency for the evacuated cavity, and the refractivity is calculated [Boudouris, 1963; Essen and Froome, 1951] using

$$\frac{\nu_g - \nu_v}{\nu_v} = n - 1 \tag{1.2.30}$$

where ν_v is the *in vacuo* resonant frequency and ν_g is the resonant frequency with the gas. For dry air, the dependence of the refractivity on density yields k_1. For water vapor, the dependence on temperature must be used to separate k_2 and k_3.

Table 1.2.2 contains a partial list of experimental determinations of the constants in the refractivity formula. A full list of determinations dating back to 1932 appears in Bean and Dutton [1966]. The precision with which the constant k_1 (dry air) has been determined is about 0.03% of its value, while the constants k_2 and k_3 (water vapor) have been determined only to about 8–15% and 1% of their respective values. The uncertainties for the water-vapor constants reflect the fact that the determinations of

k_2 and k_3 from any single experiment are correlated. This correlation, of course, tends to "inflate" the inherent uncertainty of the determination of these constants.

Realizing that the correlation between the estimates of k_2 and k_3 was a limitation, Thayer [1974] attempted to reduce its effect by constraining the value of k_2 to its optical value. (At optical frequencies, the k_3 term does not appear and the k_2 term can therefore be determined with greater precision.) Through the breaking of the correlation, the uncertainty in the k_3 term was reduced accordingly. Thayer's values for k_2 and k_3 are also shown in Table 1.2.2. It is interesting to note that the predicted value for the radio value of k_3, based on using the optical value for k_2, is quite close to the experimental determinations of k_3 in the radio range. While this result does not prove that extrapolation from the optical range to the radio range for k_2 is correct, it does mean that Thayer's values are probably no worse than the experimental determinations.

Hill *et al.* [1982] have criticized Thayer's extrapolation from the optical from a theoretical view. They claim that this extrapolation ignores the contribution of the infrared (vibrational) resonances. Based on this criticism, Hill *et al.* use tables of the infrared spectra of water vapor to calculate this contribution; these calculated values for k_2 and k_3 are shown along with the others in Table 1.2.2. The "uncertainties" associated with the values from Hill *et al.* reflect the fit of the theoretical model expressed by (1.3.21) to the calculated values, as reported by Hill *et al.*. It can be seen that these values for the refractivity coefficients are significantly different (in terms of the experimental uncertainties) from the experimental values. Hill *et al.* are unable to explain definitively this difference and recommend use of one of the set of experimental

Table 1.2.2

Determinations of Refractivity Constants

Reference	Frequency GHz	k_1 K mbar^{-1}	k_2 K mbar^{-1}	k_3 10^5 K^2 mbar^{-1}
Essen & Froome [1951]	24	77.68 ± 0.03	64.7	3.72
Birnbaum & Chatterjee [1952]	9–25	—	71.4 ± 5.8	3.747 ± 0.03
Boudouris [1963]	7–12	77.64 ± 0.08	72.0 ± 10.5	3.75 ± 0.03
Thayer [1974]	< 20	77.60 ± 0.02	64.8 ± 0.1	3.776 ± 0.004
Hill *et al.* [1982]	—	—	98.1 ± 1	3.583 ± 0.003

values. This conflict will probably remain with us until more accurate studies of the refractivity of water vapor at long wavelengths are undertaken.

1.2.v Dispersion in the atmosphere

The formula for the refractivity of moist air given in (1.2.29) contains no dependence on frequency. A formula for the refractivity which includes dispersion can be written as [Liebe, 1985]

$$N(\nu) = N_o + N_{O_2}(\nu) + N_v(\nu) + N^c_{O_2}(\nu) + N^c_v(\nu) + N^c_\ell(\nu) \qquad (1.2.31)$$

In the above equation, N_o is the nondispersive term of the refractivity, given by (1.2.29). The terms $N_{O_2}(\nu)$ and $N_v(\nu)$ are the contributions of the anomalous dispersion due to the oxygen and water-vapor lines, respectively. The terms $N^c_{O_2}(\nu)$, $N^c_v(\nu)$, and $N^c_\ell(\nu)$ are the contributions of the continuum dispersion of dry air, water vapor, and liquid water, respectively.

Formulas for each of the dispersion terms can be found in the Millimeter-wave Propagation Model (MPM), outlined in Liebe [1985]. Figures 1.2.3–1.2.6 show the individual contributions to the dispersion from the MPM , in the range 0–100 GHz and for a total pressure of 1013.25 mbar (1 atm), a temperature of 300 K, and a relative humidity of 50%. Figure 1.2.7 shows the total refractivity for these parameters, and for a liquid density of zero. The largest dispersion in this range is due to the 60 GHz oxygen line, and deviates from the zero-dispersion value by about 1.5 N-units, or about 0.5% of the nondispersive refractivity. Below 30 GHz, the dispersion is less than 0.04%

of the nondispersive value, and is due not only to the water-vapor resonance, but also to the tail of the 60 GHz oxygen resonance and the water-vapor continuum dispersion.

Although the dispersion below 30 GHz is small, it is not negligible with respect to the uncertainty with which the refractivity is known (see Section 1.2.*iv*). The dry refractivity constant k_1 is known to about 0.03% of its value. We will continue to use the nondispersive formula for the refractivity, with the understanding that it may need appending as per the MPM.

Figure 1.2.8 shows the dispersion due to liquid water from the MPM, for a liquid water density of 0.5 g cm^{-3} and a temperature of 280 K. This density is typical for clouds of thickness between 200 and 600 m [Decker *et al.*, 1978]. No existing formulas for the atmospheric propagation delay includes the effects of liquid water, since it is difficult to estimate the amount of liquid water except by remote sensing. From Figure 1.2.8, however, we can see that the zenith propagation delay through a cloud with this temperature and liquid density and a thickness of 500 m is approximately 3.5 mm for a frequency of 8 GHz. For a slant path of 10° elevation, the delay through such a cloud is approximately 2 cm. If we compare this delay to typical group-delay uncertainties of about 1 cm, we can see that this delay is not in general negligible. We will discuss this topic further in Chapter 5.

We have now completed our review of the properties of the refractive index at microwave frequencies, and turn to the presentation of formulas for the propagation delay, and the effects of errors in these formulas on the estimation of geodetic parameters from VLBI data. However, preliminary to this presentation, we will review a number of definitions which are frequently used in discussing models of the propagation delay.

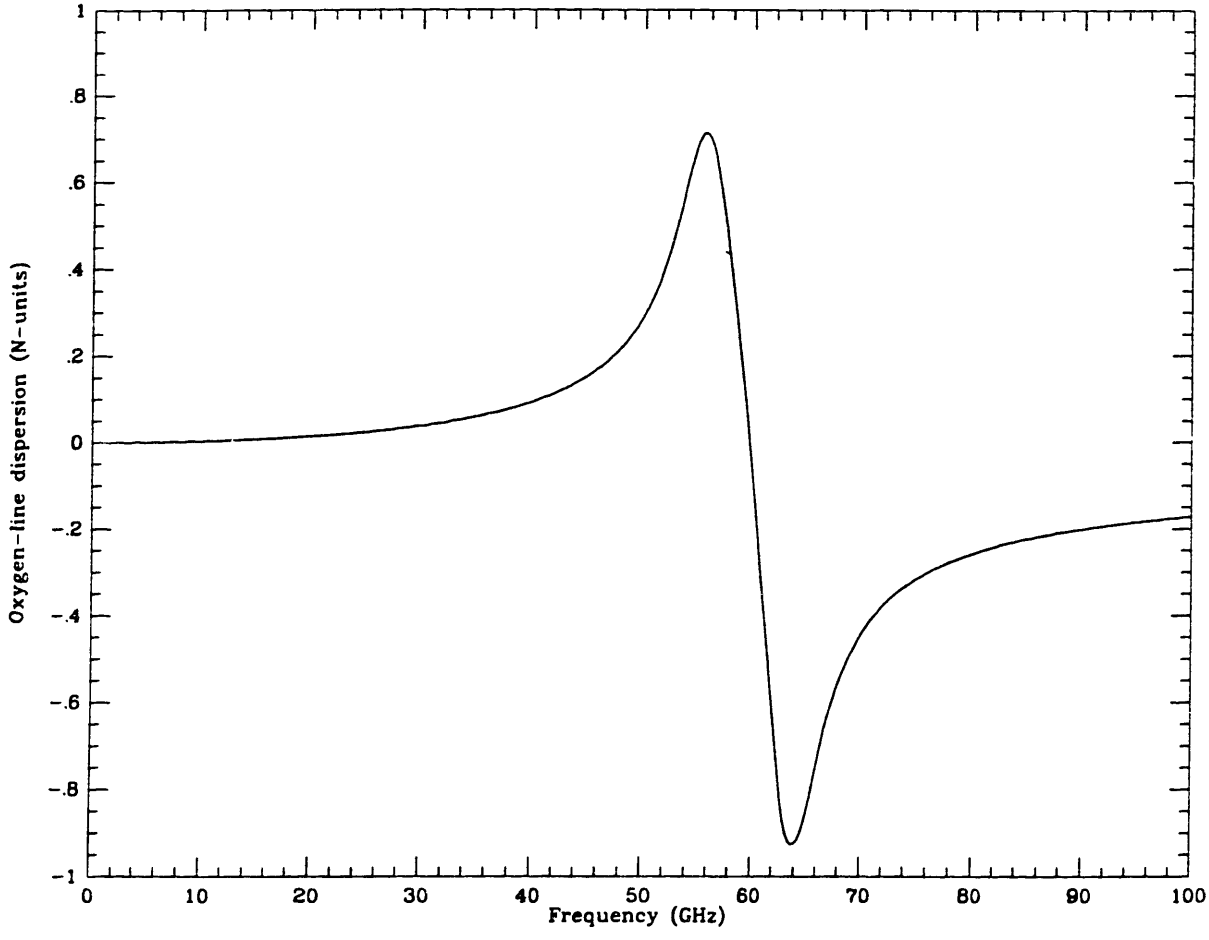


Figure 1.2.3. Contribution to the dispersive refractivity of O_2 resonances, primarily those near 60 GHz, calculated from the MPM (see text). The values of the relevant thermodynamic parameters were pressure $P = 1013.25$ mbar and temperature $T = 300$ K.

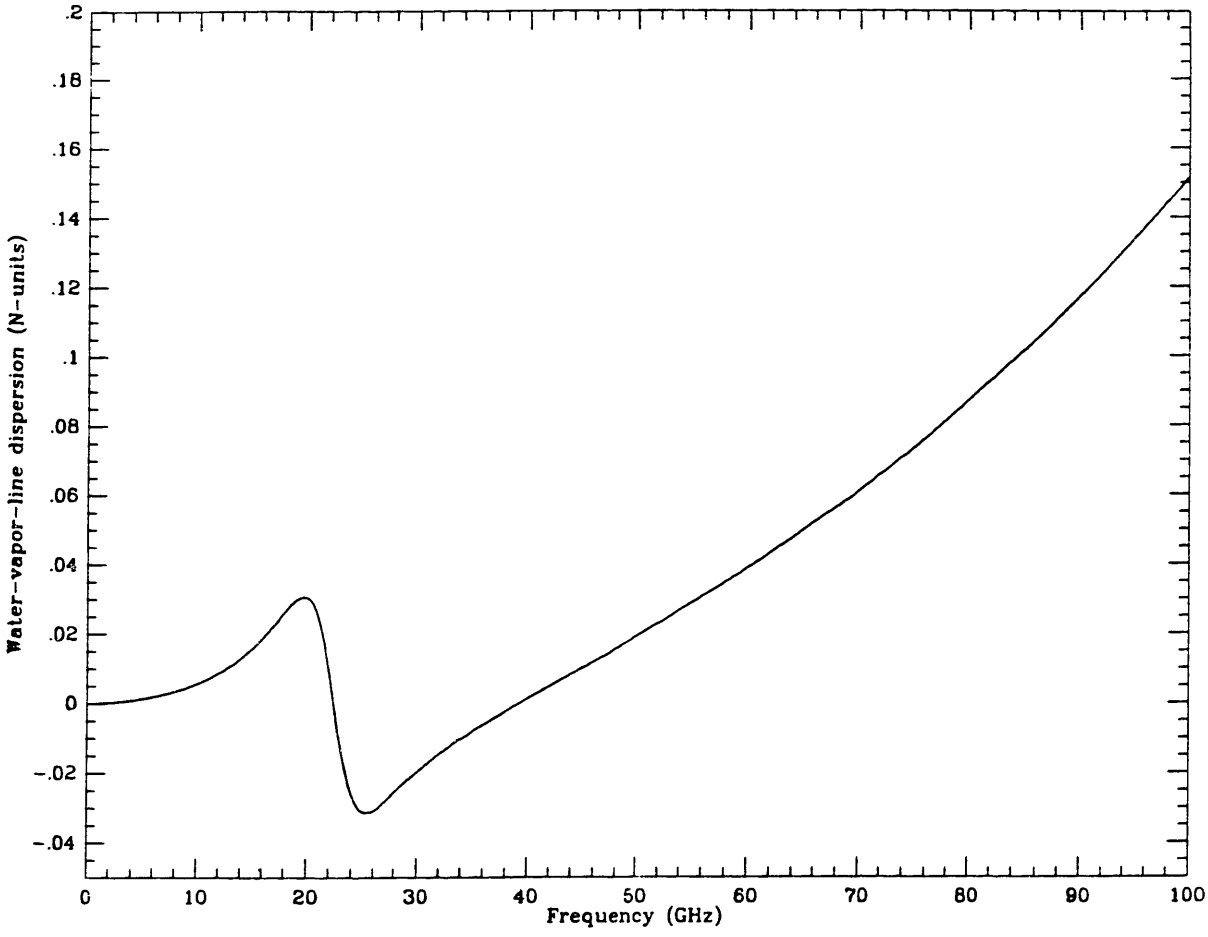


Figure 1.2.4. Contribution to the dispersive refractivity of H_2O resonances, calculated from the MPM (see text). The values of the relevant thermodynamic parameters were pressure $P = 1013.25$ mbar and temperature $T = 300$ K, and relative humidity $RH = 50\%$.

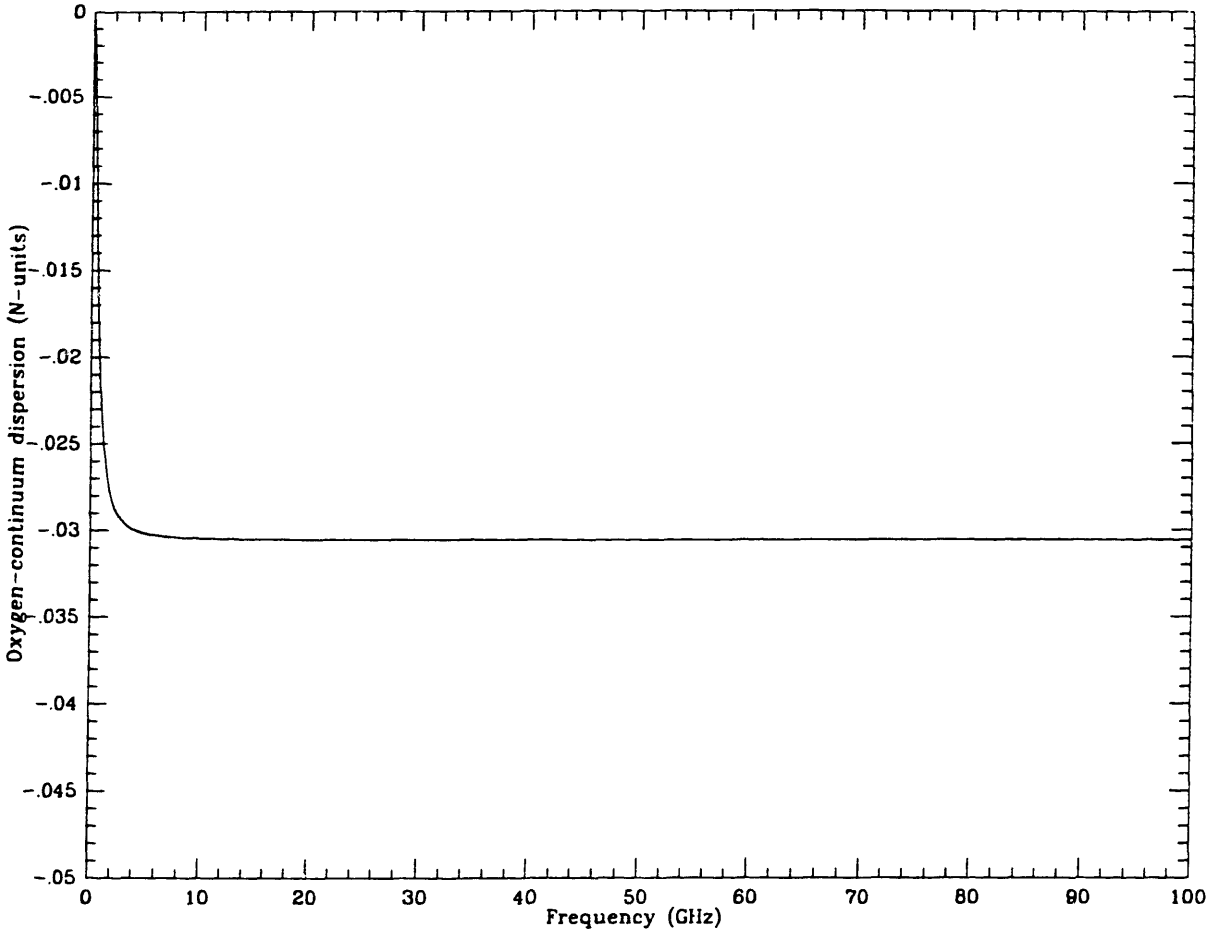


Figure 1.2.5. Contribution to the dispersive refractivity of the O_2 continuum, calculated from the MPM (see text). The values of the relevant thermodynamic parameters were pressure $P = 1013.25$ mbar and temperature $T = 300$ K.

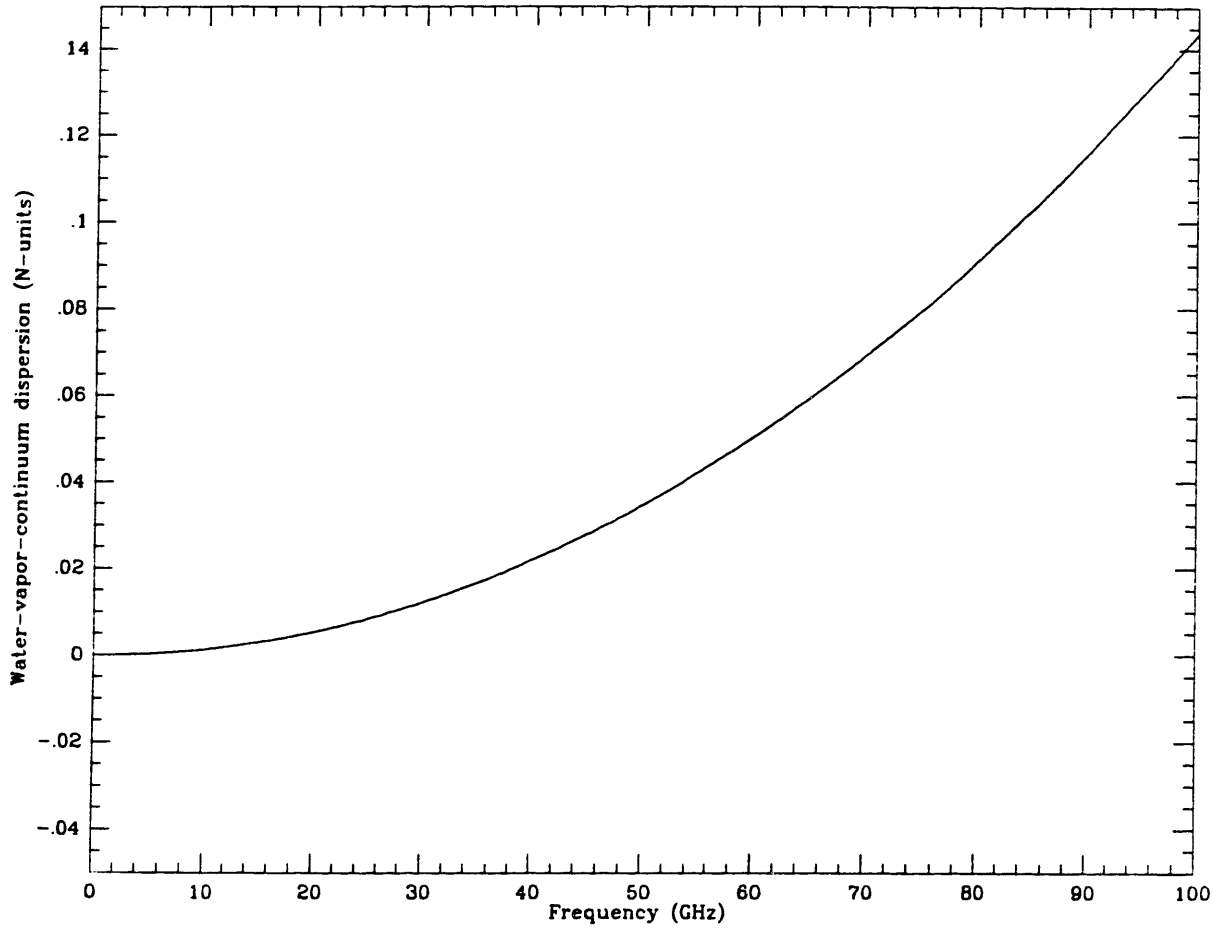


Figure 1.2.6. Contribution to the dispersive refractivity of H_2O continuum, calculated from the MPM (see text). The values of the relevant thermodynamic parameters were pressure $P = 1013.25$ mbar, temperature $T = 300$ K, and relative humidity $RH = 50\%$.

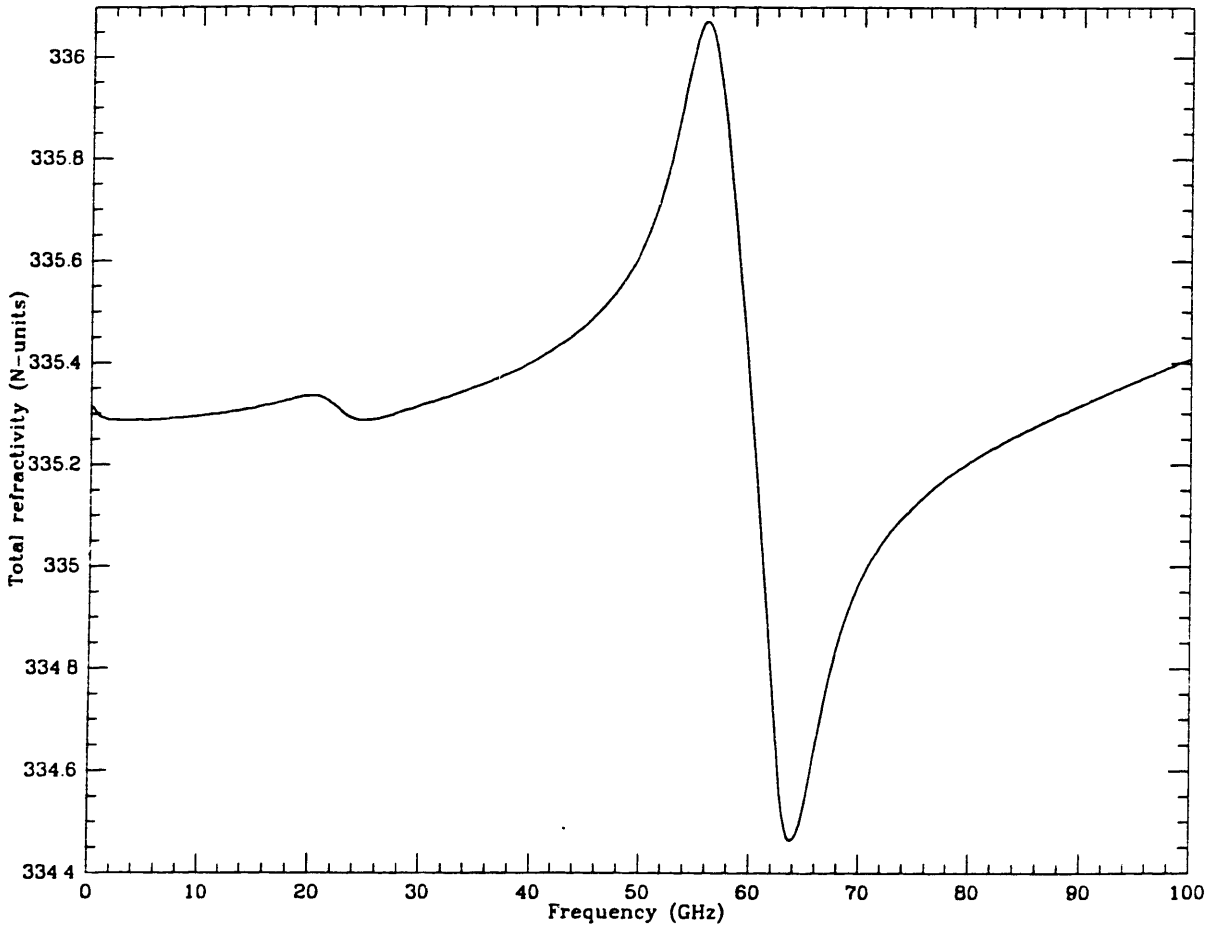


Figure 1.2.7. Total refractivity, from the the MPM (see text). The thermodynamic parameters used were those used for Figures 1.2.3–1.2.6.

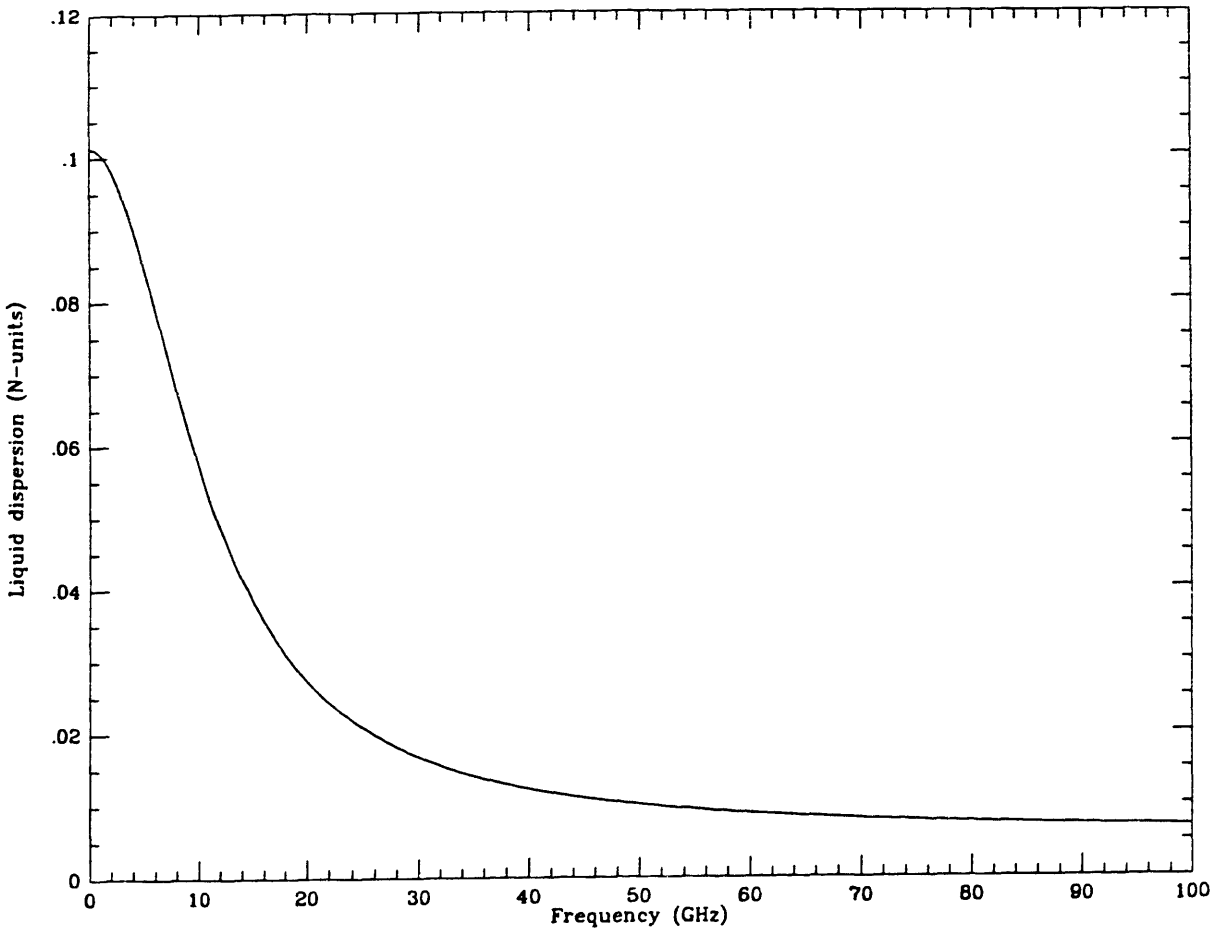


Figure 1.2.8. Contribution to the dispersive refractivity of the liquid water continuum, calculated from the MPM (see text). (This contribution is not included in the calculation for Figure 1.2.7.) Cloud-like conditions of $T = 280$ K and $\rho_\ell = 0.5$ g cm^{-1} were used.

1.3 Modeling the atmospheric delay: some definitions

In this section, we will define several terms which are often used when modeling the atmospheric propagation delay. These are *zenith delay*, *mapping function*, and *wet* and *dry delays*.

1.3.i The zenith delay

The *zenith delay* is defined to be the atmospheric propagation delay (1.1.2) for a signal arriving from the zenith direction. Although in general the path of the ray arriving from this direction may be curved, for a spherically symmetric atmosphere we find that the ray path is a straight line, because the path strikes the lines of equal refractive indices normally. We also obtain this result immediately from (1.2.14), which is the equation for the position angle. The delay equation (1.2.15) for a spherically symmetric atmosphere reduces to

$$\tau_a = \int_{r_o}^{\infty} dr' \ (n(r') - 1) \tag{1.3.1}$$

This equation can be rewritten in terms of the height z above the earth, $z = r' - r_o$, as

$$\tau_a^z = \int_0^{\infty} dz \ (n(z) - 1) = 10^{-6} \int_0^{\infty} dz \ N(z) \tag{1.3.2}$$

where $N(z)$ is the refractivity defined in (1.2.16). The superscript z indicates the zenith delay. Even though (1.2.15) is impossible to integrate analytically, as noted below this equation, it *is* possible to find analytical, closed–form solutions for the integral for the zenith delay.

1.3.ii The mapping function

In the section defining the zenith delay we saw that the integral solution for the delay (1.2.15) had a relatively simple form for the delay in the zenith direction ($\theta_\circ = 0$). The only other realistic form for the refractive index which has such a simple solution is $N(z) = constant$. Then again there is no bending, and the delay as a function of elevation ϵ can be written in terms of the zenith delay τ_a^z, and for a plane-parallel earth, as

$$\tau_a(\epsilon) = \tau_a^z \csc \epsilon \tag{1.3.3}$$

In (1.3.3), the function $\csc \epsilon$ is known as the "mapping function" because it relates the zenith delay to the delay at all other elevations. In this example, the mapping function contains no dependence on azimuth because of the assumption of spherical symmetry. The $\csc \epsilon$ mapping function is usually referred to simply as the "cosecant law." The earth's atmosphere is described by the cosecant law to some approximation. In fact, (1.3.3) motivates us to write the atmospheric delay as

$$\tau_a(\epsilon) = \tau_a^z \, m(\epsilon) \tag{1.3.4}$$

where $m(\epsilon)$ is the true mapping function, which as we noted is only approximately described by the cosecant law. Because overall the refractive index decreases with height, and because the earth is approximately spherical, the true mapping function will nearly always be less than the cosecant of ϵ, for all ϵ. (The exceptions to this rule result from horizontal variations of the refractive index.)

It is important to remember that (1.3.4) *defines* the mapping function $m(\epsilon)$. Other kinds of mapping functions can be defined. For instance, the grouping of the

right-hand side of (1.1.2) into two terms might motivate us to define two functions: the mapping function, which scales the zenith delay, and an additive "geometric delay" τ_a^g

$$\tau_a(\epsilon) = \tau_a^z \cdot m(\epsilon) + \tau_a^g(\epsilon) \tag{1.3.5}$$

Such a form has been suggested, for example, by Elgered and Lundqvist [1984]. Other definitions are, of course, possible. The form one uses depends on the manner in which one attempts to find a formula for the mapping function, and the manner in which one will ultimately use the mapping function. For example, as discussed in Chapter 2, an instrument known as a water-vapor radiometer (WVR) is in principle able to determine the "wet" path delay (which will be defined below) along the line–of–sight. Thus, a model combining the "dry" delay and the "wet" delay determined from WVR data would have the form

$$\tau_a(\epsilon) = \tau_d^z \cdot m(\epsilon) + \tau_{WVR} \tag{1.3.6}$$

where τ_d^z is the zenith "dry" delay and τ_{WVR} is the line-of-sight wet path delay determined by the WVR.

In the next section, we will describe in more detail what we mean by "wet" and "dry" delays. The definitions for these terms are important, for as we have seen these terms also define the mapping function.

1.3.iii The wet and dry delays

Terms like "wet delay" and "dry delay" are often used, but are seldom defined carefully, and frequently not at all. This state of affairs is unfortunate, since there is no obvious single way to define these terms. There are many ways to separate the

different terms for both the refractivity and the delay. We will give the definitions which will be used in this thesis, but they are not universal. We believe, however, that they are "best" in a sense that will be discussed below.

The definitions of "wet" and "dry" refractivities are unambiguous, because it is possible to write an expression for the refractivities of dry air and of water vapor. In terms of the constants k_1, k_2, and k_3 introduced in Section 1.2.*iv*, we have

$$N_d = k_1 R_d \rho_d \tag{1.3.7}$$

$$N_v = k_2 R_v \rho_v + k_3 R_v \frac{\rho_v}{T} \tag{1.3.8}$$

Let us examine the zenith delay. Equation (1.3.2) can now be written

$$\tau_a^z = 10^{-6} \int_0^\infty dz\, N_d + 10^{-6} \int_0^\infty dz\, N_w \tag{1.3.9}$$

It seems obvious to call the first term on the right-hand side of (1.3.9) the dry zenith delay, and the second term the wet zenith delay. These definitions become inconvenient, however, when one attempts to derive formulas for the zenith delay. The reason for this inconvenience is that it is necessary to know the profile of the density of dry air for the integration of the dry refractivity, and it is also necessary to know the profile of the density of water vapor for the integration of the wet refractivity. On the other hand, it is possible to estimate the integral of the *total* density without knowing its specific profile. This ability stems from the equation of hydrostatic equilibrium, from which we obtain

$$P_o = \int_0^\infty dz\, \rho(z)\, g(z) \tag{1.3.10}$$

where $P_\circ$ is the pressure at the base of the vertical column, $\rho(z)$ the total density (specified here and hereafter by the omission of any subscripts), and $g(z)$ is the acceleration due to gravity. Because gravity varies only slightly over the effective range of the integration in (1.3.10), we can expand $g(z)$ to first–order in z with negligible error, and write (1.3.10) as

$$P_\circ = g_\circ \int_0^\infty dz\, \rho(z) + g'_\circ \int_0^\infty dz\, z\, \rho(z) \tag{1.3.11}$$

where $g'_\circ$ is the derivative of g with respect to height evaluated at the surface. We can rewrite the second integral in terms of the altitude H_c of the center of mass of the vertical column, which by definition is

$$H_c \equiv \frac{\int_0^\infty dz\, z\, \rho(z)}{\int_0^\infty dz\, \rho(z)} \tag{1.3.12}$$

Substitution of (1.3.12) into (1.3.11) yields

$$P_\circ = (g_\circ + g'_\circ H_c) \int_0^\infty dz\, \rho(z) \tag{1.3.13}$$

Although we have not examined the accuracy of (1.3.13)—this is discussed in Appendix A—the implication of this equation is clear: it is possible to estimate the integral of the total density using the pressure at the base of the vertical column. This result prompts us to rearrange (1.2.29) in terms of the total density, and the remaining water vapor terms:

$$N = k_1 R_d \rho + (k_2 - \frac{M_v}{M_d} k_1) R_v \rho_v + k_3 R_v \frac{\rho_v}{T} \tag{1.3.14}$$

We will now *define* the "dry" zenith delay to be

$$\begin{aligned} \tau_d^z &= 10^{-6} k_1 R_d \int_0^\infty dz\, \rho(z) \\ &= 10^{-6} k_1 R_d (g_\circ + g'_\circ H_c)^{-1} P_\circ \end{aligned} \tag{1.3.15}$$

and the "wet" zenith delay to be

$$\tau_w^z = 10^{-6} R_v \int_0^\infty dz \left[k_2 - \frac{M_v}{M_d} k_1 + \frac{k_3}{T(z)} \right] \rho_v \tag{1.3.16}$$

Several things should be pointed out about these definitions. Firstly, they are just that: definitions. For other applications, it might be convenient to have other definitions. One might disagree with the term "dry" for the quantity defined in (1.3.15) because it depends on the *total* density. However, this term is a "dry" zenith delay in the sense that if it were known that the atmosphere contained no water vapor, τ_d^z would still be defined by (1.3.15), because it is "parametrized" by the total pressure. It is not a "dry" zenith delay in the sense that the mean molar mass of the atmospheric gas, the density of which is ρ, is not equal to the molar mass of dry air.

Let us now move on to the propagation delay for directions other than the zenith. In terms of the refractivity, we can write the delay equation (1.1.2) as

$$\tau_a = 10^{-6} \int_{atm} ds\, N(s) + \left[\int_{atm} ds - \int_{vac} ds \right] \tag{1.3.17}$$

If we express the refractivity not in terms of the wet and dry refractivity, but in terms of the total density and water-vapor density as in (1.3.15) and (1.3.16), (1.3.17) becomes

$$\begin{aligned} \tau_a &= 10^{-6} k_1 R_d \int_{atm} ds\, \rho(s) \\ &\quad + 10^{-6} R_v \int_{atm} ds \left[k_2 - \frac{M_v}{M_d} k_1 + \frac{k_3}{T(s)} \right] \rho_v \\ &\quad + \tau_a^g \end{aligned} \tag{1.3.18}$$

where the "geometric propagation-delay" τ_a^g is given by

$$\tau_a^g = \int_{atm} ds - \int_{vac} ds \tag{1.3.19}$$

The meaning of τ_a^g was discussed in Section 1.1.1. If we are to divide the delay into a "wet" delay and a "dry" delay, what are we to do with the geometric propagation delay? Elgered and Lundqvist [1984] have suggested that the propagation delay be made up of three terms: the dry and wet delays, and the geometric delay as in (1.3.6). In fact, the grouping of the propagation delay is primarily a matter of taste. It is clear, however, that the definition of the "dry" delay being proportional to the integral of the total density has the advantage of being able to integrate the dry delay "exactly" as already discussed.

In the following section, we will employ the definitions presented in this section, when we discuss the formulas for the propagation delay used in VLBI data analysis and the effects of errors in those formulas.

1.4 Treatment of the atmosphere in VLBI data analysis. I

In this section we will present the formulas commonly used to model the atmospheric propagation delay in geodetic VLBI data analysis. We will begin in Section 1.4.*i* by discussing the possible effect of the atmosphere on the estimation of parameters from VLBI data. In Section 1.4.*ii* we will present common formulas used to model the propagation delay. Then, in Section 1.5, we will discuss the effects on parameter estimates of errors in these formulas for the propagation delay. Finally, in Section 1.6,

we will again discuss the treatment of the atmosphere in VLBI data analysis, this time in light of the results of Section 1.5.

1.4.i Parameter estimation from VLBI data

In this section we will review estimation of parameters from "geodetic" VLBI data, (*i.e.*, data composed of VLBI group delays). The usual means of estimating such parameters is by weighted least squares. This estimation involves comparing the "measured" group delay to an *a priori* estimate of that delay. The *a priori* estimate of the delay is calculated from a model which uses *a priori* estimates of some set of parameters; the differences between the observed group delays and the *a priori* estimates are interpreted as being due to differences between the true values of the parameters and the *a priori* values, plus measurement noise.

What effect does the atmosphere have on this estimation procedure? In Section 1.1.i, we presented the effect of the atmospheric propagation delay on the "observed" group delay. If unaccounted for in the mathematical model for the group delay, this propagation delay will contribute to the difference between the observed and calculated group delays. As we already discussed, the magnitude of the propagation delay (single site) ranges from about 1.9–2.5 m in the zenith direction (depending on site location and meteorological conditions), and increases approximately as the cosecant of the elevation angle, yielding a 20–28 m delay at 5° elevation. If the antennas at both sites of an interferometer observed that low, then the total atmospheric delay (*i.e.*, two-site difference) could range from about -26 to +26 m.

Is the atmospheric propagation delay significant with respect to the precision with which we can "measure" the group delay? Obviously, if the delay itself were not significant, then its effects on the estimates of parameters would be insignificant, and we would not have to worry about it. However, a typical uncertainty for a group delay "measurement" is approximately 1–2 cm, based on the signal-to-noise ratio (SNR). Thus, the atmospheric delay represents a potential 2000-σ error, if we observe at an elevation of 5°.

Ideally, we would like to have a "perfect" estimate of the propagation delay, which we would then incorporate into our model for the group delay, and hence into our *a priori* estimate of the group delay. Things are not ideal, however, and so the next best thing is to have an estimate for the propagation delay which is "insignificantly different" from the true propagation delay. By "insignificantly different" we mean small with respect to the uncertainty of the measurement. (Note that errors can be made insignificant by increasing the number used for the uncertainty of the measurement: see Section 1.6.*i*.) As above, we assume that insignificant errors influence the estimates of parameters insignificantly. This assumption is not always valid if the errors are of a systematic nature.

Another option is to develop a mathematical expression for the propagation delay which can be parametrized, and the atmospheric parameters can be estimated along with the other parameters. Nevertheless, if the mathematical expression for the propagation delay is incorrect, systematic errors may result. Even if this expression is correct, the estimation of atmospheric parameters will have an influence on the estimates of other parameters. (This influence is not specific to atmospheric parameters.

The introduction of *any* parameters into a least-squares solution will affect the values of the original parameter estimates, although the magnitude of the effect depends on the correlations between the "signature" of the new parameter and those of the old parameters.)

In the following section, we will discuss how the propagation delay has typically been modeled, and how atmospheric parameters have been incorporated into the least-squares solutions. We will then discuss the effects of errors in the model for the propagation delay on the values of estimated parameters. We will also discuss the effects of including these atmospheric parameters.

1.4.ii Modeling the propagation delay

In this section, we will review the way in which the atmospheric delay has been modeled. In doing so, we will make use of the definitions of Section 1.3. The method which we will first discuss is outlined in Clark *et al.* [1985], and is implemented in the CALC-SOLVE suite of programs.

The formula used for the zenith delay is that developed by Saastamoinen [1972]. It relates the pressure, temperature, and humidity at the site, as well as the latitude and altitude of the site, to the propagation delay in the zenith direction:

$$\tau_a^z = 0.002277 \left[P_\circ + \left(\frac{1255}{T_\circ} + 0.05 \right) e_\circ \right] \left(f\left(\phi, H\right) \right)^{-1} \qquad (1.4.1)$$

where τ_a^z is the zenith delay in meters, $P_\circ$ is the total surface pressure in mbars, $e_\circ$ is the partial pressure of water vapor at the surface, in mbars, $T_\circ$ is the absolute

temperature in Kelvins, and $f(\phi, H)$ is a function which accounts for the variation of gravity with the latitude ϕ and altitude H of the site, and is given by

$$f(\phi, H) = 1 - 0.0026(1 - 2\sin^2\phi) - 0.00031H \tag{1.4.2}$$

where H is in km. The term proportional to the total pressure we have discussed; it is usually called the zenith "dry" delay. The term proportional to e_o is called the zenith "wet" delay. Although, as we have discussed, it is very difficult to determine the wet delay based only on surface measurements, Saastamoinen used an "average" exponential vertical profile of humidity to determine the wet-delay constants in (1.4.1).

The Saastamoinen zenith delay formula must be combined with some mapping function in order to yield a formula for the propagation delay in other directions. Two mapping functions figure prominently in geodetic VLBI data analysis: the Chao mapping function [Chao, 1972] and the Marini mapping function [Marini, 1972; Marini and Murray, 1974]. We will give the Marini mapping function here because it is not readily available, having been given originally in an unpublished internal memorandum. The Marini formula is based on a continued-fraction expansion for the propagation delay originally developed by Marini [1972]. This formula yields for the propagation delay $\tau_a(\epsilon)$ for a signal arriving from the elevation angle ϵ

$$\tau_a(\epsilon) = \frac{A + B}{\sin\epsilon + \dfrac{[B/(A+B)]}{\sin\epsilon + 0.015}} \tag{1.4.3}$$

where A is the Saastamoinen zenith delay given by (1.4.1) and B in meters is given by

$$B = \frac{2.644 \times 10^{-3}}{f(\phi, H)} \exp(-0.14372H) \tag{1.4.4}$$

where $f(\phi, H)$ is given in (1.4.2). We have presented (1.4.3) in a slightly different form from Marini and Murray [1974] in order to emphasize the role of the Saastamoinen zenith delay. Using (1.4.3) we can define a Marini *mapping function* given by

$$m(\epsilon) = \frac{1+\kappa}{\sin\epsilon + \dfrac{[\kappa/(1+\kappa)]}{\sin\epsilon + 0.015}} \tag{1.4.5}$$

where $\kappa = B/A$. Although it might seem as though (1.4.5) is artificial, it does give us a way to combine the Marini formula with different formulas for the zenith delay.

The other formula for the propagation delay which is frequently used for analysis of VLBI data is the Chao formula [Chao, 1972]. This formula is also based on the continued–fraction form of Marini, but the mapping function contains no dependence on meteorological variables. The Chao formula for the propagation delay is

$$\tau_a(\epsilon) = \frac{\tau_d^z}{\sin\epsilon + \dfrac{A}{\tan\epsilon + B}} + \frac{\tau_w^z}{\sin\epsilon + \dfrac{C}{\tan\epsilon + D}} \tag{1.4.6}$$

In (1.4.6), τ_d^z and τ_w^z are the dry and wet zenith delays, respectively, and A, B, C, and D are dimensionless constants, whose values are

$$\begin{aligned} A &= 0.00143 \quad B = 0.0445 \\ C &= 0.00035 \quad D = 0.0170 \end{aligned} \tag{1.4.7}$$

Chao determined the values for the constants A–D by fitting (1.4.6) to values for the delay determined by ray-tracing. The refractivity profile used in this ray-trace was an "average" profile derived from radiosonde data.

Because, as we have discussed, the Saastamoinen formula for the wet delay can be expected to be in error by several centimeters, an adjustment to A in (1.4.3) might be

one of the parameters which is estimated when this formula is used for the propagation delay. However, in the SOLVE program the Chao mapping function is used to calculate the derivatives for this adjustment, whether the Chao formula or the Marini formula is used for the *a priori* estimate. The reason for this inconsistency is that the Chao formula is somewhat older than the Marini formula and was originally used to calculate the partial derivatives when only the Chao formula could be used for the *a priori* value of the propagation delay. When the Marini formula for the propagation delay was introduced, the ability to use partial derivatives based on the Marini formula was never encoded.

The Saastamoinen, Chao, and Marini formulas were the only formulas employed for the reduction of VLBI data until about 1984. The Berman "day-night" formula [Berman, 1976] was also used, but this formula differs from the Saastamoinen formula only in the expression for the wet delay. By mid–1984, two new mapping functions had been introduced. One of these new mapping functions is described in Lanyi [1984], and is based on a third-order expansion in the refractivity of the analytical expression for the propagation delay. This new mapping function has not yet been thoroughly tested using VLBI data, but preliminary tests indicate that use of this mapping function reduces elevation-angle dependent systematic behavior in the estimates of declinations of radio sources; testing of this mapping function is continuing at present [R.N. Treuhaft, Jet Propulsion Laboratory, private communication, 1984; 1986] The other was developed by us, and its development and testing is described in Chapter 3 and Appendix A.

In the following section, we discuss the effects on estimates of geodetic parameters of errors in the models for the propagation delay. In Section 1.6, we will use the results of Section 1.5 to discuss the "optimum" treatment of the atmosphere in VLBI data analysis.

1.5 Effects of atmospheric modeling errors

In this section, we discuss the effects on the estimates of geodetic parameters (in this case baseline length) of errors in the models for the propagation delay given above. We will develop a simple error model for three types of errors in the above models: (*i*) errors in the elevation-angle dependence of the mapping function, (*ii*) errors due to azimuthal asymmetry, and (*iii*) errors in the estimate of the zenith delay. We will use these error models to investigate the effects of the respective errors on estimates of site position, and from these errors infer the errors in baseline-length estimates.

1.5.i Mapping function errors

In this section we will discuss the effects of errors in the elevation-angle dependence of the mapping function. In developing an error model we will assume that the "true" mapping function can be expressed as a Chao–type mapping function (*i.e.*, a continued fraction with the second sine replaced by tangent):

$$m(\epsilon) = \frac{1}{\sin\epsilon + \dfrac{a}{\tan\epsilon}} \tag{1.5.1}$$

The mapping function (1.5.1) is identical to the mapping functions in (1.4.6) with $B = 0$. Setting $B = 0$ will not affect our results and makes for a simpler mapping

function. Instead, this mapping function is parametrized by the single constant a. An erroneous value for a will induce an error $\Delta\tau_a$ given by

$$\Delta\tau_a \simeq -\tau_a^z \frac{\cos\epsilon}{\sin^3\epsilon}\Delta a \tag{1.5.2}$$

where $\Delta a = a - a_{true}$. Equation (1.5.2) is an approximation for small a and small Δa.

How large an error does (1.5.2) yield for a mapping function error? From (1.4.7), we know that $a \simeq 0.001$. If we have an error $\Delta a = -0.1a$, and a zenith propagation delay of 230 cm, then Figure 1.5.1 shows the predicted error between elevation angles of about 5° and 20°. From this curve, we can see that a 10% error in a can yield over a 4 cm error in the value for the propagation delay at an elevation angle of 10°.

How will such an error affect the estimate of site position? In order to determine this effect, we will need an *a priori* model for the group delay–error. This model will be the one which we, in our ignorance of mapping function errors, assume to represent the difference between the observed group delay and the calculated group delay. Of course, our *a priori* model might contain the effect of an error in the mapping function, but in general it does not, for reasons which are discussed later. We will use the simplest "realistic" model:

$$\Delta\tau_g = -\Delta\mathbf{b}\cdot\hat{e} + \Delta\tau_a^z\, m(\epsilon) + \Delta C \tag{1.5.3}$$

Here $\Delta\mathbf{b}$ is an error in the *a priori* value for the baseline vector, $\hat{e}$ is the unit vector in the direction of the radio source, $\Delta\tau_a^z$ is an error in the *a priori* value of the zenith delay, and ΔC is an error in the *a priori* value of the "clock offset" (see Robertson [1975]). The system of units employed in (1.5.3) is such that the speed of light is unity (refer to Section 1.1.ii). The signs of the baseline vector and the group delay are

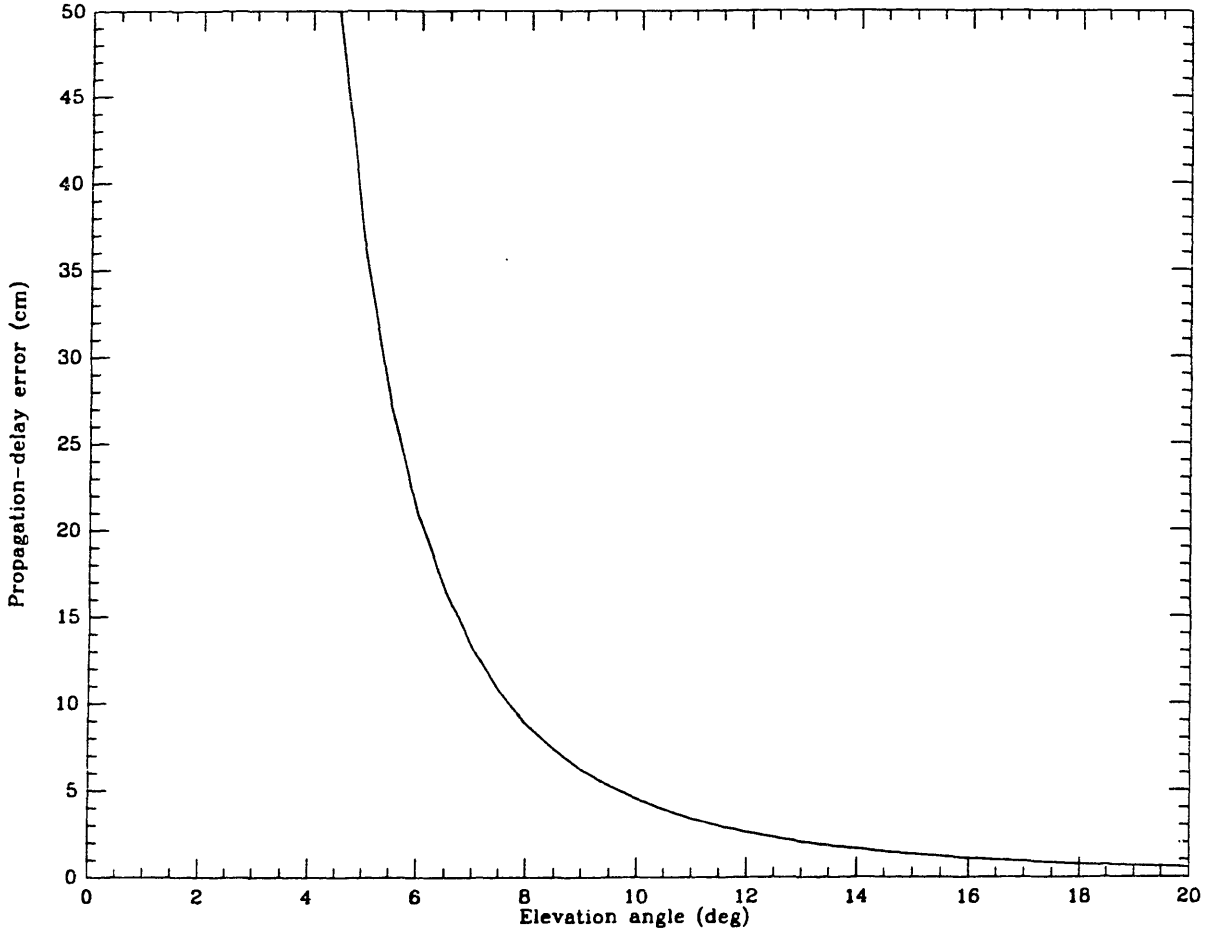


Figure 1.5.1 Error in the predicted propagation delay for a mapping-function error of $\Delta a = -10^{-4}$ and a zenith delay of 230 cm (see text).

simultaneously defined by (1.5.3), *i.e.*, if the baseline vector is defined to point from site A to site B, then the group delay is defined to be $t_B - t_A$, where t_A is the time of arrival of the signal at site A, and t_B the time of arrival of the signal at site B in the frame of reference of site A.

In our simple model,we will assume that we have data from a single baseline only. In this case we can estimate changes in the three components of the baseline vector, but we are insensitive to equal parallel displacements of the two sites. However, we are primarily interested in the effects of an atmospheric–delay error at one site on the estimates of the position of that site. It is irrelevant whether that site is the "reference" (A) or "remote" (B) site. Thus, we will arbitrarily choose the "remote" site; an error in the position of the "remote" site, expressed in a local cartesian frame, along with the clock and atmosphere errors parametrized in (1.5.3), yield for the error in the group delay

$$\Delta\tau_g = -\Delta z \sin\epsilon - \Delta x \cos\epsilon \sin\alpha - \Delta y \cos\epsilon \cos\alpha + \Delta\tau_a^z \, m(\epsilon) + \Delta C \qquad (1.5.4)$$

The local right-hand cartesian frame is defined so that the x-axis points east, the y-axis north; the azimuth α is defined in the usual manner of being measured clockwise from the north axis as viewed down along the z-axis.

In general, the analyst of VLBI data will estimate the errors in the local site coordinates $(\Delta x, \Delta y, \Delta z)$, the error in the zenith delay $\Delta\tau_a^z$, and the clock offset ΔC. In this section, however, we will assume that the "observing" schedule is sufficiently homogeneous with respect to azimuth that an error of the form (1.5.3) will induce only small errors in the local x-y plane. We will therefore estimate only Δz, $\Delta\tau_a^z$, and ΔC.

We will assume a uniform elevation coverage from the zenith direction down to some minimum elevation angle ϵ_o. In Figure 1.5.2 the solid line shows the estimate of Δz as a function of this minimum elevation angle, when the three parameters Δz, $\Delta \tau_a^z$, and ΔC are estimated simultaneously. The dotted line shows the *negative* of the estimate of Δz for the case for which no zenith delay parameter is estimated. Figure 1.5.3 shows the resulting "statistical uncertainty" for these corresponding estimates, for the assumption of 100 observations uniformly weighted with a measurement uncertainty of 1 cm. (The statistical uncertainty will scale with the measurement uncertainty and with the reciprocal of the square-root of the number of observations.) Comparison of Figures 1.5.2 and 1.5.3 illustrates a problem central to this thesis: the statistical uncertainty in the estimate of the vertical component of site position is smaller for smaller ϵ_o while the effects due to atmospheric-related systematic errors are increased. The statistical uncertainty of the estimate of the vertical coordinate is increased by the estimation simultaneously of the zenith delay parameter $\Delta \tau_a^z$. The partial derivative for this parameter looks nearly like $\csc \epsilon$, while the partial derivative for Δz looks like $\sin \epsilon$. The estimates of these parameters are therefore highly correlated. Figure 1.5.4 shows the correlation in the estimates of these parameters as a function of ϵ_o. Even for very low minimum elevation angles, the correlation is still very large.

An implication of Figure 1.5.2 is that if successive estimates of site position are made from subsets of data having increasingly larger minimum elevation angles, then the estimates of the vertical component of site position will monotonically increase or decrease, depending of course on the sign of Δa (assuming ϵ_o is below about 50°). It also tells us that we should get a systematically biased estimate for this parameter if

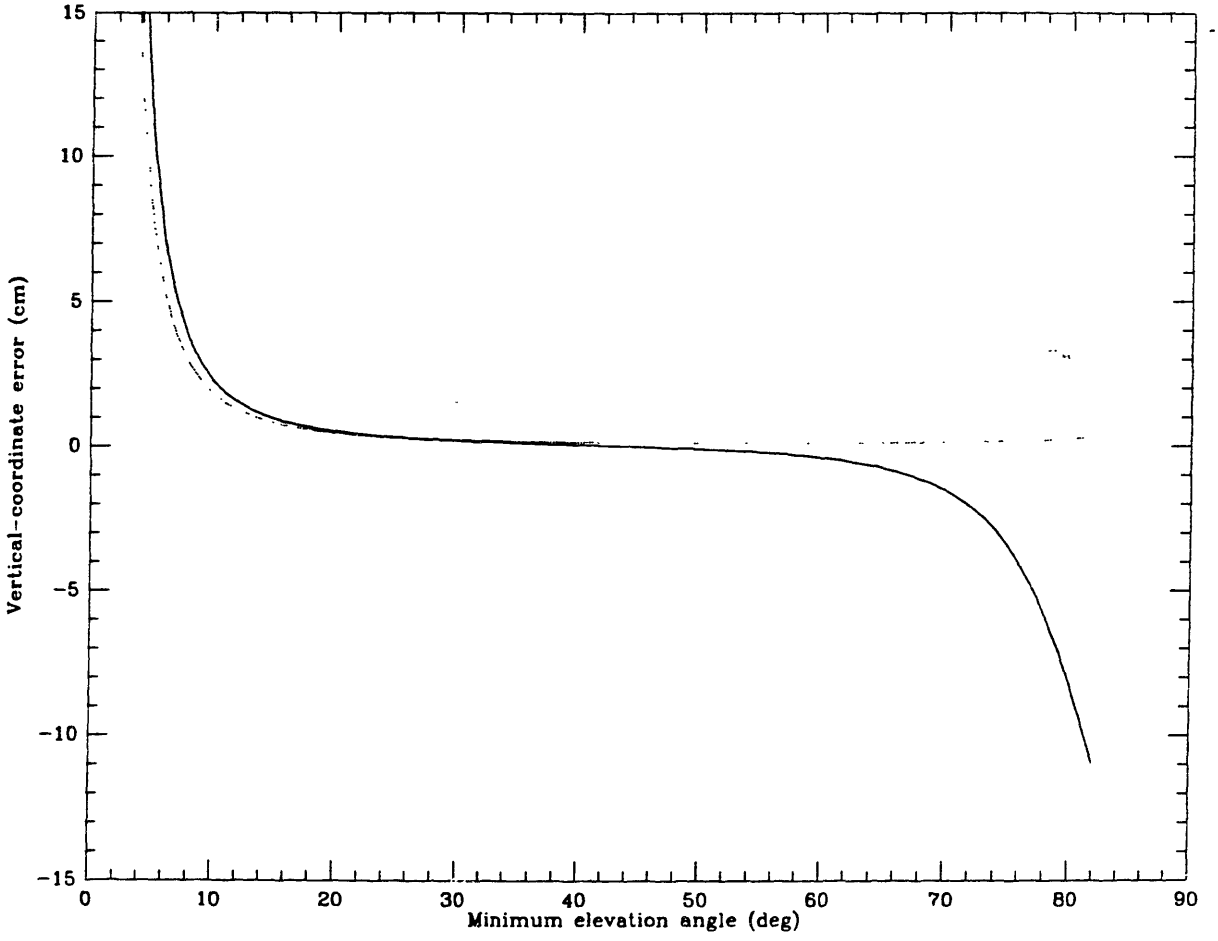


Figure 1.5.2. Error in the estimate of the vertical coordinate of site position due to the mapping-function error shown in Figure 1.5.1 (solid line). The dotted line shows the negative of the error for the case in which no zenith delay parameter is estimated. For a fuller explanation of the error model for the group delay, see text.

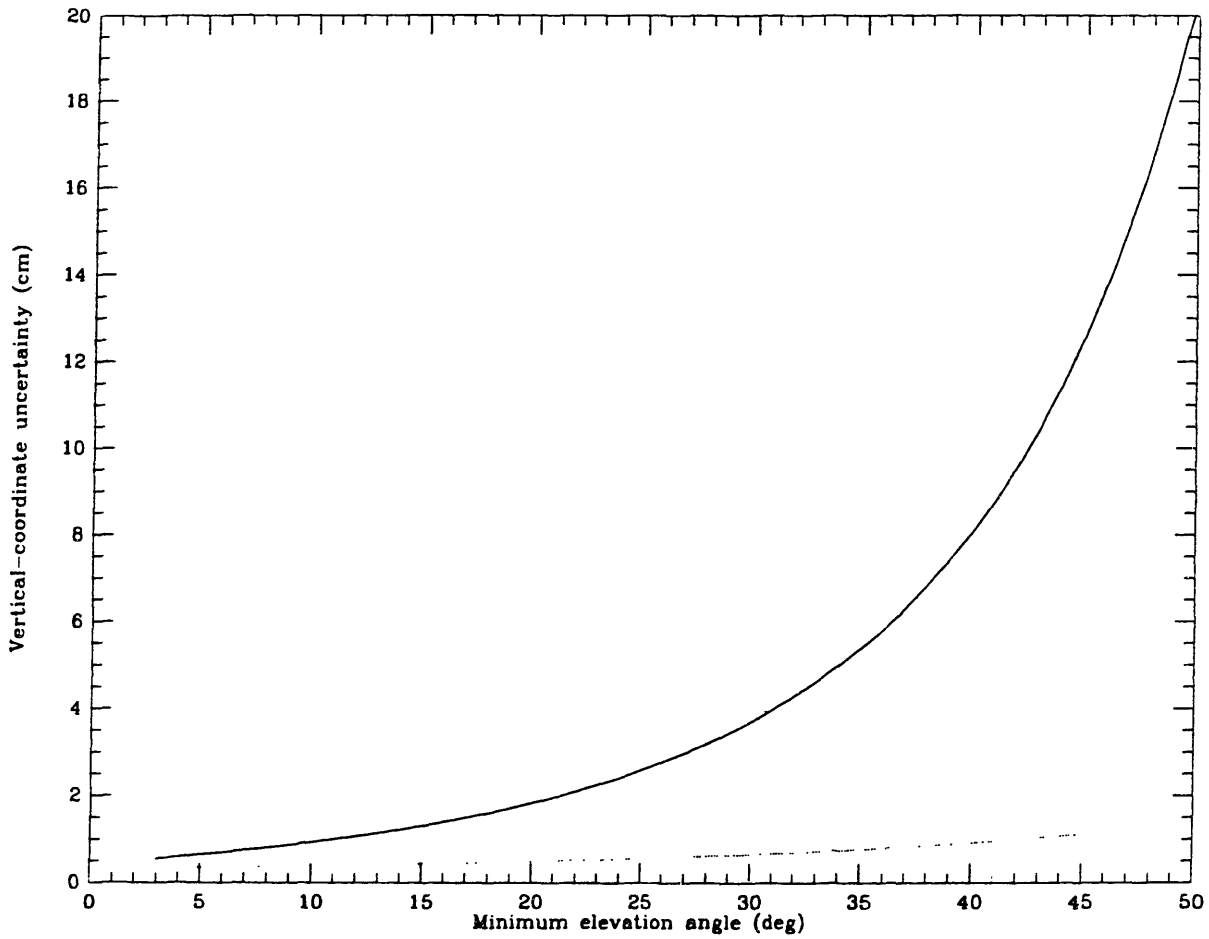


Figure 1.5.3. Statistical uncertainty of the estimate of the vertical coordinate of site position as a function of the elevation-angle lower limit (see text). Again, the solid line refers to the case for which a zenith delay parameter is estimated, while the dotted line shows the case for which no zenith delay parameter was estimated.

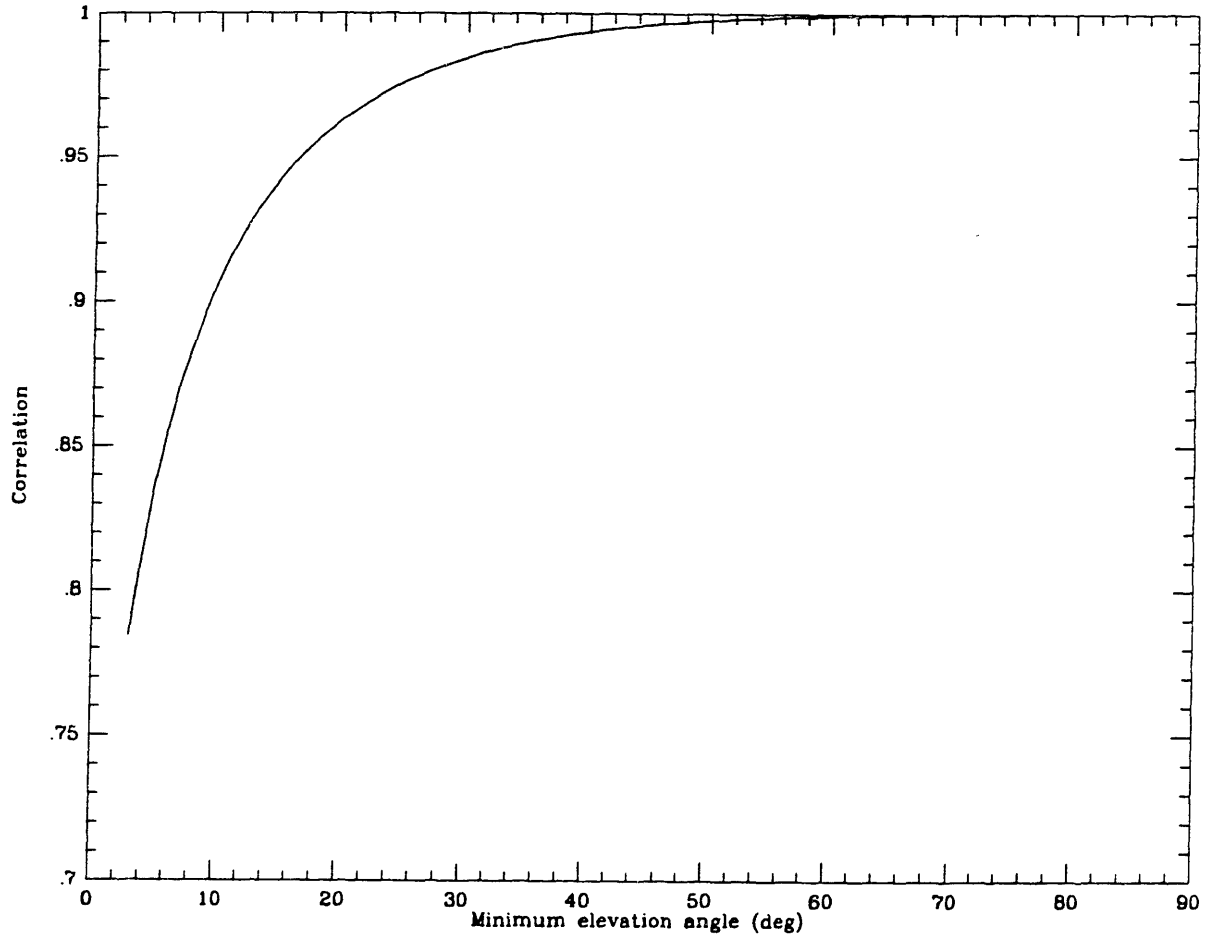


Figure 1.5.4. Correlation between the estimates of the vertical coordinate of site position and the zenith delay parameter, as a function of the elevation-angle lower limit (see text).

we remove the zenith delay parameter from the solution. We will use these facts in Chapter 3 to help us detect the presence of mapping function errors.

We now turn to another possible error in the models for the propagation delay: azimuthal asymmetry.

1.5.ii Azimuthal asymmetry

In this section we will continue our study of the effects of errors in the models for the propagation delay by investigation of the effects of azimuthal asymmetry. Both the Marini and Chao formulas given above assume that the atmosphere is azimuthally symmetric about the site, but it is reasonable to expect that the real atmosphere is not so symmetric. Our approach will be to develop a model for the asymmetry error, and to estimate its effects on the estimates of site position in a manner identical to the previous section.

The model we will use to model horizontal inhomogeneities was developed by Gardner [1977]. In this model, Gardner first expresses the refractivity in cylindrical coordinates, with the positive z–axis directed up along the local vertical; he then expands the refractivity in a one-dimensional Taylor series about the horizontal distance ρ from the vertical column above the site

$$N(\rho,z) = N(0,z) + \rho\left[\frac{\partial}{\partial\rho}N(\rho,z)\right]_{\rho=0} + \frac{\rho^2}{2!}\left[\frac{\partial^2}{\partial\rho^2}N(\rho,z)\right]_{\rho=0} + \ldots \qquad (1.5.5)$$

Assuming azimuthal symmetry is equivalent to assuming that $N(\rho,z) = N(0,z)$ for all ρ. Gardner calculated the effect of the higher order terms by integrating them along the propagation path of the signal. This integration is equivalent to including these

terms in the first integral on the right-hand side of (1.1.2), which represents the slowing of the signal, and excluding these terms from the second integral, which represents the change in propagation path. Thus, the correction terms calculated by Gardner are zero in the zenith direction, whereas if he had included the effect on the changing of the ray-path, the corrections would not be zero. We will discuss the corrections to the ray-path below.

The approximate expression developed by Gardner for the first-order effect on the propagation delay is given by GC_1, the first–order "gradient correction," the expression for which is

$$\begin{aligned} GC_1 = & \frac{C}{\sin\epsilon \tan\epsilon}\hat{n}\cdot\vec{\nabla}(P_\circ T_\circ K_\circ) \\ & + \frac{D(1+\frac{1}{2}\cos^4\epsilon)}{\sin^3\epsilon\tan\epsilon}\hat{n}\cdot\vec{\nabla}\left(\frac{P_\circ T_\circ^2 K_\circ^2}{2-K_\circ}\right) \end{aligned} \tag{1.5.6}$$

where $\hat{n}$ is the unit vector pointing in the direction of the source, projected on the local horizontal plane, and C, D, and $K_\circ$ are given by

$$C = 10^{-6}\frac{k_1 R^2}{(M_d g)^2} \simeq 6.673\times 10^{-2} \tag{1.5.7}$$

$$D = -10^{-6}k_1\frac{2}{r_\circ}\frac{R^3}{(M_d g)^3} \simeq -6.139\times 10^{-7} \tag{1.5.8}$$

$$K_\circ = \left(1-\frac{R\beta_\circ}{M_d g}\right)^{-1} \tag{1.5.9}$$

In the above, k_1 is the dry refractivity constant, R is the universal gas constant, M_d is the molar mass of dry air, g is the acceleration due to gravity at the surface of the earth, $r_\circ$ is the distance from the center of the earth to the site, which will be taken to be the mean radius of the earth, and $\beta_\circ$ is the temperature lapse rate at the surface of the earth. The dimensions of GC_1 are meters when the gradient operator is expressed

in units of m^{-1}. This model does not include inhomogeneities in the distribution of water vapor, since there is no way to describe these deterministically. We will discuss the effects of the random variation of water vapor later this section.

In order to simplify the investigation of the effects of horizontal inhomogeneities, we will make some simplifying assumptions. Since Gardner found that the largest contribution to the right-hand side of (1.5.6) is from temperature gradients, we will assume that the gradients in pressure and lapse rate are zero. This assumption reduces (1.5.6) to

$$GC_1 \simeq \left[\frac{56.8}{\sin\epsilon\ \tan\epsilon} - 0.219\frac{(1+\frac{1}{2}\cos^4\epsilon)}{\sin^3\epsilon\ \tan\epsilon}\right]\hat{n}\cdot\vec{\nabla}T_\circ \qquad (1.5.10)$$

The maximum value of GC_1 (*i.e.*, in the direction such that $\hat{n}\cdot\vec{\nabla}T_\circ = |\vec{\nabla}T_\circ|$) is plotted in Figure 1.5.5 as a function of elevation angle for $|\vec{\nabla}T_\circ| = 0.01$ K km^{-1}. This value for the temperature gradient yields average values for the horizontal gradients reported by Gardner. Thus, if the horizontal gradients he reported could be attributable solely to temperature gradients, then the average temperature gradient could be expected to be 0.01 K km^{-1}. Although Gardner did not reveal the gradients of the individual terms of (1.5.6), he did say that temperature gradients dominated over gradients of pressure and $K_\circ$.

In deriving the expression (1.5.10) for GC_1, Gardner made approximations which make the accuracy of GC_1 poor for low elevation angles; we therefore show GC_1 only above 10° elevation in Figure 1.5.5. Gardner assumed that his correction would be used in the analysis of satellite ranging data, and implicitly assumed that no ranging would be done to satellites below this elevation angle. Therefore, if in the analysis of VLBI data we wish to apply a gradient correction to the model for the propagation

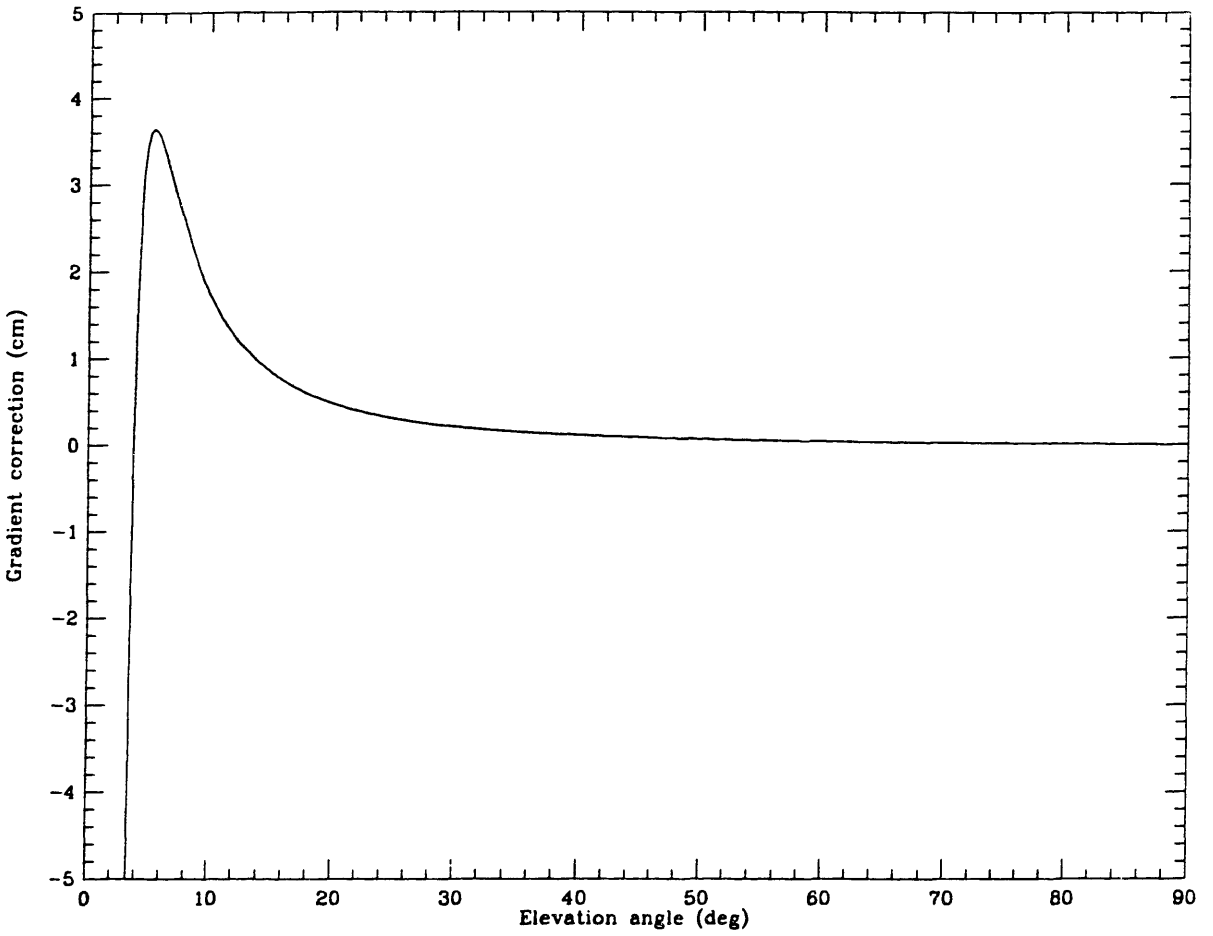


Figure 1.5.5. First order gradient correction, from Gardner [1977].

delay for observations with elevation angles below 10°, we would first have to improve the accuracy of the first–order gradient–correction formula.

In order to study the effects on the estimates of site position of unmodeled horizontal gradients, we can make use of the technique used in Section 1.5.i to study mapping function errors. However, we will make several changes to the method of analysis. We will estimate all three components of site position, since it is conceivable that the strongest effect may be in the horizontal direction. Also, to improve the realism of the study, we will use a real observing schedule instead of the idealized schedule used in the previous section. The distribution of observations used is shown in Figure 1.5.6. This distribution represents the observations made at the Mojave Base Station on 23 September 1984, when it participated in the first of the "Low Elevation" experiments which are described in Chapter 4.

The effects of horizontal gradients on the estimates of site coordinates for this experiment are shown in Figures 1.5.7 and 1.5.8, and are based on a horizontal–gradient error of the form (1.5.10), with a temperature gradient of 0.01 K km^{-1}. The results scale with the size of the gradient, but the dependence on the direction of the gradient is more complicated, so we have repeated the test for the gradient direction having azimuths of 0° through 170° in steps of 10°. (A gradient in a given direction is equivalent to a gradient in the opposite direction with the same magnitude but opposite sign.) The result from each azimuth is represented by a black circle in Figures 1.5.7 and 1.5.8. The error bars shown in these figures are the statistical standard deviations, based on the assumed observational standard deviations derived from the signal-to-noise ratio [Clark *et al.*, 1985]. These error bars have been placed on the 0°–azimuth

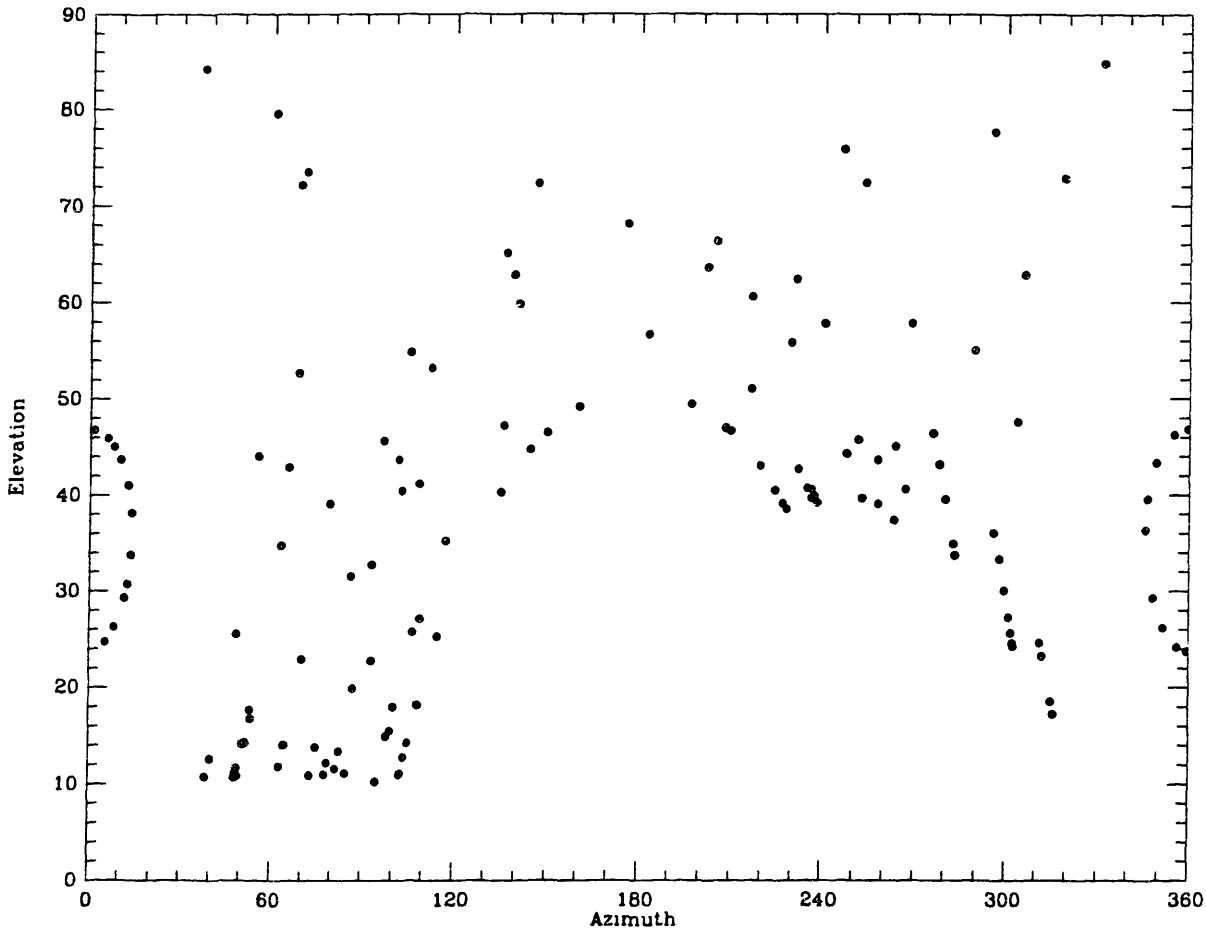


Figure 1.5.6. Distribution of VLBI observations in azimuth and elevation at the Mojave site, for the observing session used to investigate the effects of horizontal gradients on estimates of site position (see text).

results to indicate the size of the error bars as well as the 0° azimuth result. The results for successively higher values of the azimuth of the gradient error describe an ellipse, both in the X-Y plane (Figure 1.5.7) and the length-Z plane (Figure 1.5.8). The purpose of showing the error bars is so that we can compare the systematic error due to the horizontal gradient to the statistical standard deviation based upon the "known" measurement noise.

The results indicate that even though the error is "horizontal," the largest effect is in the estimate of the vertical component. The error in the vertical component is largest for the azimuth of the gradient between 50° and 60°. From Figure 1.5.6, it can be seen that this direction corresponds to a large number of low-elevation observations. The next largest overall effect is in the estimate of the north component of site position. This fact might at first be unexpected, because the east axis (*i.e.*, 90° azimuth) corresponds to the concentration of low-elevation observations, and we might believe that when the gradient is in this direction we will suffer large errors in the estimate of the horizontal component in this direction. However, it is important to recall from (1.5.4) that the elevation–dependence for an error in the the horizontal coordinates of site position scales as the cosine of the elevation angle. If the set of observations from a particular azimuth has associated with it a large range of elevation angles, then a change in the component of site position in the direction of this azimuth will not produce an effect which is correlated with the horizontal gradient effect; the horizontal gradient will show up in the estimate of the vertical coordinate. If the set of observations from a particular azimuth has associated with it a small range of elevation angles, then it is not possible to "separate" a change in the horizontal coordinate from the

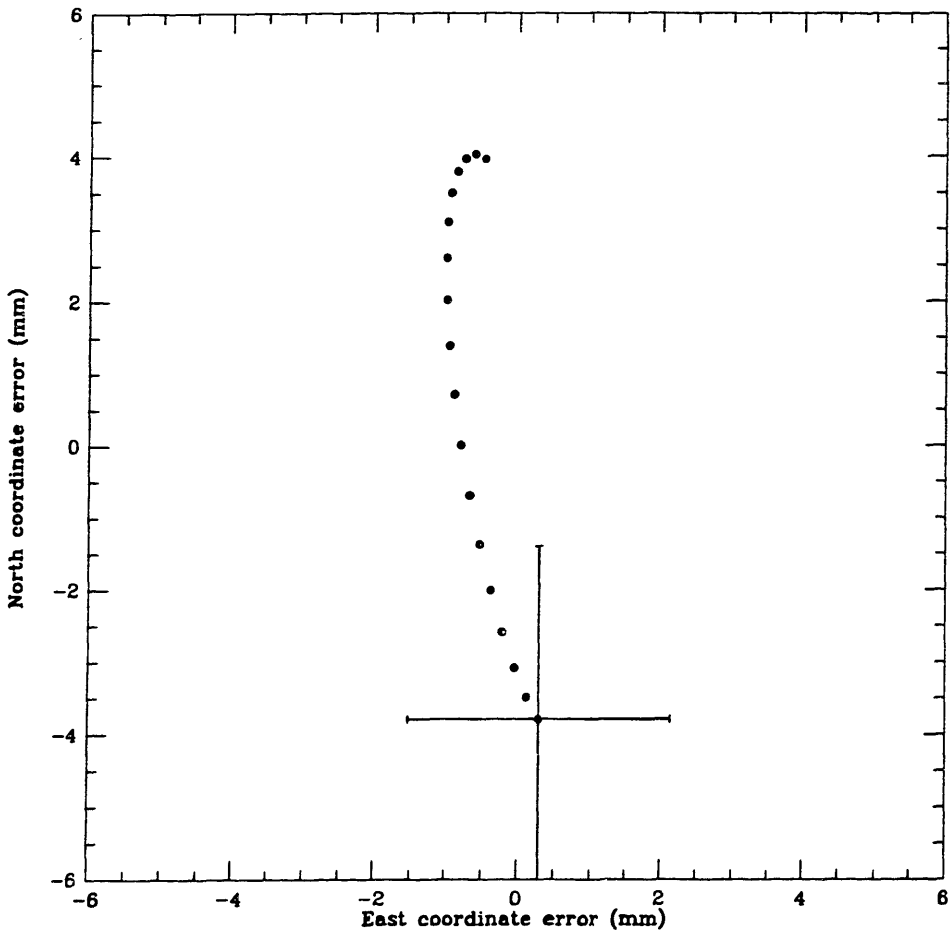


Figure 1.5.7. Errors in the estimates of the horizontal coordinates of site position due to a horizontal gradient (see text). The error bars shown are the standard deviations of the respective estimates. These error bars are shown only for the result for the azimuth of the gradient being 0°.

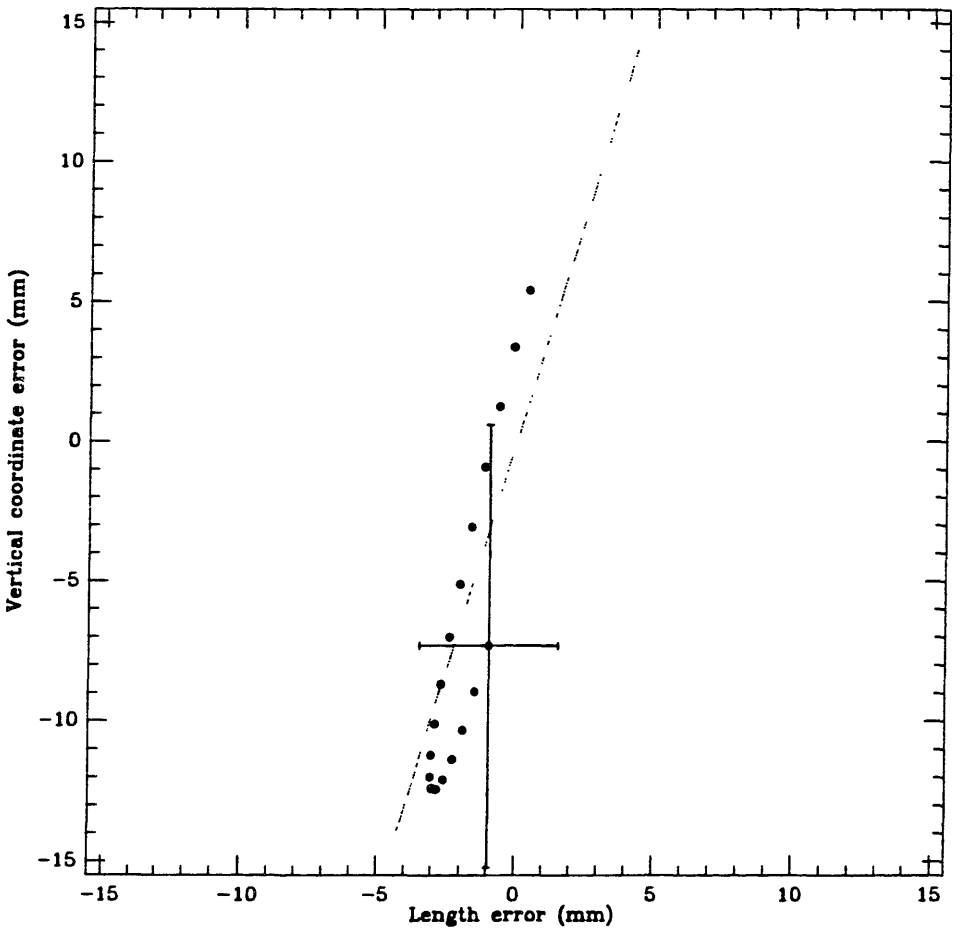


Figure 1.5.8. Errors in the estimates of the vertical coordinate and baseline length due to a horizontal gradient (see text). The meaning of the error bars is the same as in Figure 1.5.7. The dotted line indicates the error in the estimate of the baseline length due solely to an error in the respective estimate of the vertical coordinate of site position. The Mojave–Haystack baseline length is approximately 3900 km.

horizontal gradient effect. For the "Low-Elevation" experiment, the azimuths with a large range of associated elevations are roughly along an east–west axis. The azimuths with a small range of associated elevations are roughly along a north–south axis.

The dotted line in Figure 1.5.8 indicates the error in the estimate in baseline length due solely to an error in the vertical coordinate of site position. It can be seen that this line falls very close to the ellipse which is actually described by the errors.

The previous analysis assumed that horizontal variations of the refractive index could be expressed by a Taylor expansion in the horizontal coordinates, and that the terms of order higher than the first could be neglected. This assumption is not valid, however, if we attempt to describe the random variations in the distribution of water vapor. These variations are caused by the turbulence of the lower troposphere, wherein most of the water vapor exists. We will assume that the estimates of the parameters (site position, clock offset, and zenith delay), are obtained via the familiar standard least–squares solution

$$\hat{\mathbf{x}} = (A^T G^{-1} A)^{-1} A^T G^{-1} \mathbf{y} \tag{1.5.11}$$

Here $\hat{\mathbf{x}}$ is a vector containing the estimates of the differences of the parameters from their *a priori* values, $\mathbf{y}$ is a vector containing the differences of the observations and their value based on the *a priori* values for the parameters, G is a covariance matrix of the measurement noise, and A is the derivative matrix. We have, of course, already implicitly made use of (1.5.11) to study the effects of mapping function errors and horizontal gradients. In order to study the effects of the random distribution of water vapor, we will include only that contribution to $\mathbf{y}$ due to water-vapor turbulence. Because we will be discussing the effects of non-zero zenith-delay errors in the next

section, for these calculations we will assume that $\mathbf{y}$ is zero–mean, which implies that $\hat{\mathbf{x}}$ is zero mean. In order to characterize the random variations of the parameters due to the random variations of the water vapor, we can look at the covariance matrix of the estimates of the parameters, which can be obtained from (1.5.11):

$$\begin{aligned} K_x \equiv \langle \hat{\mathbf{x}}\hat{\mathbf{x}}^T \rangle &= \left(A^T G^{-1} A\right)^{-1} A^T G^{-1} \langle \mathbf{y}\mathbf{y}^T \rangle G^{-1} A \left(A^T G^{-1} A\right)^{-1} \\ &= \left(A^T G^{-1} A\right)^{-1} A^T G^{-1} K_v G^{-1} A \left(A^T G^{-1} A\right)^{-1} \end{aligned} \tag{1.5.12}$$

where $\langle\ \rangle$ indicates expectation, K_x is the covariance matrix of the parameter estimates, and K_v is the covariance of the contribution to the observations due to the random variations of water vapor.

We will now derive an expression for K_v based upon the known correlation spectrum for turbulence in the troposphere, and a simple model for the vertical distribution of water vapor. (A more sophisticated model is described in Appendix B.) We will discuss the effects of the simplifications later in this section. We will assume that all the water vapor in the troposphere is contained in a thin plane above the surface of the earth (also assumed planar) at a height H_v. The turbulence in that layer produces "frozen features" in the distribution of water vapor, which move in the horizontal plane with speed v and in a direction with azimuth α_v. The turbulence is assumed to be described by the Kolmogorov "2/3 law" [Tatarskii, 1961], which yields for the two-dimensional structure function D of zenith delay

$$D(\rho) \equiv \langle [\tau_a^z(\mathbf{x}_1) - \tau_a^z(\mathbf{x}_2)]^2 \rangle = \begin{cases} 2.91 C_n^2 \Delta H \rho^{5/3}, & \rho \le \rho_o \\ 2.91 C_n^2 \Delta H \rho_o^{5/3}, & \rho > \rho_o \end{cases} \tag{1.5.13}$$

Here $\tau_a^z(\mathbf{x})$ is the zenith delay measured up through the layer over the point $\mathbf{x}$ (see Figure 1.5.9), and $\rho = |\mathbf{x}_1 - \mathbf{x}_2|$. The "strength" of the turbulence is characterized

by the terms $C_n^2 \Delta H$ and ρ_o. The term $C_n^2 \Delta H$ comes from the integration of the refractive–index structure function vertically through the layer of water vapor, which is assumed to have thickness ΔH. Values for $C_n^2 \Delta H$ have been determined experimentally using interferometry; the values for the determinations range from about 2×10^{-13} to 2×10^{-9} $\mathrm{m}^{1/3}$ (see Armstrong and Sramek [1982], *e.g.*). The larger this number, the more turbulent the troposphere. The term ρ_o is the "outer scale" of the turbulence, and has never been measured; its value is thought to be about 2 km.

We can derive a correlation function $R(\rho)$ for the zenith delay from (1.5.13) if we assume that the turbulence is a stationary process. We then find

$$R(\rho) = R(0) - \frac{1}{2} D(\rho) = \begin{cases} \sigma_v^2 - \sigma_o^2 \left(\frac{\rho}{\rho_o} \right)^{5/3} & \rho \leq \rho_o \\ \sigma_v^2 - \sigma_o^2 & \rho > \rho_o \end{cases} \tag{1.5.14}$$

In (1.5.14), instead of writing the structure function in terms of the refractive–index structure factor C_n^2 and layer–thickness ΔH, we have expressed it in terms of the outer–scale length ρ_o and a corresponding "outer–scale variance" σ_o^2. The correlation at zero length $R(0)$ has been expressed in terms of a zero–length variance σ_v^2.

We can now write the general expression for an element of the covariance matrix K_v. For the element with the indices i and j we have

$$K_v(i,j) = R(\rho_{ij}) \csc \epsilon_i \csc \epsilon_j \tag{1.5.15}$$

where ρ_{ij} is the distance between the point in the layer of water vapor which is pierced by an imaginary line connecting the site and the source observed at time t_i, and the point which is pierced by an imaginary line connecting the site and the source observed at time t_j, in a frame of reference which is stationary with respect to the moving layer

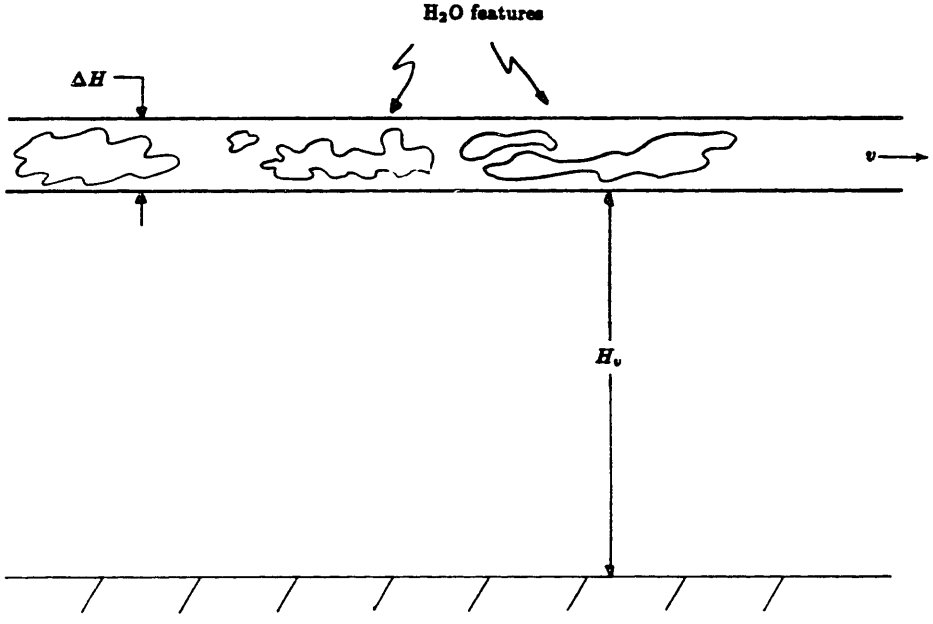


Figure 1.5.9. Geometry for the investigation of the effects of the random horizontal distribution of water vapor (see text).

of water vapor. The elevation angle of the source is ϵ_i at time t_i, and ϵ_j at t_j. (The sources observed at t_i and t_j need not be the same ones.) The separation ρ_{ij} can be written

$$\rho_{ij} = \left[(\Delta x_{ij})^2 + (\Delta y_{ij})^2\right]^{1/2} \tag{1.5.16}$$

where

$$\Delta x_{ij} = H_v[\cot \epsilon_i \sin \alpha_i - \cot \epsilon_j \sin \alpha_j] - v(t_i - t_j) \sin \alpha_v \tag{1.5.17}$$

$$\Delta y_{ij} = H_v[\cot \epsilon_i \cos \alpha_i - \cot \epsilon_j \cos \alpha_j] - v(t_i - t_j) \cos \alpha_v \tag{1.5.18}$$

Therefore, in order to calculate (1.5.12), we must specify the values for six parameters: H_v, α_v, v, ρ_o, σ_v, and σ_o.

In our calculation of K_x, we chose values for these parameters which we believed to be representative of "mild" conditions. (The effect of these choices are discussed below.) We assumed complete decorrelation of the variations for $\rho > \rho_o$, so that $\sigma_v^2 = \sigma_o^2$. We took σ_v to be 3 mm, which from the analysis of water-vapor radiometer data (see Chapter 2) we know is an upper bound for the minimum value of σ_v. For the outer-scale length ρ_o we took $\rho_o = 1.5$ km. (These values imply a value for $C_n^2 \Delta H$ of 3.1×10^{-11} $m^{1/3}$.) We chose a height of 1.5 km for the water-vapor layer, and a horizontal wind velocity of 2 m s^{-1}. The calculations were performed for wind-directions of from 0° to 320° in steps of 40°. The distribution of observations was that from the Mojave station, used in the previous section.

The results indicate no difference at the level of 0.2 mm between the effect on site position of using a full matrix of Kolmogorov-type covariances and of using a diagonal, *i.e.*, white-noise, covariance matrix. (The effect of such a white-noise covariance matrix

will be discussed in the following section.) This result is due to the sparseness of the Kolmogorov covariance matrix: only observations closely separated in time and in the approximately same direction will for all intents and purposes be correlated.

Performing the analysis with different model parameters would probably not be enlightening. For more turbulent atmospheres, both σ_v^2 and σ_o^2 would likely increase, but so would v, so that the these atmospheres would tend toward white noise. For milder atmospheres, both σ_v^2 and σ_o^2 would decrease, so that even though the correlations increase, the covariances decrease. Furthermore, although we realize that our model is a simplified version of reality, we expect our results to be correct to within an order of magnitude, putting a very realistic upper bound of a few millimeters on these effects. It should be noted, though, that the observing schedule used to perform the preceding analysis is a so-called "geodetic observing schedule," with many widely separated sources being observed. For observing schedules which have one or two sources being "tracked" continuously or nearly so, one would expect the atmospheric correlations to have a greater effect.

We will now discuss a final aspect of possible errors in the formulas for the propagation delay, namely zenith delay errors.

1.5.iii Zenith delay errors

The effects of errors in the *a priori* estimate of the zenith delay can be calculated easily using the same methods as were used in Sections 1.5.i and 1.5.ii. We will assume that the errors in the estimates of the horizontal components are small, and estimate only the vertical coordinate of site position.

The results obtained are shown in Figure 1.5.10. The solid curve is the error in the estimate of the vertical coordinate, for a zenith delay error of 1 cm, plotted against minimum elevation angle. Herring [1986] has also examined the effect on the estimate of the vertical coordinate of site position for a randomly changing zenith delay. His results are also shown in Figure 1.5.10. The dashed curve labeled I is for a solution with a zenith delay parameter estimated, while curve II is for a solution with no zenith delay parameter estimated. Unlike the effect of a constant zenith delay error, the effect of a random zenith delay error is greater when the zenith delay parameter is estimated, because of the correlation between the estimates of the parameters of zenith delay and vertical coordinate of site position. By estimating a zenith delay parameter, we make our estimates of the vertical coordinate of site position less sensitive to a zenith-delay bias (*i.e.*, an error in the zenith delay which is constant for the duration of the observing session), while increasing their sensitivity to short-term random variations of the zenith delay.

We have now concluded our discussion of the effects of errors in the models for the atmospheric propagation delay. In Section 1.5.*iv*, we will briefly summarize these results. In Section 1.6, we will use these results to discuss future treatment of the atmosphere in VLBI data analysis. Later, in Chapter 3, we will examine VLBI data for evidence of such errors.

1.5.iv Summary

We have now examined the effect of three types of errors in the model for the atmospheric propagation delay: mapping function errors, horizontal gradient errors,

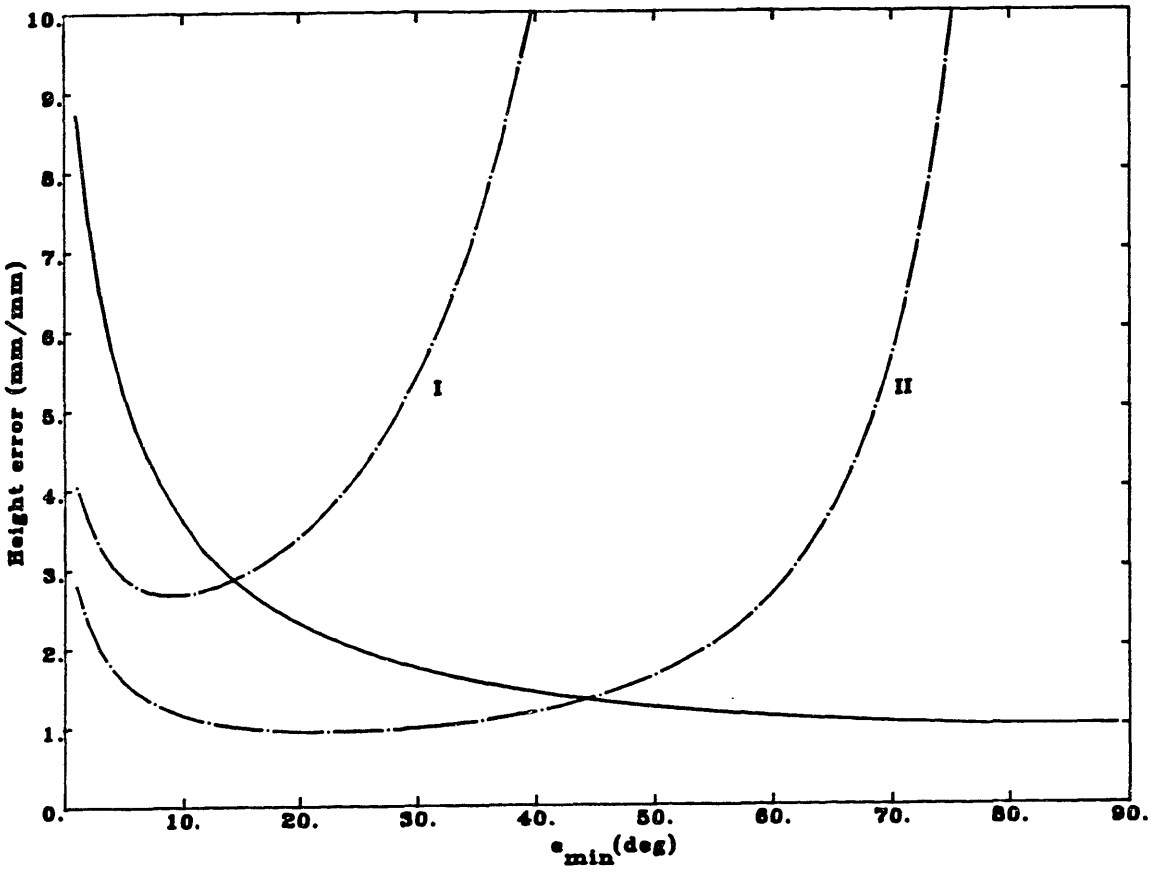


Figure 1.5.10. Effects on the estimate of the vertical coordinate of site position of errors in the zenith delay, as a function of the elevation-angle lower limit, from Herring [1986] (see text). The solid curve represents the effect of a constant error in the zenith delay when no zenith delay parameter is estimated. The dash-dot curves represents the effects of uncorrelated random variations in the zenith delay for the cases in which a zenith delay parameter is estimated (I) and in which no zenith delay parameter is estimated (II).

and zenith delay errors. We can summarize our knowledge of the effects of these errors on the estimates of site position:

- Of the three components of site position (east–north–vertical), the estimate of the vertical component is most seriously affected. This statement is true even for errors due to horizontal gradients. The primary reason that the estimate of the vertical component suffers the most is that an error in the vertical coordinate induces an error in the model of the group delay which is proportional to $\sin\epsilon$, which correlates highly with all forms of errors in the model for the propagation delay.
- In general, the effect of an error in the model for the propagation delay increases with decreasing elevation angle. Estimates of site position made from VLBI data taken using an observing schedule with many low elevations will therefore suffer more heavily than if all the observations had been in directions with high elevation angles. However, this rule is not necessarily valid for time-dependent errors, as can be seen from Figure 1.5.10.
- Estimation of a zenith delay parameter can affect the estimate of site position, even if the *a priori* zenith delay is correct, because other types of systematic errors in the model for the propagation delay (*e.g.*, mapping function errors and horizontal gradients) are also elevation dependent.
- The statistical standard error of the estimate of the vertical coordinate of site position can be increased several-fold by the simultaneous estimation of a zenith

delay parameter, due to the correlations between the estimates of those quantities. The amount of the increase depends on the elevation angles of the observations. If observations from low elevations are included, the correlations, and hence the increase, are less than if no such observations are included.

In the next section we will add to our discussion of Section 1.4 dealing with the treatment of the atmosphere in VLBI data analysis, based on what we have learned in this section.

1.6 Treatment of the atmosphere in VLBI data analysis. II

In this final section of Chapter 1, we will discuss several more topics dealing with treatment of the atmosphere in VLBI data analysis. In Section 1.6.*i*, we will discuss the motivation and the possible consequences of "reweighting" the data. In Section 1.6.*ii*, we will discuss the estimation of atmospheric parameters from VLBI data. The rest of this thesis will be devoted to methods of improving the atmospheric models.

1.6.i Treatment of the time-variability of the atmosphere: Reweighting

In Section 1.4, we discussed how the atmospheric propagation delay is treated in VLBI data analysis. First, a model for the propagation delay is used to generate the *a priori* value for the delay, and then a zenith delay is estimated in the least-squares analysis. In practice, if after this analysis is performed the postfit residuals of group delay still exhibit obvious elevation-angle-dependent behavior, the correction

to the *a priori* model for the propagation delay can be represented as a polynomial in time and the coefficients estimated, or an independent zenith delay can be estimated for different time intervals within the experiment, as the analyst sees fit. (The two techniques may, of course, be combined. Moreover, a polynomial representation is not necessarily the most sensible representation.) Regardless of which of these techniques is used, it is not unreasonable to believe that the true behavior of the atmosphere with time is much more complicated. The horizontal structure of the refractive index is in reality dominated by the turbulence in the lower troposphere; this turbulence is "fully developed" and can be described by the Kolmogorov "2/3 Law" [Tatarskii, 1961].

In order to account for the inadequacy of the simple atmospheric models, a method known as "data reweighting" is used. To understand how this method works, consider the following simple situation. Suppose we wish to estimate some deterministic, time-independent quantity x given a set of n measurements y_i, $i = 1, \ldots, n$, where the y_i are given by

$$y_i = x + \epsilon_i + z_i \tag{1.6.1}$$

In (1.6.1), ϵ_i is a random Gaussian number, with zero mean and variance σ_y^2, and represents measurement noise. (We will assume the ϵ_i are independent.) The term z_i in (1.6.1) is some Gaussian random process which is interfering with our "direct" measurement of x; the z_i are independent and have a mean of zero and a variance of σ_z^2. The z_i and ϵ_i are independent of each other. The z_i are analogous to the time-dependent, random atmosphere.

Since the z_i are zero mean and independent of the ϵ_i, we can write (1.6.1) as

$$y_i = x + \epsilon_i' \tag{1.6.2}$$

where ϵ_i' is a zero mean Gaussian number with variance $\sigma_y^2 + \sigma_z^2$. That is, the presence of the random process represented by the z_i is equivalent to having no such random process and increasing the variance of the measurement noise by σ_z^2.

The effect on the precision of the estimate of x is obvious. If the z_i did not exist, the uncertainty in the estimate of x would have been $\sqrt{\frac{\sigma_y^2}{n}}$; with the existence of the z_i, the uncertainty of the estimate of x is $\sqrt{\frac{\sigma_y^2+\sigma_z^2}{n}}$.

Thus we can "account for" the presence of the z_i by ignoring them in all ways except for increasing the effective measurement variance by the variance of the z_i. If we had some independent estimates of the z_i, then we could include these in the observation equations; these observations would help to decrease the uncertainty in the estimate of x if the observation error for the measurements of the z_i were less than or of order of our *a priori* uncertainty, σ_z.

If σ_z is unknown, then some other information must be supplied. In current methods of VLBI data analysis, that information is in the form of a criterion that after addition of the variance for the atmosphere, the χ^2 per degree of freedom for the postfit residuals for a particular pair of sites be unity. (The atmospheric variance is added in a baseline–dependent manner.) Since the postfit residuals depend on the weights used in the least–squares analysis, the analysis is usually performed with the new atmospheric weighting, until some convergence criteria is met. This whole process is known as reweighting. (In reality, reweighting is meant not only to account for the

atmospheric mismodeling, but also for mismodeling of the behavior of the hydrogen maser "clocks.")

The point which we wish to make concerning reweighting is that although the process is statistically "valid" if all the assumptions concerning the statistical noise process are correct, if these assumptions are incorrect then reweighting may not have the desired result, and may have undesirable results. For example, let us assume that our *a priori* value (or values, for a time-varying atmosphere) for the zenith delay is (are) correct, but that we have used an incorrect mapping function. We can reweight this data set, and one can argue that the net effect is to increase the uncertainty in the estimates of site position, which will in fact account for our "uncertainty" in the atmospheric model. However, this argument does not take into consideration that a mapping-function error may be of a systematic nature, thereby introducing errors in the estimates of site position which are biased or may have long-period (*e.g.*, seasonal) correlations. Also ignored is the possibility that the increase in the uncertainty of the estimates may not be a true reflection of the variability of the estimates.

Reweighting is really the process of adding noise to the data so that errors in the model effectively go "unnoticed." If the nature of the errors can be represented by a white-noise random process, then reweighting is an entirely adequate procedure. However, the net effect is to decrease the "signal," or the amount of information, in the data. If a more accurate model can be found, then presumably less reweighting would be necessary, and some of the signal may be recovered. Parametrizing the model further uses the signal in the data to determine the errors in the *a priori* model; the uncertainties of all the estimated parameters increase to reflect the correlations with

the new parameters. If feasible, either of these techniques is preferable to reweighting. In the following section, we will discuss the estimation of atmospheric parameters from VLBI data. In the chapters that follow Chapter 1, we will discuss the possible improvement of the atmospheric model.

1.6.ii Estimation of atmospheric parameters

In this section we will discuss the possible parametrization of the model for the propagation delay. In the previous section, we noted that it is usual in the analysis of VLBI data to estimate a correction to the *a priori* value for the zenith delay. However, we can also introduce the parameter Δa for mapping function errors, and from (1.5.6) any horizontal gradients may be parametrized approximately by a direction and an amplitude. Why then not estimate these parameters from the VLBI data in the same way a zenith delay is estimated? In principle, this estimation is possible. In practice, the estimation of these parameters is undesirable because they will all be very highly correlated among themselves and with the estimate of the vertical coordinate of site position. These correlations, of course, cause the statistical uncertainty of all the parameter estimates to increase. Particularly affected are those parameters, such as the vertical coordinate parameter, which have associated high correlations. To get an idea of the effect, refer to Figure 1.5.3, which shows the increase in the statistical uncertainty of the estimate of the vertical coordinate when a zenith delay parameter was estimated. For a minimum elevation of 10°, the uncertainty in that coordinate, when a zenith delay parameter was estimated, was more than twice the uncertainty when a zenith delay parameter was not estimated.

If we limit ourselves to one atmospheric parameter, why do we choose to estimate a zenith delay parameter rather than, say, a mapping function parameter? The main reason is that we *know* that our *a priori* estimates of the "wet" zenith delay are very inaccurate (see Chapter 2). It was not until recently, however, that clear evidence was presented for mapping function errors (see Chapter 3). In practice, if a zenith delay error did exist one could probably estimate the mapping function parameter and obtain estimates of site position fairly close to those that would be obtained if one estimated a zenith delay parameter, provided the observations were not from elevation angles that were too low, because an error which scales as $\csc \epsilon$ can be represented fairly well by the sum of a term which scales as $\frac{\cos \epsilon}{\sin^3 \epsilon}$ (*i.e.*, the mapping-function parameter) and a constant term (*i.e.*, the clock-offset parameter).

The above considerations play an important role in the design of many VLBI experiments. In designing an experiment, one must decide "how low" to observe. Since errors in the model for the propagation delay tend to increase with decreasing elevation angle, the tendency is often to observe above about 10° elevation. However, from Figure 1.5.3 it can be seen that the effect on the uncertainty of the estimate of the vertical coordinate increases rapidly as the minimum elevation-angle is increased.

In later chapters, we will detail attempts to increase the accuracy of the models for the propagation delay. In Chapter 2, we will discuss the estimation of the "wet" delay from ground-based radiometry. In Chapter 3 we will discuss attempts to improve the mapping function. In Chapter 4, we will discuss the planning and results from a set of experiments which have been used to test the models for the propagation delay.

Chapter 2

Water-Vapor Radiometry

Introduction

In Chapter 1 we spoke several times of the unpredictability of the wet path delay. This unpredictability arises directly from the "unmixed" condition of water vapor in the atmosphere. The term "unmixed" refers to the presence under normal conditions of localized sources of water vapor, usually in the form of liquid water. These sources along with turbulence in the lower troposphere lead to local variations in the concentration of water vapor which are not correlated with variations even a short distance away. In practical terms, this decorrelation means that the vertical profile of humidity is unknown, and that the wet delay—even in the zenith direction—is not predicted with any accuracy by formulas which are based solely on surface conditions.

In this chapter, we will discuss the means by which measurements of the integrated intensity of emission of water vapor at microwave frequencies may be used to infer the wet delay, and discuss the precision and accuracy of the instruments, known as water-vapor radiometers, which make these measurements. We will begin in Section 2.1 by describing how exactly the wet delay may be related to such radiometric measurements. In doing so, we will introduce the single-frequency weighting function. Then in Section 2.2, before we discuss the dual-frequency algorithm, we will present some necessary background dealing with the microwave absorption spectrum of the atmosphere.

We will then be in a position in Section 2.3 to introduce the dual-frequency algorithm. In Section 2.4, we will review several important instrumental features of radiometers, and the way these features affect calibration via internal calibration loads. In Section 2.5, we discuss a common alternative to calibration via internal loads, namely calibration by the tip-curve method. We follow these discussions of instrumental effects by a discussion of the overall accuracy of the dual-frequency algorithm presented in Section 2.3. We end this chapter by discussing alternate algorithms for estimation of the wet delay from radiometric measurements, including alternate dual-frequency algorithms (Section 2.7) and multichannel water-vapor radiometry (Section 2.8).

2.1 Relating the wet path delay to radiometric quantities

In this section, we will discuss the means by which radiometric quantities may be related to the wet propagation delay. In order to discuss these methods, however, we must first define the wet delay. In Chapter 1, we discussed possible definitions for the both the "dry" and the "wet" delay, and how these definitions are related to each other. Hereafter, we will define the wet delay L_w to be the second term of (1.3.18):

$$L_w = 10^{-6} R_v \int_{atm} ds \left(k_2 - \frac{M_v}{M_d} k_1 + \frac{k_3}{T} \right) \rho_v \tag{2.1.1}$$

In the above equation, k_1, k_2, and k_3 are the constants in the refractivity formula discussed in Section 1.2.*iv*; R_v is the specific gas constant for water vapor; M_v is the molar mass for water and M_d the molar mass of "dry air," discussed in Section 1.2.*iii*; ρ_v is the density of water vapor; and T is the absolute temperature. The path indicated by *atm* is the path of the ray through the atmosphere, as discussed in Section 1.1.

The use of the symbol "L" for delay—rather than "τ," which was used throughout Chapter 1—enables us to distinguish delay from opacity, which is introduced below. Hereafter, we will combine the constants k_1 and k_2 used in (2.1.1) above and define k_2':

$$k_2' \equiv k_2 - \frac{M_v}{M_d} k_1 \tag{2.1.2}$$

The use of radiometers to estimate the wet path delay is motivated by the equation for the optical depth τ_v of water vapor in the atmosphere, along the path of the ray [Chandrasekhar, 1960]

$$\tau_v = \int_{atm} ds\, \kappa_v \rho_v \tag{2.1.3}$$

where κ_v is the mass absorption coefficient of water vapor, which is approximately independent of ρ_v, but which depends heavily on P, T, and frequency (see Section 2.2). Equation (2.1.3) prompts us to write (2.1.1) in the form

$$\begin{aligned} L_w &= 10^{-6} R_v \int_{atm} ds \left[\left(k_2' + \frac{k_3}{T} \right) \kappa_v^{-1} \right] \kappa_v \rho_v \\ &= W_\circ^{-1} \tau_v \end{aligned} \tag{2.1.4}$$

where the constant $W_\circ$ has been defined as

$$W_\circ \equiv \left[\frac{10^{-6} R_v \int_{atm} ds \left[\left(k_2' + \frac{k_3}{T} \right) \kappa_v^{-1} \right] (\kappa_v \rho_v)}{\int_{atm} ds\, \kappa_v \rho_v} \right]^{-1} \tag{2.1.5}$$

An alternate interpretation of the constant $W_\circ$ can be found. If the "weighting function" $W(s)$, given by

$$W(s) = 10^6 R_v^{-1} \left(k_2' + \frac{k_3}{T(s)} \right)^{-1} \kappa_v(s) \tag{2.1.6}$$

is constant for all s, then we have $W_\circ = W(s)$. Furthermore, it is simple to evaluate $W_\circ$, using values for the temperature in (1.2.6), and calculating the mass absorption coefficient from formulas given later in this chapter. Estimates for the opacity of water vapor can then be used in (2.1.4) to yield estimates of the wet path delay.

Before examining the weighting function $W(s)$ in detail, let us first look for a means for determining the opacity τ_v. The total opacity τ of the atmosphere can be determined radiometrically by measuring the atmospheric brightness temperature. The relationship between the atmospheric brightness temperature T_a and the opacity τ of the atmosphere is expressed by the equation of radiative transfer [Chandrasekhar, 1960]:

$$T_a = T_{bg}e^{-\tau} + T_{eff}\left(1 - e^{-\tau}\right) \tag{2.1.7}$$

where $T_{bg} = 2.7$ K, and T_{eff} is an "effective temperature," defined below. Equation (2.1.7) is for a radiometer pointed at the atmosphere in a direction in which there are no sources of radiation except the 2.7 K microwave background and the atmosphere itself. The effective temperature is defined by

$$T_{eff} \equiv \frac{\int_{atm} ds\, T(s)e^{-\tau(s)}\kappa(s)\rho(s)}{(1 - e^{-\tau})} \tag{2.1.8}$$

where $\kappa(s)$ is the *total* mass absorption coefficient (not just that due to water vapor), $\rho(s)$ and is the total density. The function $\tau(s)$ is defined to be the optical depth of the atmosphere were the atmosphere to be terminated at that point:

$$\tau(s) = \int_{s'<s} ds'\, \kappa(s')\rho(s') \tag{2.1.9}$$

where the integral, as before, is to run along the path *atm*.

As mentioned above, the primary "observable" of a radiometer is the atmospheric brightness temperature (although it does not directly "observe" this quantity). However, from (2.1.4) we can see that we do not need the total opacity, but the opacity due solely to water vapor, and the contributions to the opacity from other atmospheric constituents (mainly oxygen and liquid water) are not negligible, at least at those frequencies at which one would observe in order to estimate the opacity of water vapor. Thus there must be some way to determine the opacity of water vapor only, before we can make use of (2.1.4). In the following sections, we will develop the so-called dual-frequency algorithm, in which observations at two frequencies are used to determine the opacity of water vapor. We will discuss algorithms which use more than two frequencies in Section 2.7. Before we discuss these other algorithms, however, we must first first discuss absorption in the atmosphere for the individual constituents.

2.2 Absorption of microwaves by the atmosphere

Figure 2.2.1 shows the microwave absorption spectrum for frequencies below 300 GHz. The main features of this spectrum are the water-vapor absorption lines near 22 and 183 GHz, O_2 absorption lines near 60 and 118 GHz, and a number of sharp lines due to ozone. (Not included in this figure is the effect of liquid water, although liquid water plays an important role in microwave absorption in the atmosphere, as we will see.) In this section, we will discuss the properties of the microwave spectrum near the 22 GHz water-vapor line. It is near this line that existing WVR's operate, in order to take advantage of the relatively low atmospheric opacity. There are radiometers that use the 60 GHz O_2 lines to obtain information concerning the oxygen opacity,

but this use will not be discussed until Section 2.8. We will first examine the 22 GHz water-vapor absorption line.

2.2.i The 22.235 GHz line of water vapor

Unlike any of the primary constituents of dry air, water vapor possesses a permanent electric dipole moment. Due to the rotational transitions associated with this dipole moment, water vapor possesses a microwave absorption spectrum. (See, *e.g.*, Debye [1929].) The lines in this spectrum are collision-broadened, and so possess a Lorentzian shape [Van Vleck and Weisskopf, 1945]. Van Vleck [1947b] showed that the accumulated effect of all the rotational lines above the 22.235 GHz line could be written as a single additive term to the absorption near this line, because this line is well separated from the other water-vapor lines and the tails of these lines all have approximately the same shape. Barrett and Chung [1962] found that Van Vleck's theoretical expression for this accumulated effect disagreed with experimental results from, *e.g.*, Becker and Autler [1946] possibly due to Van Vleck's assumption of a single line width for all of the lines. Barrett and Chung therefore introduced an empirical correction factor for this accumulative-effects term. Their expression for the water-vapor absorption coefficient α_v, expressed by Staelin [1966] in convenient units, is

$$\begin{aligned} \alpha_v = \kappa_v \rho_v = 3.43 \times 10^{-3} e^{-644/T} \frac{\nu^2 \rho_v}{T^{2.5}} \\ \times f(\nu; \nu_o, \Delta\nu) \\ + 2.55 \times 10^{-8} \rho_v \nu^2 \frac{\Delta\nu}{T^{3/2}} \quad \text{cm}^{-1} \end{aligned} \tag{2.2.1}$$

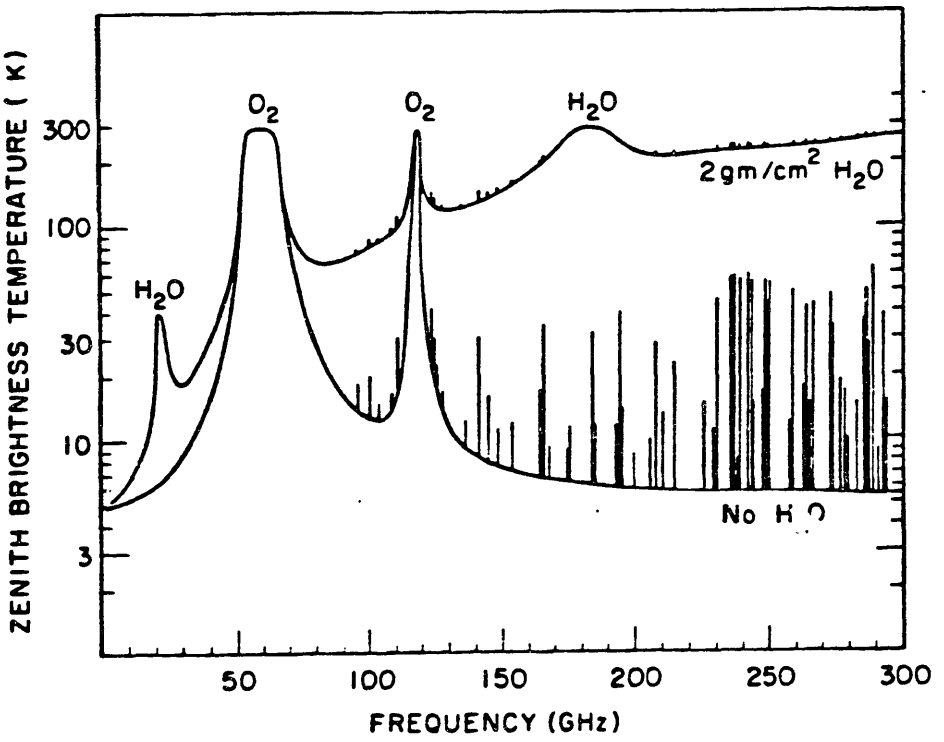


Figure 2.2.1. Atmospheric attenuation spectrum, from Waters [1976]. The lower curve has been calculated for a dry atmosphere, the upper for an integrated column density of 2 g cm^{-2} assumed distributed exponentially with a scale height of 2 km.

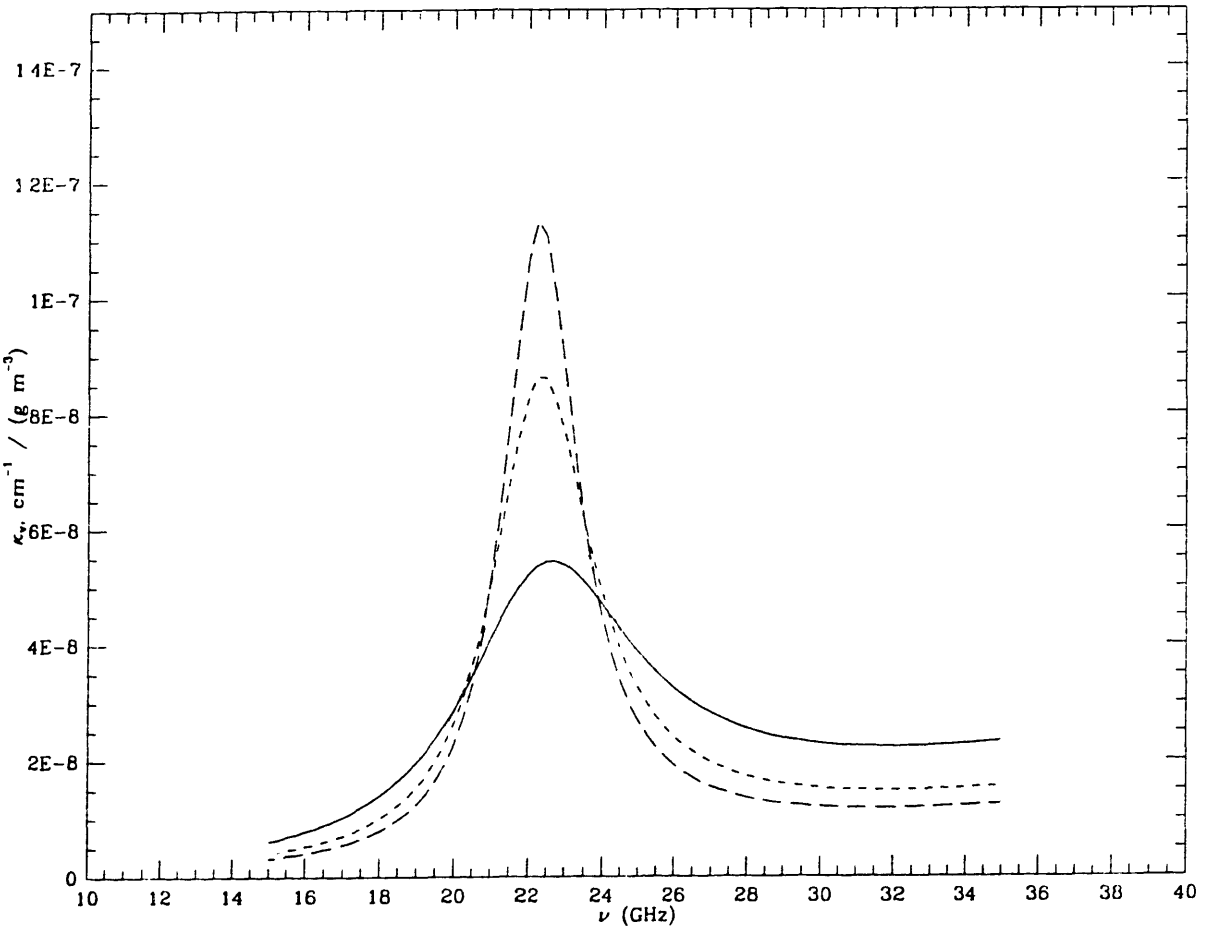


Figure 2.2.2. Mass absorption coefficient κ_υ of water vapor, at an altitude of 0 km (solid line), 2.5 km (dotted line), 5 km (short dashes), and 7.5 km (long dashes). The density of water vapor is taken to be zero, and the temperature and pressure are taken from the U.S. Standard Atmosphere [1976].

where T is the absolute temperature in Kelvins, ρ_v is the density of water vapor in g m^{-3}, and ν is the frequency in GHz. The line half-width $\Delta\nu$ in GHz is given by

$$\Delta\nu = 2.58 \times 10^{-3} \left(1 + 0.0147 \frac{\rho_v T}{P}\right) \frac{P}{(T/318)^{0.625}} \tag{2.2.2}$$

where P is the total pressure in mbars. The shape function $f(\nu; \nu_\circ, \Delta\nu)$ is given by the expression

$$f(\nu; \nu_\circ, \Delta\nu) = \left[\frac{\Delta\nu}{(\nu - \nu_\circ)^2 + (\Delta\nu)^2} + \frac{\Delta\nu}{(\nu + \nu_\circ)^2 + (\Delta\nu)^2}\right] \tag{2.2.3}$$

where $\nu_\circ \simeq 22.235$ GHz. Plots of κ_v for typical atmospheric parameters at several altitudes are shown in Figure 2.2.2. Note that κ_v does not depend on the density of water vapor except weakly through the half-width. Thus, the weighting function $W(s)$ in (2.1.6) has little water-vapor density dependence.

Waters [1976] has discussed the accuracy of the Van Vleck–Weisskopf line shape as well as alternative line shapes, such as the "kinetic" line shape. The differences between these two theoretical line shapes is small (~10%) compared to the disagreement between both theoretical line shapes and the experimental values. Thus, regardless of the line shape chosen one must add the empirical correction term proportional to ν^2 discussed for (2.2.1). However, as we shall see, the dual-frequency weighting function is insensitive to additive terms with such frequency dependence.

2.2.ii Oxygen

While O_2 possesses no permanent electric dipole moment, it does possess a permanent magnetic dipole moment, with which microwave absorption is associated. This

absorption gives rise to a large number of lines clustered near 60 GHz, appearing in Figure 2.2.1 as a single wide line. Van Vleck [1947a] showed that for frequencies below about 45 GHz, the collection of lines can be treated as a single broad line centered at 60 GHz. Snider and Westwater [1969] give the expression for the absorption below 45 GHz:

$$\begin{aligned}\alpha_{O_2} &= 2.6 \times 10^{-8} \nu^2 \left(\frac{P}{1013.25}\right)\left(\frac{293}{T}\right)^3 \\ &\times \left[\frac{1}{(\nu-\nu_o)^2+(\Delta\nu_{O_2})^2} + \frac{1}{(\nu+\nu_o)^2+(\Delta\nu_{O_2})^2} + \frac{1}{\nu^2+(\Delta\nu_{O_2})^2}\right] \\ &\times \Delta\nu_{O_2} \text{ cm}^{-1}\end{aligned} \tag{2.2.4}$$

where $\nu_o = 60$ GHz, P is the pressure in mbar, T is the temperature in Kelvins, ν is the frequency in GHz, and $\Delta\nu_{O_2}$ is the oxygen line half-width in GHz, which Snider and Westwater give as

$$\Delta\nu_{O_2} = 0.75\left(\frac{P}{1013.25}\right)\left(\frac{293}{T}\right)^{0.85} \tag{2.2.5}$$

For frequencies above 45 GHz, see Rozenkranz [1975].

Of more use to us than the absorption coefficient of oxygen is the opacity due to oxygen in the atmosphere, which is the integral of α_{O_2} through the atmosphere. This use is important because in the dual-frequency WVR algorithm presented in Section 2.3, we must apply an "oxygen correction" based on the opacity of oxygen. We will therefore integrate (2.2.4) with respect to height to determine the zenith opacity for an atmosphere which has the following properties: the atmosphere is in hydrostatic equilibrium with surface pressure P_o, and the temperature decreases linearly with constant lapse rate β ($\beta < 0$). The surface temperature is T_o, and the minimum value

of the temperature, 0 K, is reached at an altitude of $-T_\circ/\beta$. (We will discuss the effect of using this "special" atmosphere as opposed to a physically sensible atmosphere below.) Substituting (2.2.5) into the numerators of (2.2.4) yields

$$\alpha_{O_2} = 1.95 \times 10^{-6} \left(\frac{P}{1013.25}\right)^2 \left(\frac{293}{T}\right)^{3.85} \nu^2 K_\nu \text{ m}^{-1} \tag{2.2.6}$$

where K_ν represents the shape functions of (2.2.4); K_ν still depends on the line half-width, but only very weakly if we are many half-widths from the line center: K_ν varies by less than 5% for $0 \leq \nu \leq 45$ GHz and $0 \leq \nu \leq 2$ GHz. Thus, we may consider K_ν to be constant with height for the purpose of integration of (2.2.6).

To integrate (2.2.6) with respect to height and derive an expression for the zenith opacity, we must evaluate the integral

$$\int_0^\infty dz \left(\frac{P}{1013.25}\right)\left(\frac{293}{T}\right)^{3.85} \tag{2.2.7}$$

By combining the equation of hydrostatic equilibrium and the ideal gas law (see Hess [1959], for example), the pressure can be written as

$$P = P_\circ \left(\frac{T}{T_\circ}\right)^\gamma \tag{2.2.8}$$

where

$$\gamma = -\frac{M_d g}{R\beta} > 0 \tag{2.2.9}$$

In the above equation, we assume that the air is dry. This assumption will introduce a small bias in our result on the order of a few per cent, which we will see is negligible.

If we now substitute (2.2.8) into (2.2.6), and change the variable of integration by using $dT = \beta dz$, then after lengthy but straightforward algebra we arrive at

$$\tau_{O_2} = -\left(\frac{P_o}{1013.25}\right)^2 \left(\frac{293}{T_o}\right)^{2.85} \times 293\beta^{-1}(2\gamma - 2.85)^{-1}\nu^2 K_\nu \qquad (2.2.10)$$

If we use standard values of $M_d \simeq 29$ kg kmol^{-1}, $g \simeq 9.8$ m s^{-2}, $R \simeq 8314$ J (kmol K)$^{-1}$, and $\beta = -0.0065$ K m^{-1}, we find

$$\tau_{O_2} = 1.15 \times 10^{-2} \left(\frac{P_o}{1013.25}\right)^2 \left(\frac{293}{T_o}\right)^{2.85} \nu^2 K_\nu \text{ nepers} \qquad (2.2.11)$$

Equation (2.2.11) shows that we can predict the opacity due to oxygen for frequencies less than about 45 GHz for this special atmosphere if we know the pressure and temperature at the surface. How well does (2.2.11) predict the oxygen opacity for a "real" atmosphere? Up to about 11 km, the model for the temperature in our special model and in the U.S. Standard Atmosphere are identical, except that in our model the surface temperature may vary. Above this height only about 20% of the mass of the atmosphere remains, so one may take this number as an (extreme) upper limit to the error in (2.2.11). As we will see, the "oxygen correction" needed in the dual-frequency WVR algorithm has an amplitude of only a few millimeters, so an error of 20% amounts to about 1 mm at most.

In Section 2.8, we will discuss algorithms for WVR's which make use of a measurement of brightness temperature on the 60 GHz oxygen "line." Now, however, we will complete our discussion of absorption in the atmosphere.

2.2.iii Liquid water

Liquid water has a continuum spectrum due to absorption as well as to scattering in the microwave region. Goldstein [1951] gives expressions for both the absorption cross section σ_a and the scattering cross section σ_s for a single drop of water with radius a. The ratio of these cross sections is

$$\frac{\sigma_a}{\sigma_s} \propto \frac{(a^3/\lambda)}{(a^6/\lambda^4)} = (\lambda/a)^3 \tag{2.2.12}$$

Note that the scattering cross section used in (2.2.12) assumes Rayleigh scattering. Equation (2.2.12) tells us that absorption dominates over scattering for $\lambda \gg a$. For $\lambda = 1$ cm and $a = 1$ mm (2.2.12) gives $\sigma_a/\sigma_s \simeq 10^3$. Staelin [1966] used the data presented by Goldstein and determined

$$\kappa_\ell \rho_\ell = 10^{-6} \frac{\rho_\ell}{\lambda^2} e^{0.0281(18-T_c)} \quad \text{cm}^{-1} \tag{2.2.13}$$

where ρ_ℓ is the liquid water density in g m^{-3}, λ is in cm, and T_c is the temperature in °C. Because of the restriction on a/λ, (2.2.13) is not valid in cases where $a \simeq \lambda$, such as might occur in rain. For cases of clouds or fog or dense clouds, (2.2.13) begins to break down, and the algorithm presented in the next section is less accurate.

2.3 The dual-frequency algorithm

Having discussed briefly the frequency dependence of atmospheric absorption, we will in this section derive an expression for the wet path delay based on having available a two-frequency radiometer. Such a radiometer when used for the purpose of estimating the integrated water-vapor content is known as a water-vapor radiometer

(WVR). One of the channels of the WVR is centered at frequency ν_1, the other at ν_2. (In the following, whenever we omit the subscript indicating the frequency channel, then that equation holds for both channels.)

The total opacity can be written as

$$\tau = \tau_{O_2} + \tau_v + \tau_\ell \tag{2.3.1}$$

where the subscript O_2 indicates oxygen, v indicates water vapor, and ℓ indicates liquid water. As discussed in Section 2.2.*i*, it should be possible to model the O_2 opacity with some accuracy. However, it is very difficult, using surface-based measurements, to determine the liquid-water content of the atmosphere, because even on dry days there may be a large amount of liquid water suspended in clouds several kilometers in altitude. It is therefore desirable to eliminate the dependence on τ_ℓ. The total opacities at the two different frequencies can be combined in a manner which eliminates the liquid water term. Integrating (2.2.13) along the line–of–sight path in order to obtain the opacity, we find that

$$\tau_\ell \propto \nu^2 \tag{2.3.2}$$

where the constant of proportionality depends upon the profiles of temperature and liquid water, but not on frequency. The frequency dependence of (2.3.2) enables us to construct a liquid-independent "observable":

$$\tau_1 - \frac{\nu_1^2}{\nu_2^2}\tau_2 = \left(\tau_{v1} - \frac{\nu_1^2}{\nu_2^2}\tau_{v2}\right) + \left((\tau_{O_2})_1 - \frac{\nu_1^2}{\nu_2^2}(\tau_{O_2})_2\right) \tag{2.3.3}$$

where the subscripts "1" and "2" refer to the two frequency channels of the dual-frequency WVR. From (2.2.11) we have

$$\tau_{O_2} = f(P_\circ, T_\circ)\nu^2 K_\nu \tag{2.3.4}$$

where the function $f(P_\circ, T_\circ)$ depends only on the surface pressure and temperature but not on frequency, and K_ν is the frequency-dependent shape term, which has only a small dependence on temperature and pressure for ν below about 45 GHz. Thus (2.3.3) becomes

$$\tau_1 - \frac{\nu_1^2}{\nu_2^2}\tau_2 = \left(\tau_{v1} - \frac{\nu_1^2}{\nu_2^2}\tau_{v2}\right) + f(P_\circ, T_\circ)\nu_1^2[K_1 - K_2] \tag{2.3.5}$$

Rearranging gives

$$\tau_{v1} - \frac{\nu_1^2}{\nu_2^2}\tau_{v2} = \left(\tau_1 - \frac{\nu_1^2}{\nu_2^2}\tau_2\right) - f(P_\circ, T_\circ)\nu_1^2[K_1 - K_2] \tag{2.3.6}$$

From (2.1.3), the left-hand side of (2.3.6) can be written

$$\begin{aligned} \tau_{v1} - \frac{\nu_1^2}{\nu_2^2}\tau_{v2} &= \int_{atm} ds \left(\kappa_{v1} - \frac{\nu_1^2}{\nu_2^2}\kappa_{v2}\right)\rho_v \\ &= \int_{atm} ds\, \mathcal{W}(s)\, 10^{-6} R_v \left[k_2' + \frac{k_3}{T}\right]\rho_v \end{aligned} \tag{2.3.7}$$

where

$$\mathcal{W}(s) = 10^6 R_v^{-1}\left(k_2' + \frac{k_3}{T}\right)^{-1}\left(\kappa_{v1} - \frac{\nu_1^2}{\nu_2^2}\kappa_{v2}\right) \tag{2.3.8}$$

is the dual-frequency weighting function, as compared to the single-frequency weighting function of (2.1.6). Note that, from (2.2.1), the dual-frequency weighting function does not depend on the contribution from the tails of all the water-vapor lines above the

22.235 GHz line. Thus, any question as to the strength of these contributions is irrelevant.

If the assumption is made, as with the single frequency weighting function, that $\mathcal{W}(s)$ is constant, then (2.3.7) yields

$$\mathcal{W} \cdot L_w = \tau_1 - \frac{\nu_1^2}{\nu_2^2}\tau_2 - f(T_o, P_o)\nu_1^2[K_1 - K_2] \tag{2.3.9}$$

Solving for the wet path delay and using (2.3.4) then gives

$$L_w = \mathcal{W}^{-1}\left\{\tau_1 - \frac{\nu_1^2}{\nu_2^2}\tau_2 - (\tau_{O_2})_1\left[1 - \frac{K_2}{K_1}\right]\right\} \tag{2.3.10}$$

Equation (2.3.10) expresses a relationship between the total opacities at two frequencies and the wet propagation delay. As we mentioned in Section 2.1, the total opacity can be inferred from radiometric measurements. Thus, the dual-frequency algorithm would seem to be an improvement over the single-frequency algorithm in the sense that the single-frequency algorithm depends on knowing the opacity due solely to water vapor, which is not "directly" obtainable from a measurement at a single frequency. However, we have not yet discussed the "inherent" accuracy of the dual-frequency algorithm. Both the determination of the dual-frequency weighting function and the accuracy of the dual-frequency algorithm will be discussed in Section 2.6. In the next two sections, we will discuss effects of instrumental error on the estimates of the opacity.

2.4 Instrumental effects

In this section we will discuss several effects of instrumentation which can affect the accuracy of wet path delay "retrieval," as it is commonly called. We will not present

an exhaustive study of the instrumentation of radiometers, but instead include those sources of error which have been identified for the available radiometers. For the time being, we will discuss the accuracy of the WVR in terms of the accuracy with which the atmospheric brightness temperature or the opacity may be determined. Thus, we are for now discussing the *radiometer* part of WVR. We will assume that the two channels are equivalent. Later, when discussing the accuracy of the weighting function, we will put the accuracy of the WVR in terms of wet path delay.

In this section and in Section 2.5, we will separate the discussion into two parts. In the first part, we will discuss the effect of experimental error on the precision of determinations of atmospheric brightness temperature and opacity. We will assume that the radiometer is lossless and reflectionless, so that the radiometric temperatures are represented by physical temperatures of calibration loads. In the second part, we will discuss the effect on the accuracy of the determinations of brightness temperature and opacity of the losses and reflections of the radiometer, while assuming that the system is noiseless, so that there is no experimental error. Of course, both sources of error exist in a real radiometer, but they can be analyzed separately.

This section and the following section represent a review of some of the principles of radiometry, and do not contain any new insights on that subject. However, the discussion is geared toward a "user" of one of the several WVR's that are used today in conjunction with "geodetic" radio interferometry. These WVR's include:

- R-series. These instruments were constructed at the Jet Propulsion Laboratory (JPL) in the late 1970's, and are described in Resch *et al.* [1985], for example.

These WVR's are mechanically unreliable (having been built for R&D purposes), and most of them have been or are being, at this date, "retrofitted."

- R-series retrofit. These are the retrofit versions of the R-series which are mechanically more reliable and electronically more stable than the original R-series.
- Onsala WVR. This WVR was built at the Onsala Space Observatory in Sweden, where it has operated since late 1980. This WVR, described in Elgered [1983], is of slightly different internal design than the R-series. Some of these differences are discussed in this section.
- J-series. At the time of this writing, there is only one prototype J-series WVR, which was built in 1983-1985, also at JPL. This WVR is a somewhat different design than the R-series and the Onsala WVR, and much of the discussion in these sections will not apply to the J-series. See Janssen [1985] for a description of this instrument.

2.4.i Instrumental precision

Figure 2.4.1(a) shows a schematic of a simple radiometer. The signal, which represents the radiant energy from the sky, enters the antenna horn and passes through a series of components, which in Figure 2.4.1(a) have been all lumped together in the box marked "radiometer system." This box in fact represents such components as square law detectors, filters, mixers, and amplifiers. We will consider the effect of all these components together. After passing through the radiometer system, the analog

signal is transformed to a digital signal by an A/D converter, after which the resulting number is presumably recorded for use.

We now must ask, how do we relate the output number to the atmospheric brightness temperature? If the radiometer system and A/D converter are linear, then we can write

$$T_a = G^{-1} N_a + T_{off} \tag{2.4.1}$$

where T_a is the atmospheric brightness temperature, N_a is the resulting output number, and G and T_{off} are instrumental "constants" which may in fact vary with time and, quite frequently, with temperature. Linearity of the radiometer system is something which must be built into the radiometer design; in fact, today this linearity is not a very difficult requirement to meet.

A problem separate from the designing of a linear radiometer is the determination of G and T_{off}. There are several ways these quantities can be determined. The method which is illustrated in Figure 2.4.1(b) is that used in the R-series. This method consists of a switch placed between the antenna and the radiometer system. The switch allows the input to the radiometer to be varied between the antenna and either of two resistive calibration loads maintained at separate temperatures. For the R-series, the calibration loads consist of a "hot" load near 100° C ($\approx$ 370 K) and a "base" load near 40° C ($\approx$ 310 K). The instrumental "gain" G can be determined from the measurement of the temperatures and corresponding digital output ("counts") from these loads using

$$G = \frac{N_h - N_b}{T_h - T_b} \tag{2.4.2}$$

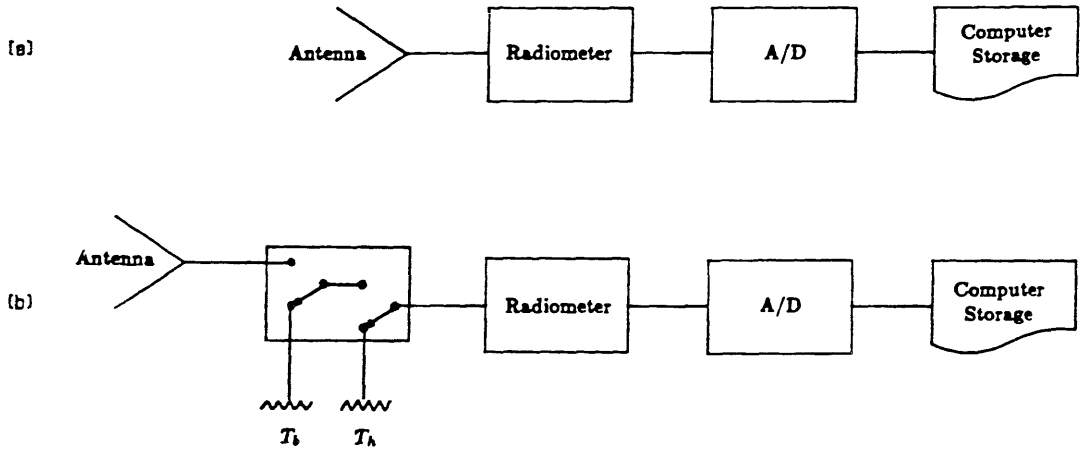


Figure 2.4.1.(a) Simple schematic of one channel of a WVR. (b) Same, except showing calibration loads.

where the subscript h refers to the hot load and b to the base load. The "offset temperature" T_{off} likewise can be determined using

$$T_{off} = T_b - G^{-1}N_b \tag{2.4.3}$$

The three equations (2.4.1), (2.4.2) and (2.4.3) allow us to calculate the effects of instrumental noise on the estimate of atmospheric brightness temperature. By using (2.4.2) and (2.4.3), we can write (2.4.1) in terms of the calibration loads:

$$T_a = \left(\frac{T_h - T_b}{N_h - N_b}\right)(N_a - N_b) + T_b \tag{2.4.4}$$

Thus the error in the antenna temperature σ_{T_a} is related to the instrumental errors in the measurement of the temperatures of the calibration loads and the output counts from the calibration loads and the antenna (all assumed Gaussian white noise and uncorrelated) by

$$\begin{aligned}\sigma_{T_a}^2 \simeq \frac{\sigma_{N_a}^2}{G^2} + (T_h - T_b)^{-2}\Bigg[&(T_a - T_b)^2\left(\sigma_{T_h}^2 + \frac{\sigma_{N_h}^2}{G^2}\right) \\ &+(T_a - T_h)^2\left(\sigma_{T_b}^2 + \frac{\sigma_{N_b}^2}{G^2}\right)\Bigg]\end{aligned} \tag{2.4.5}$$

Note that the second term on the right-hand side of (2.4.5) is inversely proportional to the square of the separation of the temperatures of the two calibration loads. For typical atmospheric brightness temperatures (between 10 K and 100 K), and for the values for T_b and T_h given above, this second term is much larger than the first term. Table 2.4.1 gives values for the various terms in (2.4.5) based on typical values from the R-series WVR's. With these values, (2.4.5) becomes

$$\sigma_{T_a} \simeq 0.15\Big[1.0 + 4.0 \times 10^{-4}\left[(T_a - 310)^2 + (T_a - 370)^2\right]\Big]^{\frac{1}{2}} \tag{2.4.6}$$

where both T_a and σ_{T_a} are in Kelvins.

In Figure 2.4.2, the solid line shows σ_{T_a} versus T_a in the region $0 < T_a < 100$ K, for the R-series. The dashed line represents the error in the determination of the brightness temperature were the hot load to be replaced by a cold load operating near 80 K. We can see that σ_{T_a} is nearly five times smaller for the case of the cold–base pair than for the case of the hot–base pair. From (2.4.5) we can see that this situation arises from two contributions: (*i*) the increase in the factor $T_h - T_b$ (which for the cold load may be written $T_c - T_b$), and (*ii*) the decrease in the factor $T_a - T_h$, which is now $T_a - T_c$. Such a cold load is built into the design of the Onsala WVR. The precision of this radiometer is about 0.2 K, while the precision of the R-series is about 1 K. In Section 2.5, we will show how the precision of the R-series WVR's can be improved.

2.4.ii Instrumental accuracy

In writing (2.4.4), we assumed that the gain defined in (2.4.2) was the gain of the signal propagating from the radiometer antenna through the switch and the following system (see Figure 2.4.1). In fact this is not the case. The waveguides transmitting the power from the loads to the radiometers will be lossy, and there will be reflections at junctions, *e.g.*, at the switch in Figure 2.4.1(b). Another problem is that the thermistors used to measure the temperatures of the calibration loads will have been incorrectly calibrated, mainly due to instrumental errors during calibration. In order to allow us to examine the effects of these errors, let us denote the radiometric

Table 2.4.1

R-Series Parameters

Parameter	Typical value (R-series)
T_b	310 K
T_h	370 K
$\sigma_{T_h} = \sigma_{T_b}$	0.1 K
$\sigma_{N_h} = \sigma_{N_b} = \sigma_{N_a}$	1–3 "counts"
G	6–10 counts K^{-1}

Uncertainty in "counts" depends on integration time, which is software selectable in steps of 0.1 sec.

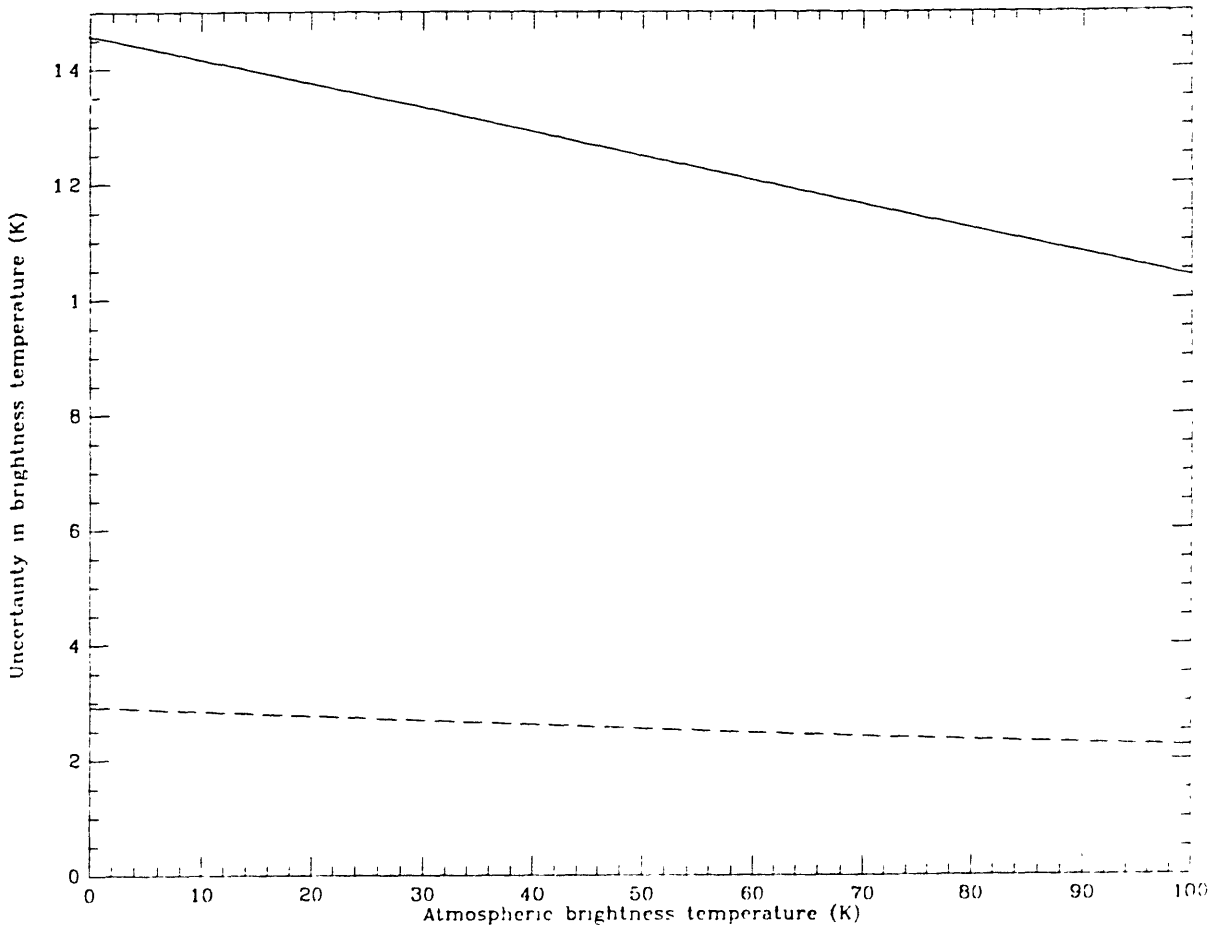


Figure 2.4.2. Atmospheric brightness temperature error (1-σ) *vs.* brightness temperature for a hot–base calibration scheme (solid line), and for a cold–base calibration scheme (dashed line).

temperature at the output of the switch with an asterisk. Then the output N_a when the switch is connected to the antenna can be written

$$N_a = G\,(T_a^* - T_b^*) + N_b \tag{2.4.7}$$

where G is the instrumental gain of the signal at the output of the switch, and is given by

$$G = \frac{N_h - N_b}{T_h^* - T_b^*} \tag{2.4.8}$$

The "gain at the output of the switch" means the gain of the signal after it leaves the switch propagating through the rest of the radiometer, including the A/D converter. A "gain" may be defined from any point in the radiometer, for instance, from the input to the antenna. This gain is related to the gain defined in (2.4.8) by $G' = G_a G$, where G' is the gain from the input to the antenna, and G_a is the gain of the signal from the input to the antenna to the output of the switch. The gain at the output of the switch is a convenient quantity to use since the path from the output of the switch through the radiometer is the same for the signals from the antenna and both calibration loads. However, the path from the antenna and each of the loads are all physically different paths, and so the gains from the source of the signal to the output of the switch may be different.

Let us introduce the terms ϵ_a, ϵ_h, and ϵ_b, defined by

$$\begin{aligned} T_a^* &= T_a + \epsilon_a \\ T_b^* &= T_b + \epsilon_b \\ T_h^* &= T_h + \epsilon_h \end{aligned} \tag{2.4.9}$$

Here T_h and T_b are the temperatures reported by the thermistors, which we will mistakenly use to calculate the brightness temperature, and T_a is the true atmospheric brightness temperature. (In this analysis, the temperatures reported by the thermistors will be the true physical temperatures of those loads. That is, we will not yet consider errors in the thermistors.) We will now calculate the error in the estimated atmospheric brightness temperature incurred by assuming that the ϵ_i ($i = a, b, h$) are zero, *i.e.*, that the radiometer is lossless and reflectionless. The quantity we calculate for the atmospheric brightness temperature will be in error by an amount δT_a, which to first order in the ϵ's is

$$\delta T_a \simeq \epsilon_a + \left(\frac{T_b - T_a}{T_h - T_b}\right)\epsilon_h - \left(\frac{T_h - T_a}{T_h - T_b}\right)\epsilon_b \tag{2.4.10}$$

What value can we expect for δT_a? The following analysis follows Corey [1983] and Elgered [1985]. The radiometric temperature at the output of the switch can be related to the true physical temperature (in the case of the calibration loads) or to the true sky brightness temperature (in the case of the antenna) by

$$T_i^* = \beta_i T_i + \Delta T_i \quad (i = a, b, h) \tag{2.4.11}$$

where the β_i represent ohmic and other losses, and the ΔT_i represent additive noise due to reflections and losses at junctions. Using (2.4.11), we can write the ϵ_i of (2.4.9) in terms of the β_i and ΔT_i as

$$\epsilon_i = (\beta_i - 1)T_i + \Delta T_i \tag{2.4.12}$$

For a waveguide in which the effects of attenuation dominate those of reflection (a state which can be achieved with decent engineering), then the additive noise ΔT_i is given by [Evans and McLeish, 1977]

$$\Delta T_i \simeq (1 - \beta_i) T_i^p \tag{2.4.13}$$

where T_i^p is the physical temperature of the transmission line (waveguide). With (2.4.13), (2.4.12) becomes

$$\epsilon_i \simeq (\beta_i - 1)(T_i - T_i^p) \tag{2.4.14}$$

For the R-series WVR's (including the retrofits), the physical temperatures T_a^p, T_b^p, and T_h^p can all be replaced by the base-load temperature because the base-load temperature is equal to the internal temperature to within about 10° C. Typical examples of base-load and internal temperatures are shown in Figure-2.4.3. Substituting the resulting expressions for the ϵ_i into (2.4.10) yields

$$\delta T_a \simeq \beta_a \left(\frac{\beta_h}{\beta_a} - 1 \right) (T_b - T_a) \tag{2.4.15}$$

According to Corey [1983; private communication, 1986], for the ~21 GHz channel of the R-series, the antenna signal propagates through roughly 30 cm more of silver-plated waveguide than the hot-load signal, for a relative attenuation compared to the hot-load signal of 0.2–0.3 dB. The antenna signal also encounters one more ferrite switch than the hot-load signal, for a relative attenuation of ~ 0.2 dB. These two differences lead to a relative attenuation of $\beta_a/\beta_h \simeq 0.9$. Then using $\beta_a \simeq 1$ in (2.4.15), we find that for T_a between 0 and 100 K, we have δT_a between 15 K and

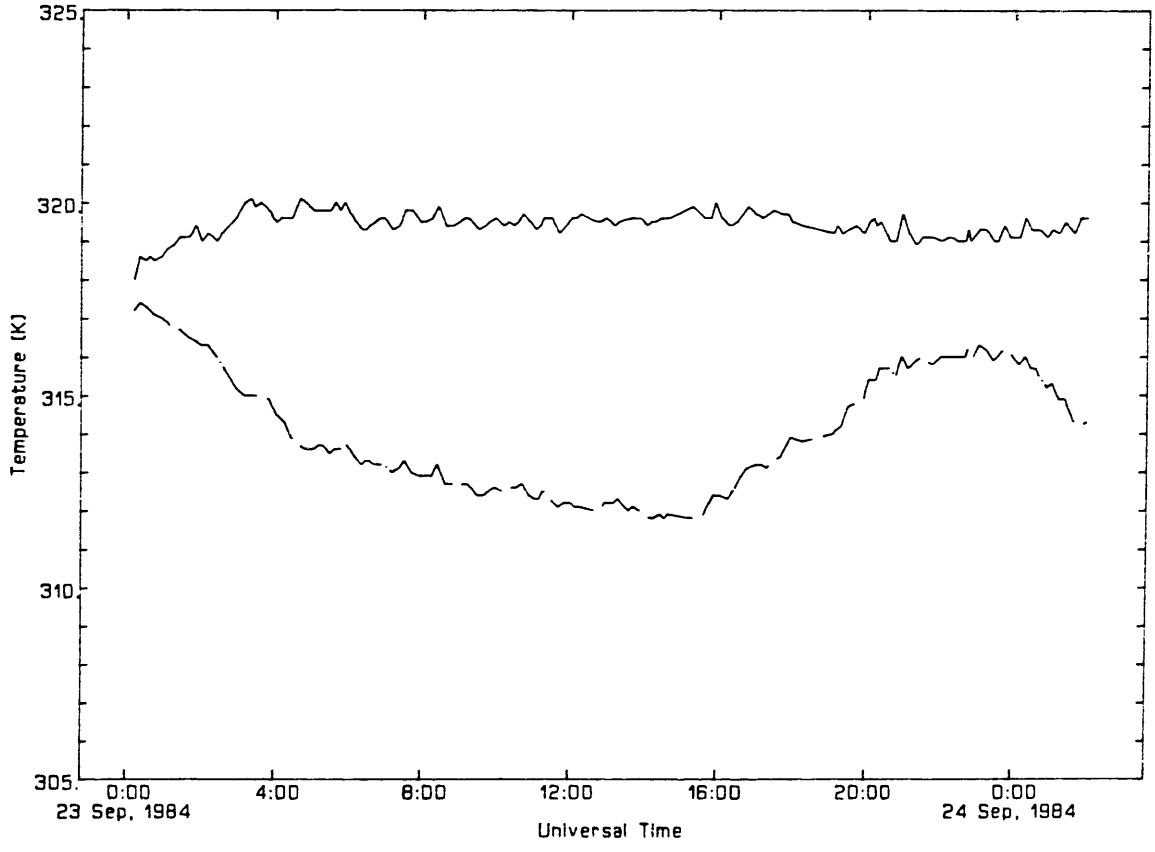


Figure 2.4.3. Typical time series showing base-load temperatures (dash-dot) and internal WVR temperatures (solid) for the R-series.

20 K. Since typical zenith atmospheric brightness temperatures are 10–30 K, this error represents possibly a 150% error in the measurement of brightness temperature.

Obviously, if the various ϵ_i are not known, the calibration of the WVR by use of the internal calibration loads is totally inadequate. It is possible to determine these parameters. One problem is that they are temperature-dependent, and so the temperatures of the waveguides must be monitored at all times. Another possibility is to parametrize the temperature of the waveguides using the internal temperature. However, since a much easier (and cheaper) method of calibrating the WVR's is available, no one has ever attempted to determine the ϵ_i. This easier method of calibration is discussed in the next section.

2.5 Calibration of the WVR

In this section we will discuss an alternate means of calibrating the radiometers which make up the WVR's. This method is the tip-curve method, and it yields a more precise calibration of the WVR because in effect it makes use of the atmospheric brightness temperature itself as a calibration source. In a tip curve, the radiometer takes several measurements at different elevations. In terms of the quantities we have already defined, the output N_{ai} of the radiometer at the elevation ϵ_i can be modeled as

$$N_{ai} = G\left[T_{bg}e^{-\tau(\epsilon_i)} + T_{eff}\left(1 - e^{-\tau(\epsilon_i)}\right) - T_b^{obs}\right] + N_b^{obs} + \varepsilon_i \qquad (2.5.1)$$

where $\tau(\epsilon_i)$ is the opacity in the direction of elevation ϵ_i, and ε_i is the random measurement error, which we will assume is zero mean, white-noise Gaussian, and with variance independent of elevation. The superscript *obs* on the base load temperature

and radiometer output indicates that these quantities are observations, subject to experimental error, which, as in the previous section, is assumed zero-mean, white-noise Gaussian. Also, for purposes of analyzing the effects of experimental error, we have assumed that $T_i = T_i^*$ $(i = a, b, h)$. Note that (2.5.1) *per se* requires no assumptions concerning the structure of the atmosphere; there may be a directional dependence of T_{eff}, defined in (2.1.8), or likewise an azimuthal dependence of τ. We will later assume, however, that the gain G is constant for the duration of the tip curve, as are T_b and N_b.

The usefulness of (2.5.1) for calibrating the WVR comes from the assumptions that (*i*) T_{eff} is known, and (*ii*) the opacity can be written as (*c.f.* Section 1.3.*ii*)

$$\tau(\epsilon) = \tau^z m(\epsilon) \tag{2.5.2}$$

where τ^z is the zenith opacity and $m(\epsilon)$ is the mapping function for the opacity. (We will discuss the effect of an error in the assumption (*i*) in Section 2.6.) Since all of the WVR's we consider are limited to angles of elevation above about 20°, we will for the moment assume that

$$m(\epsilon) = \csc \epsilon \tag{2.5.3}$$

The zenith opacity and the gain can now be treated as parameters and estimated by, *e.g.*, weighted least squares. In principle, it is possible also to treat the effective temperature and background temperature as parameters to be estimated. However in our experience, the correlations among the parameters become so high as to render meaningless the estimates of all the parameters. The reason for this correlation can be seen from the expansion of the exponential terms of (2.5.1) after replacing τ with

(2.5.2). If we ignore the small background radiation term, then to first order in τ^z the radiometer output depends only on the product of T_{eff} and τ^z.

In the following two sections, we will parallel our discussions in Section 2.4 of the effects of instrumental errors on the precision and accuracy of brightness temperature determinations. In this section, however, we will examine these effects in the context of calibration done with tip curves.

2.5.i Instrumental precision

In order to determine the precision with which the zenith opacity and gain may be obtained from a tip curve, let us perform a simple covariance analysis. For this analysis, we will assume that T_{bg} and T_{eff} are known. (The effects of errors in these quantities are discussed in Section 2.6.) Our observation equation will be (2.5.1). We will assume that one reading from the calibration loads is taken during the tip curve, and that the gain and zenith opacity are constant over the duration of the tip curve. In order to take into account the measurement uncertainties of the calibration load readings, we treat T_b and N_b as parameters to be estimated, and add the "observation equations"

$$T_b^{obs} = T_b + \varepsilon_{T_b}$$
$$N_b^{obs} = N_b + \varepsilon_{N_b} \tag{2.5.4}$$

The values for the various parameters used to perform the covariance analysis are shown in Table 2.5.1. If the observation equations (2.5.1) were linear in the parameters to be estimated, then the results of the covariance analysis would be independent of the values of the parameters; since the equations are not linear, typical values for G

and τ^z are shown in Table 2.5.1 and are used in this analysis. We will assume we are making tip-curve observations at air masses of 1, 2, and 3 (elevation angles of 90°, 30°, and 19.5°).

The results of the covariance analysis are shown in Table 2.5.2. These results are not immediately comparable with the results of the instrumental error analysis from the last section, because in the last section we found the error associated with the line-of-sight brightness temperature, and in this section we have found the error associated with the zenith delay. In order to compare the results, let us first calculate the error in the line-of-sight opacity due to the error in the line-of-sight brightness temperature. This calculation can be done by using the equation of radiative transfer to solve for the opacity:

$$\tau = -\log\left(\frac{T_a - T_{eff}}{T_{bg} - T_{eff}}\right) \tag{2.5.5}$$

so that

$$\sigma_\tau \simeq \frac{\sigma_{T_a}}{|T_a - T_{eff}|} = \frac{e^{m\tau^z}}{|T_{bg} - T_{eff}|}\sigma_{T_a} \tag{2.5.6}$$

On the other hand, we can obtain the line-of-sight opacity from the tip-curve solution using (2.5.2). Thus for the tip curve we have

$$\sigma_\tau(m) = m\sigma_{\tau^z} \tag{2.5.7}$$

where we have expressed the error in the line-of-sight opacity as a function of air mass.

Using (2.4.6) for σ_{T_a} in (2.5.6), and substituting the value for σ_{τ^z} from the covariance analysis into (2.5.7), we have

$$\sigma_\tau \simeq \begin{cases} 5.7 \times 10^{-4} e^{m\tau^z}\{1 + 0.0004\left[(T_a(m) - 310)^2 + (T_a(m) - 370)^2\right]\}^{\frac{1}{2}}, & \text{(loads)} \\ 4.6 \times 10^{-4} \cdot m, & \text{(tip curve)} \end{cases} \tag{2.5.8}$$

Table 2.5.1

Values used in covariance analysis

Parameter	Value used
T_b	310 K
N_b	3100 "counts"
σ_{T_b}	0.1 K
$\sigma_{N_b} = \sigma_{N_a}$	1.5 "counts"
G	10 counts K^{-1}
τ^z	0.05 nepers
T_{bg}	2.7 K
T_{eff}	265 K

Table 2.5.2

Results of covariance analysis

Parameter	Uncertainty
T_b	0.1 K
N_b	1.5 counts
G	9.92×10^{-3} counts K^{-1}
τ^z	4.55×10^{-4} nepers

The correlation matrix is found to be

$$\begin{array}{c} \\ G \\ \tau^z \\ T_b \\ N_b \end{array} \begin{array}{c} \begin{array}{cccc} G & \tau^z & T_b & N_b \end{array} \\ \begin{pmatrix} 1.0000 & & & \\ 0.7029 & 1.0000 & & \\ -0.3268 & 0.0394 & 1.0000 & \\ 0.4902 & -0.0591 & 0.0000 & 1.0000 \end{pmatrix} \end{array}$$

These two expressions are plotted as a function of elevation angle in Figure 2.5.1. It can be seen that the error in the opacity is always less for the tip-curve calibrated WVR.

The values obtained for the standard deviations represent a "lower limit" in terms of radiometer stability and weather conditions, corresponding approximately to a standard deviation of the zenith delay of 1 cm (internal calibration) and 0.1 cm (tip-curve calibration). The standard deviation used for the output radiometer is actually the sample standard deviation of the average value of the radiometer output for n integration periods of 0.1 sec, where n is user-selectable. The choice of the value of n reflects a balance between the stability of the radiometer and the weather. Choosing a large value for n may not necessarily decrease the standard deviation of the average if the instrument is unstable or the weather is unstable. For the R-Series WVR's, which are very unstable, we have found that $n = 50$ seems to yield standard deviations of between 1.5 and 4. Furthermore, the resulting integration time seems to be short enough that instrumental effects dominate atmospheric effects. We base this conclusion on our experience that the standard deviations of the output of the radiometer is about the same regardless of whether the input is the antenna or either calibration load. If the dominating effect were atmospheric, then the standard deviation of the output from the antenna would be greater than those from the calibration loads. Furthermore, the standard deviation of the output of the antenna would be elevation-angle dependent, a

behavior we see no evidence for. Therefore, both curves of Figure 2.5.1 may be scaled depending on the actual standard deviations.

In the following, we discuss the effect of instrumental accuracy on calibration by tip curves.

2.5.ii Instrumental accuracy

Analogously to Section 2.4.*ii*, we can evaluate the effect on the tip curve results of the radiometer-sensed temperatures being not equal to the physical temperatures. A simple way of evaluating this effect is by assuming $T_a \simeq mT_{eff}\tau^z$, which is an approximation of the equation of radiative transfer for small opacity, and by assuming $T_{bg} \simeq 0$. (These approximations will affect our interpretation of the results very little, if at all.) If we were to assume (erroneously) that the $\epsilon_i = 0$, we would write (2.5.1) as

$$N_a = G(mT_{eff}\tau^z - T_b) + N_b \tag{2.5.9}$$

where G is the gain of the signal at the output of the switch. If our assumption concerning the ϵ_i were correct, G would equal the gain of the signal at the entrance to the antenna. For illustration we now assume that the experimental errors are zero. From (2.5.9), we find that our tip-curve analysis will be equivalent to fitting a straight line to a plot of N_a *vs.* m, and determining $\hat{G}$ and $\hat{\tau}^z$, our estimates of the gain and zenith opacity, from

$$\begin{aligned} slope &= \hat{G}T_{eff}\hat{\tau}^z \\ intercept &= -\hat{G}T_b + N_b \end{aligned} \tag{2.5.10}$$

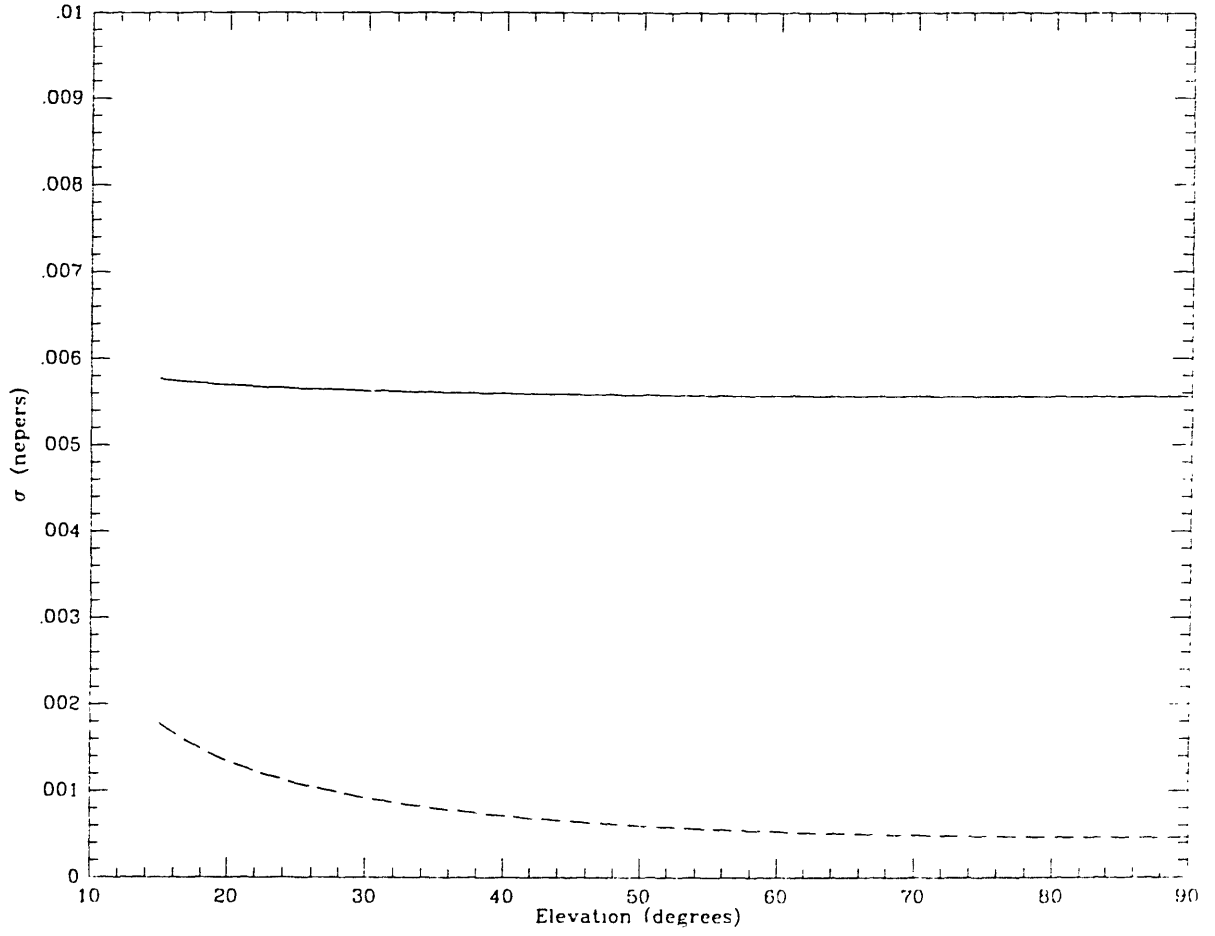


Figure 2.5.1. Error in opacity (1-σ) *vs.* elevation, for determination of opacity by internal hot–base calibration (solid) or tip-curve (dashed). For conversion to delay error, see section *2.6.iv*.

However, in reality, due to the losses and reflections of the signal in each of the channels incurred prior to the switch, the correct observation equation, from (2.4.7) and (2.4.9), is

$$N_a = G(mT_{eff}\tau^z + \epsilon_a - T_b - \epsilon_b) + N_b \tag{2.5.11}$$

If we use (2.4.12) to express ϵ_a, and assume as in Section 2.4 that the physical temperatures could be approximated by the base load temperature, then we can express (2.5.11) as

$$N_a = G(m\beta_a T_{eff}\tau^z - \beta_a T_b) + N_b \tag{2.5.12}$$

where m is again the air mass. Therefore, in the absence of noise the slope and intercept will be related to the true gain G and true zenith opacity τ^z by

$$\begin{aligned} slope &= G\beta_a T_{eff}\tau^z \\ intercept &= -G\beta_a T_b + N_b \end{aligned} \tag{2.5.13}$$

We can solve (2.5.10) and (2.5.13) simultaneously to find the expression for the estimated zenith opacity and gain in terms of the true opacity and gain. We then find

$$\begin{aligned} \hat{G} &= \beta_a G \\ \hat{\tau}^z &= \tau^z \end{aligned} \tag{2.5.14}$$

Thus the estimate of the zenith opacity is not affected by the temperature "errors." In fact, we have ignored certain factors in reaching (2.5.14) which virtually rule out having zero error (even for a noiseless radiometer). The values T_b and T_h which we have taken for the calibration load temperatures are reported to us from thermistors located near the corresponding load. Heretofore, we have assumed that the readings from the thermistors were perfect, so that the T_b and T_h reflected the true radiometric

temperatures of the loads. This in fact is nearly impossible to assure, primarily because the thermistors themselves are calibrated only to the tenth or few hundredths of a degree. We have also assumed that for the base load the attenuation is exactly canceled by the added noise; in reality this cancellation is never perfect. In practice, it is very difficult to achieve standard errors in brightness temperature measurements of less than a few tenths of one Kelvin [A.K. Wu, Bendix Field Engineering Corporation, personal communication, 1985].

Let us for a moment look again at the estimate of the gain made only from the physical temperatures of the calibration loads. Such an estimate has come to be known as the "raw" gain. In terms of the true gain and the "errors" ϵ_b and ϵ_h, the raw gain can be written

$$\begin{aligned} G_{raw} &= \frac{N_h - N_b}{T_h - T_b} \\ &= \frac{N_h - N_b}{(T_h^* - T_b^*)} \left(1 - \frac{\epsilon_h - \epsilon_b}{T_h^* - T_b^*}\right)^{-1} \\ &\simeq G\left(1 + \frac{\epsilon_h - \epsilon_b}{T_h - T_b}\right) \end{aligned} \tag{2.5.15}$$

Using (2.5.15) and (2.5.14), we can introduce a new quantity γ which is defined to be the the ratio of the gain estimated from a tip curve to the raw gain:

$$\gamma \equiv \frac{\hat{G}}{G_{raw}} \simeq \beta_a \left(1 + \frac{\epsilon_h - \epsilon_b}{T_h - T_b}\right)^{-1} \tag{2.5.16}$$

If we again make the assumption that the physical temperatures of the waveguides can be represented by the base load temperature, then using (2.4.14) we can show that

$$\frac{\epsilon_h - \epsilon_b}{T_h - T_b} \simeq \beta_h - 1 \tag{2.5.17}$$

which substituted into (2.5.16) yields

$$\gamma \simeq \frac{\beta_a}{\beta_h} \tag{2.5.18}$$

Thus γ depends only on the attenuations of the hot load and antenna signals. This quantity is in principle a constant for the instrument which can be estimated easily by performing a tip curve and measuring the raw gain simultaneously. In Section 2.4, we assumed that $\gamma \simeq 0.9$. In fact, γ has been measured for several radiometers of the upgraded R-series (each WVR yields two values of γ, one for each channel) by the method described here. The measured values range from about 0.85 to 0.95 [M.W. Hayes, Interferometrics Inc., private communication, 1985]. Time series of γ from a single radiometer channel show time variabilities of 2–3% of the corresponding value of γ, for all radiometer channels, for time scales of several months. This variability is probably due to small drifts in the calibration of the thermistors measuring the temperatures of the calibration loads.

How can we make use of the knowledge of γ? Suppose that we have at some time previously performed a tip curve, or many tip curves, and have an estimate, or an average of estimates, of γ. It is some time later, and we have made a "line-of-sight" measurement which means we have measured N_a with the WVR pointed in one specific direction. We have also the calibration data T_b, T_h, N_b, and N_h. The four calibration quantities allow us to calculate the raw gain G_{raw}. Furthermore, our knowledge of γ allows us to calculate a quantity which we will call the "system gain" G_{sys}

$$G_{sys} = \gamma G_{raw} \tag{2.5.19}$$

From (2.5.18), we can see that G_{sys} is the gain we would have measured had we performed a tip curve. Thus, from (2.5.14), we have

$$G_{sys} = \beta_a G \tag{2.5.20}$$

where as before, G is the gain at the output of the switch.

We wish to use our data to estimate the atmospheric brightness temperature T_a, but so far we have no good estimate of the gain. Instead, we have only a "close" estimate of the gain which we have called the system gain, which is in error by the factor β_a. Suppose, though, that we use the system gain along with T_b to calculate an estimate of the atmospheric brightness temperature which we will call $\hat{T}_a$:

$$\hat{T}_a = G_{sys}^{-1}(N_a - N_b) + T_b \tag{2.5.21}$$

Equation (2.5.21) might be thought of as being very naive. On the one hand, it assumes that $G_{sys} = G$, which we know is not true. On the other hand, this equation assumes that $T_a = T_a^*$, which from the analysis of the previous section we know is wildly inaccurate. The cumulative effect of these two assumptions can be easily calculated:

$$\begin{aligned}
\hat{T}_a &= (\beta_a G)^{-1}(N_a - N_b) + T_b \\
&= \beta_a^{-1}(T_a^* - T_b^*) + T_b \\
&= \beta_a^{-1}(\beta_a T_a + (1 - \beta_a)T_b - T_b) + T_b \\
&= \beta_a^{-1}(\beta_a T_a - \beta_a T_b) + T_b \\
&= T_a
\end{aligned} \tag{2.5.22}$$

The errors in the two assumptions compensate. This compensation in fact is not totally unexpected because, as we pointed out earlier, G_{sys} is the gain we would have found had we performed a tip curve, and we earlier showed that the estimate of the zenith opacity was unaffected even though the tip-curve-estimated gain was in error by a factor β_a. The basic reason for the compensation is that the estimate of the gain was obtained by a tip curve, and hence by actually observing the sky. The estimate of the gain is therefore in error by a factor which represents the attenuation of the antenna signal. When that gain is used with the radiometer output, that output is "boosted" by the amount which it was originally attenuated, *i.e.*, by the inverse of the antenna attenuation. Of course, as discussed above, the compensation is really not perfect, and in practice it is very difficult to reduce the standard error of the brightness temperature to less than a few tenths of one Kelvin.

We have mentioned two techniques for determining γ, which is the ratio $\frac{\beta_a}{\beta_h}$. The first technique involved estimating the excess attenuation of the antenna signal compared to the hot-load signal, based upon the known materials and lengths of the wave guides. The second method involved taking the ratio of the tip-curve estimated gain to the "raw" gain. The estimates of $\frac{\beta_a}{\beta_h}$ obtained by these two methods agreed well, with radiometer-dependent variations of about 5% from a nominal value of 0.9. A third method of determining this ratio can be found by observing that (2.5.1), which used the base load as a reference load, still holds if T_b is replaced by T_h and N_b by N_h. If we perform the analysis leading up to (2.5.4) with the base load replaced by the hot load, we find

$$\hat{\tau}_h^z \simeq \left[1 - \left(1 - \frac{\beta_a}{\beta_h}\right)\left(1 - \frac{T_b}{T_h}\right)\right] \tau^z \qquad (2.5.23)$$

where $\hat{\tau}_h^z$ is the estimate of the zenith opacity obtained from a tip-curve analysis using the hot load as a reference. Denoting the estimate of the zenith opacity obtained by using the base load as $\hat{\tau}_b^z$ we find

$$\Delta\hat{\tau}^z \equiv \hat{\tau}_h^z - \hat{\tau}_b^z \simeq \left(1 - \frac{\beta_a}{\beta_h}\right)\left(1 - \frac{T_b}{T_h}\right)\tau^z \tag{2.5.24}$$

Solving for $\frac{\beta_a}{\beta_h}$, we find

$$\frac{\beta_a}{\beta_h} \simeq 1 - \left(1 - \frac{T_b}{T_h}\right)^{-1}\frac{\Delta\hat{\tau}^z}{\tau^z} \tag{2.5.25}$$

Figure 2.5.2 shows a series of estimates of $\frac{\beta_a}{\beta_h}$ taken from the 20.7 GHz channel of the WVR known as R-05 (fifth WVR from the R-series). Because we do not know the zenith delay τ^z, the averages of $\hat{\tau}_b^z$ and $\hat{\tau}_h^z$ were used for τ^z in (2.5.25). The standard deviations of the estimates have not been shown in Figure 2.5.2 because they are difficult to calculate since $\hat{\tau}_b^z$ and $\hat{\tau}_h^z$ are correlated. However, from the short term scatter of the estimates we can see that the standard deviation is less than 0.01.

The estimates of $\frac{\beta_a}{\beta_h}$ from Figure 2.5.2 agree quite well with the estimates made using the two methods previously discussed. A systematic trend, however, is evident from this figure. This trend is probably due to our assumption concerning the equality of the physical temperatures of the wave guides and the temperature of the base load. In Figure 2.5.3 we show the temperature of the base load as a function of time. The same systematic behavior is evident in this figure as in the previous figure, indicating that these behaviors are correlated.

Which of the three methods presented for the determination of $\frac{\beta_a}{\beta_h}$ is the "best" method? One of the problems with the old R-Series WVR's was that they were unstable and the value of this ratio could change with time. However, this variability did

not pose a problem if we used a tip-curve to determine the gain for every observation, and due to the imprecision of calibration via the internal loads such a procedure was necessary. Of course, using such a procedure, we do not need to know the value of $\frac{\beta_a}{\beta_h}$. A different situation exists with the upgraded R-Series, which are stable enough that calibration may be obtained using smoothed or averaged values of $\frac{\beta_a}{\beta_h}$ and the "raw" gain, as discussed above. In this case, the best method for obtaining $\frac{\beta_a}{\beta_h}$ is from the ratio method expressed by (2.5.16). The two other methods which we used are approximate methods which we included only to compare with this more accurate method.

2.6 Accuracy of the dual-frequency algorithm

In the preceding sections, we have examined the effects of errors in the WVR calibration as well as several instrumental effects. In this section we will look at the effects of other errors, including errors in obtaining the numerical value of the weighting function. We will begin by reviewing the process by which one obtains delays from WVR measurements. We will assume that the instrumental errors are zero, so that we have perfect estimates of the gain, and the calibration load parameters. The brightness temperature for each channel is determined by using (2.4.7) in the form

$$T_a = G^{-1}(N_a - N_b) + T_b \tag{2.6.1}$$

In practice, we will probably determine the raw gain using the calibration information, and then use the known value of γ to obtain G_{sys}, which we will use in place of G in (2.6.1). However, as we have shown, one still obtains the correct value of the

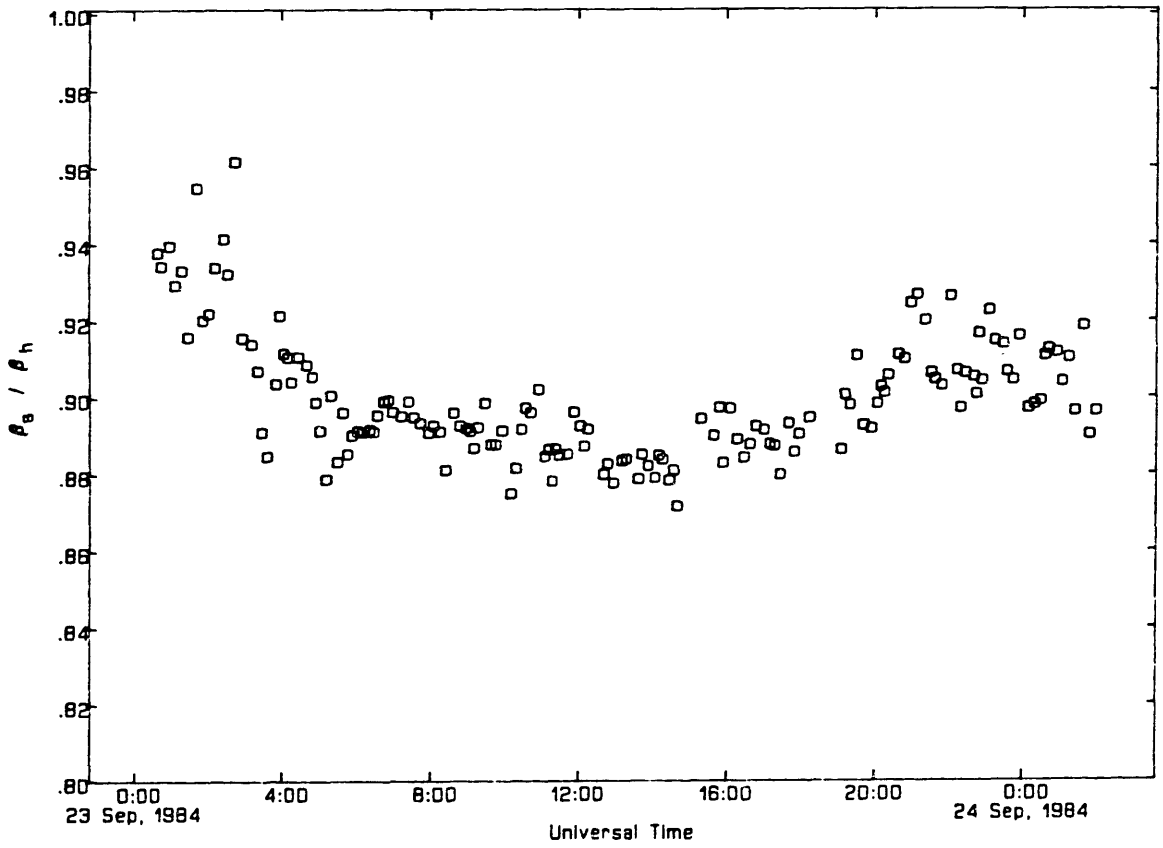


Figure 2.5.2. Estimates of $\frac{\beta_a}{\beta_h}$ for the 20.7 GHz channel, determined by differencing estimates of the zenith opacity obtained from tip-curve analyses using the hot load and the base load for reference loads (see text).

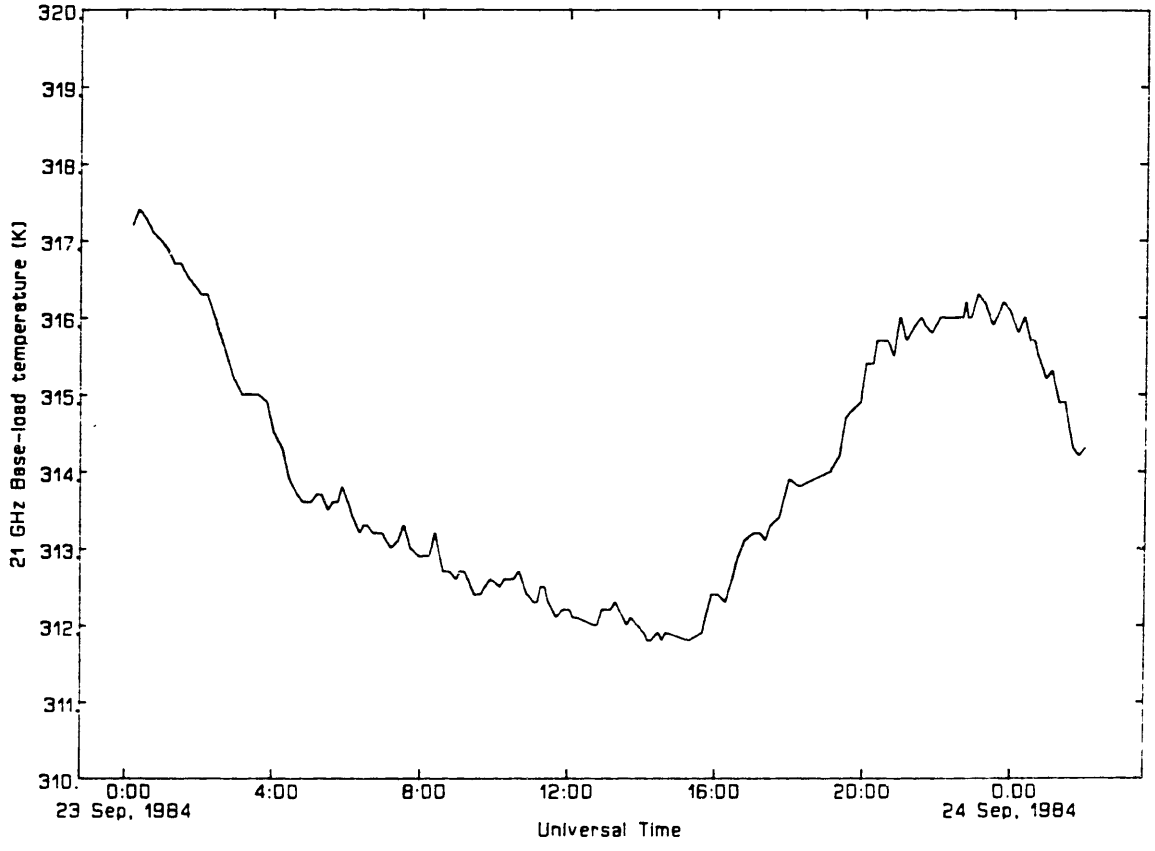


Figure 2.5.3. Base-load temperatures corresponding to Figure 2.5.2.

brightness temperature. The equation of radiative transfer (2.1.7) is then used to obtain the opacity τ in each channel

$$\tau = -\log\left(\frac{T_a - T_{eff}}{T_{bg} - T_{eff}}\right) \tag{2.6.2}$$

Alternatively, a tip curve may be performed, and the opacity in the direction of interest obtained from the zenith opacity. However, the former procedure is preferred if possible since the possible inhomogeneous distribution of water vapor makes the prediction of a "line-of-sight" opacity from the zenith opacity less sure.

The opacities from the two channels are then combined to yield the wet path delay, using a weighting function which has been determined previously:

$$L_w = \mathcal{W}^{-1}\left[\tau_1 - \frac{\nu_1^2}{\nu_2^2}\tau_2 - (\tau_{O_2})_1\left(1 - \frac{K_2}{K_1}\right)\right] \tag{2.6.3}$$

We will now describe the usual method for obtaining the dual-frequency weighting function $\mathcal{W}$. There are in fact two methods, both of which involve the use of radiosonde data. A radiosonde is a meteorological-sensor package carried aloft by a balloon. The sensors report the pressure, temperature, relative humidity, horizontal wind speed direction, and current height (usually derived from the pressure) for a number of epochs as the balloon is rising, until the balloon bursts.

Both methods for determining the dual-frequency weighting function $\mathcal{W}$ have the following procedure in common:

Given a set of radiosonde-launch data, it is possible to construct for each launch a profile of pressure P, temperature T, and relative humidity RH. Although the

radiosonde drifts horizontally, a "simulated" vertical path must be created; the errors due to the radiosonde drift are discussed below. Each profile is used to compute a profile of the integrand of (2.1.1). This profile can then be numerically integrated to yield a value for the (zenith) wet path delay.

There are at this point two methods which can be used to complete the determination of $\mathcal{W}$:

Method 1: If a WVR is available for the site being studied, a measurement in the zenith direction may be taken simultaneously with the radiosonde launch. (To avoid possible errors associated with the drift of the WVR, it is also possible to take a WVR measurement in the general direction of the radiosonde drift. To our knowledge, this technique has never been used.) A dual-frequency WVR observable is then calculated using the right-hand side of (2.3.9). A regression fitting the liquid-independent, WVR-derived "observable" (2.3.3) as a linear function of the radiosonde-derived wet propagation delay, using a suitable number of WVR-radiosonde pairs, then yields $\mathcal{W}$. A constant parameter may be added to the model for the wet propagation delay (2.6.3) in order to test the theory and the data.

Method 2: If a WVR is not available (either not yet built, or perhaps the study is for a site which will only be occupied for a short time, such as for a GPS experiment), one must use the same radiosonde data to "simulate" WVR data.

The procedure is first to use the measurements of P, T, and RH to calculate the absorption coefficient of water vapor, given by (2.2.1), and of oxygen, given by (2.2.4). The amount of liquid water present, if any, must be inferred from the radiosonde data and used to calculate the liquid-water absorption coefficient given by (2.2.6). (Usually, one infers the presence of liquid water from the relative humidity being greater than some predetermined value: see Decker *et al.* [1978].) The individual absorption coefficients are added to yield a total absorption coefficient, which can be integrated using (2.1.7), (2.1.8), and (2.1.9) to yield a simulated brightness temperature. The process is repeated for the second frequency channel. The equation of radiative transfer can be used then to infer the opacity at each frequency, which can be used to calculate the right-hand side of (2.3.9), just as the opacities from the real WVR data were used. The procedure for determining the weighting function $\mathcal{W}$ is hereinafter the same as for Method 1.

A reasonable question to ask is: Why go to the trouble of using the absorption to calculate the brightness temperature, which is then inverted to yield the opacity, when the absorption can be directly integrated using (2.1.9) with $s \rightarrow \infty$ to calculate the opacity directly (and with much less calculation)? The answer is related to the fact that when we reduce the true WVR data, we in general do not have *measurements* of the effective temperature T_{eff}, but rather we have a *model* based on surface temperature. If there are errors in this model, then there will be errors in the estimation of opacity from brightness temperature. Since we want to take these possible errors into account when we determine the weighting function, we use the radiosonde data only to simulate

the WVR observable which is (relatively) independent of any atmospheric-model; this observable is the brightness temperature.

Which of the two methods is "better"? The determinations of $\mathcal{W}$ via the first method are usually thought to be better because the radiosonde data are used to determine only one of the quantities necessary to obtain $\mathcal{W}$. Any algorithm errors can be observed during the derivation of $\mathcal{W}$. Furthermore, unknown instrumental offsets are automatically absorbed into the value of $\mathcal{W}$. However, this "absorption" can be a disadvantage, if the WVR changes sites or if a new WVR is brought to a site. The "Method 2" determinations have the advantage that the value of $\mathcal{W}$ goes with the site, and not the WVR.

Assuming now that we have a numerical value in hand for the dual-frequency weighting function $\mathcal{W}$, from where can errors in estimating the wet path delay from WVR data arise? Assuming an ideal radiometer (no instrumental error, G and T_{off} in (2.4.1) known perfectly, *etc.*), the only possibilities are:

1. Error in T_{bg}
2. Error in T_{eff}
3. Error in τ_{O_2}
4. Error in $\mathcal{W}$

Assuming the frequencies of the WVR are known, the above are the only sources of error. We will examine these sources of error shortly. Before proceeding, however, it is necessary to establish the value of some quantities, so that we can finally interpret

these errors in terms of errors in the estimate of the wet path delay. We will assume that the two frequencies for the WVR are approximately 21 and 31 GHz. (These are within 1 GHz of currently available dual-frequency WVR's.) From (2.6.3), it can be seen that the contribution of the 31 GHz opacity is therefore about 0.43 times that of the 21 GHz channel. For the value of $\mathcal{W}$, we can use an approximate value of $\mathcal{W}^{-1} \simeq 170$, as reported by Resch [1984]. Although this number is site—and possibly season—dependent the variations in this number are less than 20% of its value. With these numbers in mind, and using (2.6.1) through (2.6.3), we can now calculate the effects of errors in any of these quantities in terms of delay.

2.6.i Error in T_{bg}

An error δT_{bg} in the background temperature, from (2.6.2), induces an error $\delta\tau$ in the estimate of the opacity

$$\delta\tau \simeq \frac{\delta T_{bg}}{T_{eff}} \tag{2.6.4}$$

The temperature of the microwave background resulting from the creation of the universe is known to within a few tenths of a Kelvin, so that the contribution of this uncertainty is less than about 1 mm (independent of elevation angle). However, because the widths of the WVR beams are not infinitely narrow, radiation leaks into the sidelobes which effectively produces another source of elevation-dependent "background" radiation. If the shape of the beams are known, then one can correct for radiation from the atmosphere in directions other than the line-of-sight entering the sidelobes [Lundqvist and Potash, 1985]. If the WVR is pointed too close to the horizon, then some fraction of the ground temperature of approximately 300 K will go into T_{bg}.

For example an obstruction of angular size $5° \times 5°$, physical temperature 300 K, and distance of 15° from pointing direction will increase the estimated sky brightness temperature by ~ 0.3 K, which (assuming both channels were affected by approximately the same amount) would increase the estimated wet path delay by ~ 1 mm. (For this reason the WVR's rarely point below 20° elevation.) It is also true that as the number of earth-orbiting artificial satellites increases, the probability that a satellite transmitting near 22 GHz increases, since this region of the spectrum is not within a protected frequency band. There will therefore be times when there are "glitches" in the WVR data for seemingly unknown reasons.

2.6.ii Error in T_{eff}

An error in T_{eff} is due to the mismodeling of this quantity for use in (2.6.2). Typically, when the radiosonde studies for the weighting function are performed, values for parameters in a model for the effective temperature are also determined. The simplest model for the effective temperature is a linear model based on surface temperature. The first to use such a model apparently was Wu [1979], who reported a "good" fit using radiosonde data from a single site. Resch *et al.* [1985] using data

from several U.S. sites also found a strong linear correlation with surface temperature, but no details were reported.

We have used radiosonde data to study the modeling of the effective temperature. This study involved using the radiosonde estimates of P, T, and RH to perform the integration on the right-hand sides of (2.1.8) and (2.1.9) numerically. The results for the study using radiosonde data from the radiosonde launch site at Portland, Maine from the year 1981 are shown in Table 2.6.1. Only the results for the frequency 20.7 GHz are shown, but the results for the frequency 31.4 GHz are nearly the same, with a slight decrease of a few Kelvins in the magnitude of the effective temperatures, as expected. The first column indicates which subset of the launches were used. The label "all" indicates that all of the launches were used which reported relative humidity data up to a height of at least 2 km. The second column of this table indicates the cloud model used. The cloud models used are those given by Decker *et al.* [1978]. An additional model is cloud model zero, which indicates that no liquid was assumed to be present. The third column gives the average surface temperature and the standard deviation of surface temperature. The fourth column gives the average effective temperature and its standard deviation. The fifth and sixth columns give the coefficients A and B determined for

$$T_{eff} = A \cdot T_o + B \tag{2.6.5}$$

where T_o is again the surface temperature. Since all the data were weighted equally, the relationship

$$\bar{T}_{eff} = A \cdot \bar{T}_o + B \tag{2.6.6}$$

Table 2.6.1

T_{eff} studies for Portland, Maine

Site: Portland, Maine
Year: 1981
Frequency: 20.7 GHz

Data set	Cloud model	$\bar{T}_o$ (K)	$\bar{T}_{eff}$ (K)	A (K/K)	B (K)	$A(\bar{T}_o)_1 + B$ (K)	RMS (K)	N
All	I	282.4(9.9)	262.4(7.2)	0.6565	77.0	262.4	3.1	422
Winter	I	270.7(7.0)	253.3(4.7)	0.4864	121.6	259.0	3.2	84
Spring	I	285.7(6.4)	263.9(5.2)	0.6414	80.7	261.8	3.1	127
Summer	I	290.8(4.4)	269.0(2.6)	0.3891	155.8	265.7	1.9	132
Fall	I	275.5(6.4)	258.6(4.6)	0.5083	118.6	262.1	3.3	79
00Z	I	287.0(8.3)	264.1(6.7)	0.7215	57.1	260.9	3.0	179
12Z	I	279.0(9.6)	261.1(7.3)	0.7022	65.1	263.4	2.0	243
All	0	282.4(9.9)	262.3(7.3)	0.6551	77.3	262.3	3.1	422
All	II	282.4(9.9)	262.4(7.2)	0.6578	76.6	262.4	3.2	422
All	III	282.4(9.9)	262.5(7.3)	0.6603	76.0	262.5	3.2	422

where $\bar{T}_{eff}$ is the average effective temperature and $\bar{T}_o$ is the average surface temperature, is obeyed. This relationship gives us a handy method of comparing the values found for A and B using the subsets of data. The seventh column contains values for $A \cdot (\bar{T}_o)_1 + B$, where the subscript "1" indicates that the value for the mean surface temperature from the first solution was used. Because different data subsets were used, and because the estimates A and B are correlated, differing values of A and B may give nearly the same effective temperature for the same surface temperature, and this column allows us to compare the differences. (Another method would have been to constrain the parameter A to its value of 0.657 from the first solution. The estimates of B would then be directly comparable.) The eighth column gives the root-mean-square (rms) residual effective temperature, and the final column gives the number of profiles used in that subset. (There were originally 722 profiles, of which 300 were rejected due to the criterion given above.)

These results show that there is a variation in the constants over the year equivalent to a 7 K difference in the estimated effective temperature. The RMS fit is better in the summer (1.90 K for the summer, compared to 3.11 K in the spring, for example), probably due to the fact that the surface temperature in summer more nearly represents the profile of temperature above the ground because of the lessened effect of inversions. A *constant* value for the effective temperature in the summer could also be used, and the RMS variation (2.6 K) would still be less than the RMS fit to the linear model for the other seasons.

The results obtained here are not necessarily inconsistent with the experimental evidence that the variation about the true value of the wet delay predicted by the

dual-frequency algorithm using a constant weighting function is smaller during the summer [Schaper *et al.*, 1970; Moran and Rosen, 1981]. Their results probably reflect the greater amount of water vapor in the summer, whereas our results depend mainly on the profile of temperature, and to a lesser extent also on the *profile* of water vapor. Indeed, smaller amounts of water vapor may contribute to large variations in T_{eff} since small variations in the amount of water vapor effect large changes in the overall profile if the total amount of water vapor is small.

The rows labelled (in the first column) as "00Z" and "12Z" indicate the launch times of the radiosondes: 00Z = 00 hr GMT. The comparison between these two rows shows a slight (2.6 K equivalent) difference between the resulting effective temperatures. Use of the different cloud models has a very small (0.1 K) effect. The effect on the RMS is to increase the RMS as the number representing the cloud model increases, which is expected (since the higher this number, the greater the cloud thickness).

In order to test the dependence of these results on location, we have repeated the study for the radiosonde launch site at Hilo, Hawaii. The results are shown in Table 2.6.2. The effective temperature at Hilo displays much less variation: 1.4 K RMS for all launches (cloud model I). In reference to the discussion for the Portland site, this smaller variation is probably due to Hilo being a year-round "summer site." However, it is interesting that seasonal differences can still be discerned. The effective temperature undergoes a 2 K peak-to-peak annual variation, with the high being in the summer, as expected. Furthermore, the RMS residual predicted effective temperature is much less for the summer than for the winter–0.55 K *vs.* 1.25 K. This smaller variation seems strange in that the mean surface temperature itself for those seasons

differs by only about 2 K; these facts imply that temperature inversions may have an effect on the prediction of T_{eff} even in tropic climates.

Another similarity with the results for the Portland site is the effect of the (assumed) liquid water. The differences in cloud models tends to change the effective temperature by only about 0.2 K, and the RMS of the postfit residuals increases as the amount of liquid water increases.

The differences between the 00Z and 12Z launches are about 3 K for Portland and about 1 K for Hilo, with the sense of both being $T_{eff}(00Z) - T_{eff}(12Z)$. At Portland, local times for the launches are approximately 8 AM and 8 PM, while at Hilo they are 11 AM and 11 PM. Thus, there is a morning launch and an evening launch at both sites (although the morning launch at Portland corresponds to the evening launch at Hilo). The fact that the Hilo AM-PM difference is less than the Portland AM-PM difference is no doubt due to the difference in climates.

Table 2.6.3 shows the expected errors in the estimates of effective temperatures for several combinations of derived coefficients and their application. The entries in this table indicate the average error (*i.e.*, bias) which would occur from using the constants derived from one set of data to predict the effective temperature for another set of data. We can see that except for a few cases, the errors are less than 1%. The exceptions arise from using constants derived (*i*) at one site to predict the effective temperature at another, and (*ii*) for one season to predict the effective temperature for another. In case (*i*), the large errors occur only for using the Hilo constants to predict the Portland constants, and not vice versa. This effect is due to there being such a small range of effective temperatures at Hilo, that for most of the time the

Table 2.6.2

T_{eff} studies for Hilo, Hawaii

Site: Hilo, Hawaii
Year: 1981
Frequency: 20.7 GHz

Data set	Cloud model	$\bar{T}_o$ (K)	$\bar{T}_{eff}$ (K)	A (K/K)	B (K)	$A(\bar{T}_o)_1 + B$ (K)	RMS (K)	N
All	I	296.9(3.3)	272.0(1.3)	0.2317	203.1	272.0	1.1	574
Winter	I	295.9(3.8)	271.2(1.4)	0.1633	222.9	271.4	1.3	146
Spring	I	296.9(3.2)	271.7(1.2)	0.2295	203.5	271.6	1.0	144
Summer	I	298.0(2.8)	273.1(0.7)	0.1523	227.7	272.9	0.5	152
Fall	I	296.8(2.9)	272.1(1.2)	0.2504	197.8	272.1	1.0	132
00Z	I	299.8(1.7)	272.5(1.3)	0.5083	120.1	271.0	1.0	285
12Z	I	294.1(1.5)	271.6(1.2)	0.4142	148.8	271.8	1.0	289
All	0	296.9(3.3)	272.0(1.4)	0.2376	201.4	271.9	1.1	574
All	II	296.9(3.3)	272.1(1.4)	0.2260	205.0	272.1	1.2	574
All	III	296.9(3.3)	272.2(1.4)	0.2151	208.3	272.2	1.2	574

effective temperatures at Portland fall well outside this range; the reverse, however, is not true because of the relatively wide range of effective temperatures at Portland. (These ranges are inferred from the rms variation in the effective temperatures from Table 2.6.1.) For the same reason, the seasonal variations of case (*ii*) are found at Portland only, because at Portland the constants for each season are determined from an approximately disjoint set of temperatures. At Hilo, however, where the seasonal variation of temperature is much less, this effect does not occur.

How do the errors in effective temperature affect the estimates of delay? From (2.6.2), we find that an error δT_{eff} induces an error in the opacity given by

$$\frac{\delta\tau}{\tau} \simeq \frac{\delta T_{eff}}{T_{eff}} \tag{2.6.7}$$

Because the results for the two frequencies were so similar, we will assume that the fractional error in the effective temperatures is approximately the same for both frequencies. Then from (2.6.3) and the equation of radiative transfer we have

$$\frac{\delta L_w}{L_w} \simeq \frac{\delta T_{eff}}{T_{eff}} \tag{2.6.8}$$

The results of this section imply that the rms errors in the delay due to errors in the effective temperature can be kept to less than 1%, if site-dependent coefficients are used for the model of effective temperature. It is important to remember that the error due to an error in the effective temperature will be changing on time scales of 1 day, so that this error will appear as a bias for a 24 hour observing session.

Table 2.6.3

Errors from using erroneous T_{eff}

Use A,B derived from	To predict T_{eff} for	Mean error (K)
All, Hilo	All, Portland	-6.2
All, Portland	All, Hilo	0.1
Winter, Hilo	Summer, Hilo	1.5
Summer, Hilo	Winter, Hilo	-1.6
Winter, Portland	Summer, Portland	6.0
Summer, Portland	Winter, Portland	-7.8
All, Hilo	Summer, Hilo	0.9
All, Hilo	Winter, Hilo	-0.6
All, Portland	Summer, Portland	1.1
All, Portland	Winter, Portland	-1.4
All, Hilo	00Z, Hilo	-0.2
All, Hilo	12Z, Hilo	0.3
All, Portland	00Z, Portland	-1.3
All, Portland	12Z, Portland	0.9

2.6.iii Error in τ_{O_2}

In Section 2.2.*ii*, we presented the form for the absorption for oxygen, and derived an expression for the zenith opacity for oxygen. The contribution from oxygen to the wet path delay for the dual-frequency algorithm can be calculated by from (2.6.3) and (2.2.1). Doing so, we find that this contribution (the so-called "oxygen correction") is

$$L_w(O_2) = -0.0038\mathcal{W}^{-1}\left(\frac{P_o}{1013.25}\right)^2\left(\frac{293}{T_o}\right)^{2.85} m \qquad (2.6.9)$$

where m is the air mass, and frequencies of 20.7 GHz and 31.4 GHz were used. For $\mathcal{W}^{-1} \simeq 170$ cm neper^{-1}, we find that for standard temperature and pressure, the zenith contribution of oxygen is approximately 0.6–0.7 cm. As we discussed in Section 2.2.*ii*, an extreme upper limit on the error of the expression for the oxygen opacity is about 20%, implying that any error in the oxygen correction at zenith is less than 1.5 mm (corresponding to about 9 mm at 10° elevation).

Resch *et al.* [1985] report the results of an effort to obtain a weighting function for frequencies identical to the above. They assumed a form for the zenith path delay of

$$L_w = A_o + A_1\left[\tau_1 - \frac{\nu_1^2}{\nu_2^2}\tau_2 - A_2\left(\frac{P_o}{1013.25}\right)^2\left(\frac{293}{T_o}\right)^{2.85}\right] \qquad (2.6.10)$$

Note that in this algorithm, A_1 takes the place of $\mathcal{W}^{-1}$. If the algorithm we have presented is correct, than Resch *et al.* should have found $A_o \simeq 0$ and $A_2 \simeq 0.0038$. In fact, A_2 was constrained in such a way as to produce $A_o \simeq 0$. However, the value they obtained for A_2 was 0.0016, a value half as large as the theoretical value obtained here. This difference could indicate a bias in the humidity values for the radiosondes used,

which if unconstrained would cause a nonzero value for A_o. However, this comparison clearly indicates that more work is needed at this level of algorithm development.

2.6.iv Error in $\mathcal{W}$

Earlier in this section, we discussed the method by which the value for the dual-frequency weighting function $\mathcal{W}$ has historically been obtained. Obviously one of the greatest potential sources of error is error in the radiosonde data used to obtain this value. The contribution of this error is probably between 5% and 10%, and is discussed briefly in Elgered [1983].

When we formulated the dual-frequency algorithm we assumed that $\mathcal{W}$ was constant with height. In practice, this assumption causes no errors since the methods given above for determining $\mathcal{W}$ are equivalent to determining a *mean* dual-frequency weighting function $\mathcal{W}_o$ which corresponds to the mean single-frequency weighting function W_o defined in (2.1.5). However, no one has made a comprehensive study of the dual-frequency weighting function presented here to determine whether there are a pair of frequencies for which $\mathcal{W}$ is nearly constant. The dual-frequency algorithm appearing in this thesis has never before been presented explicitly; it is however based on the dual-frequency algorithm formulated by Wu [1979] for atmospheric brightness temperatures, or alternatively for "linearized" brightness temperatures. In fact, the dual-frequency weighting functions for opacity and for linearized brightness temperatures are equivalent in the sense that they both have properties which might be exploited in determining whether a constant weighting function can be found. They are both independent of the water-vapor density, except for a weak dependence through the line

half-width given by (2.2.2). Thus, calculation of profiles of the weighting function is unaffected by the poor precision of the radiosonde's humidity measurements [Elgered, 1983]. Both the opacity and linearized brightness temperature weighting functions are also unaffected by the cumulative strength of the water vapor lines above 22 GHz, so any deficiencies in the theory regarding this strength is unimportant (see Section *2.2.i*).

The benefits of finding a dual-frequency weighting function which is constant with height (to within a few per cent) is obvious. Measurements of pressure and temperature at the surface can then be used to determine the value of the weighting function at the time of the WVR observations. Thus, overall variations in the value of the weighting function can be taken into account.

Although no exhaustive search for "optimum" frequency-pairs has been conducted, Elgered *et al.* [1985] have performed a series of "Method 2" studies, wherein the value for $\mathcal{W}$ is determined for a nearly continuous range of first frequencies, for a fixed second frequency. For each pair of frequencies, the value of $\mathcal{W}$ is determined using Method 2, and the rms fit of the actual delay to the predicted delay is calculated. Figure 2.6.1 shows their results for the site at Landvetter, near Onsala, Sweden. The results were compared for two different fixed frequencies, 18 and 31.5 GHz. The calculations were repeated for the two fixed frequencies, for two different assumed WVR instrumental errors, 0.2 K and 1 K. Figure 2.6.1 shows the results broken down by season. The results of Elgered *et al.* imply that "Method 2" determinations of $\mathcal{W}$ can

be improved very little by choosing different frequency pairs. Instead, the instrumental error appears to dominate.

What seasonal and site dependent variations of the dual-frequency weighting function can be expected? Figure 2.6.2 shows preliminary results from Johansson [1985]. Shown are values of $\mathcal{W}^{-1}$ determined using Method 2, for Portland, ME and Landvetter, Sweden. Since Johansson used linearized brightness temperatures, the units are given as cm K^{-1}, rather than cm $neper^{-1}$. An approximate conversion of linearized brightness temperatures to opacity is found in Section 2.7. Johansson's results show that $\mathcal{W}$ can undergo a 10–20% annual variation. The site-to-site variations in this study are about 10%, but variations can be expected to be larger for sites with much different climates.

In this section we have discussed possible errors in the WVR algorithm. In Chapter 5 we discuss experimental tests of the accuracy of WVR's. We will show how VLBI data can be used to study the accuracy of WVR's, and present results. These results will be interpreted in terms of the errors discussed in this section. In the next several sections, we will discuss briefly alternative WVR algorithms, including the algorithm for linearized brightness temperatures, already mentioned above, and algorithms for multifrequency WVR's.

2.7 Alternative dual-frequency algorithms

The dual-frequency algorithm presented in Section 2.3 was motivated by the similarities between the expressions for the wet path delay and the opacity of water vapor: they both involve the integral of the density of water vapor. A dual-frequency

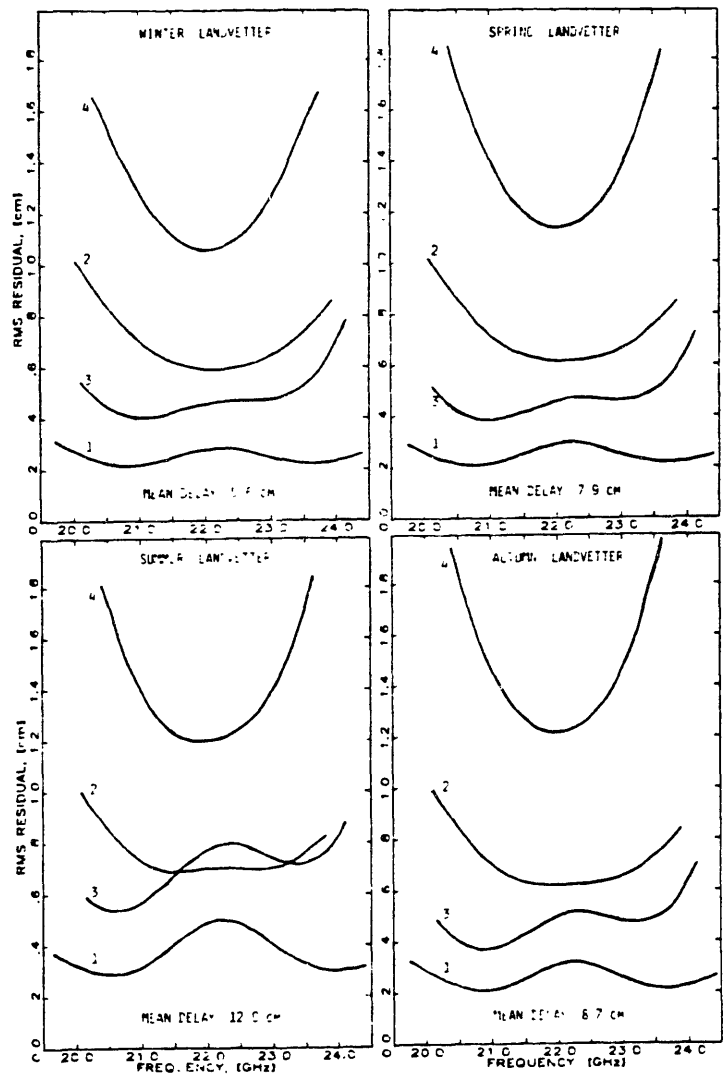


Figure 2.6.1. RMS residual wet path delay for "Method 2" determination of weighting function (see text), from Elgered *et al.* [1985]. The RMS residual is plotted as a function of frequency of the channel closest to the water-vapor line for different instrumental errors and for two fixed frequencies—31.5 GHz (curves 1 and 2) and 18.0 GHz (3 and 4). The assumed instrumental errors are 0.2 K (1 and 3) and 1 K (2 and 4).

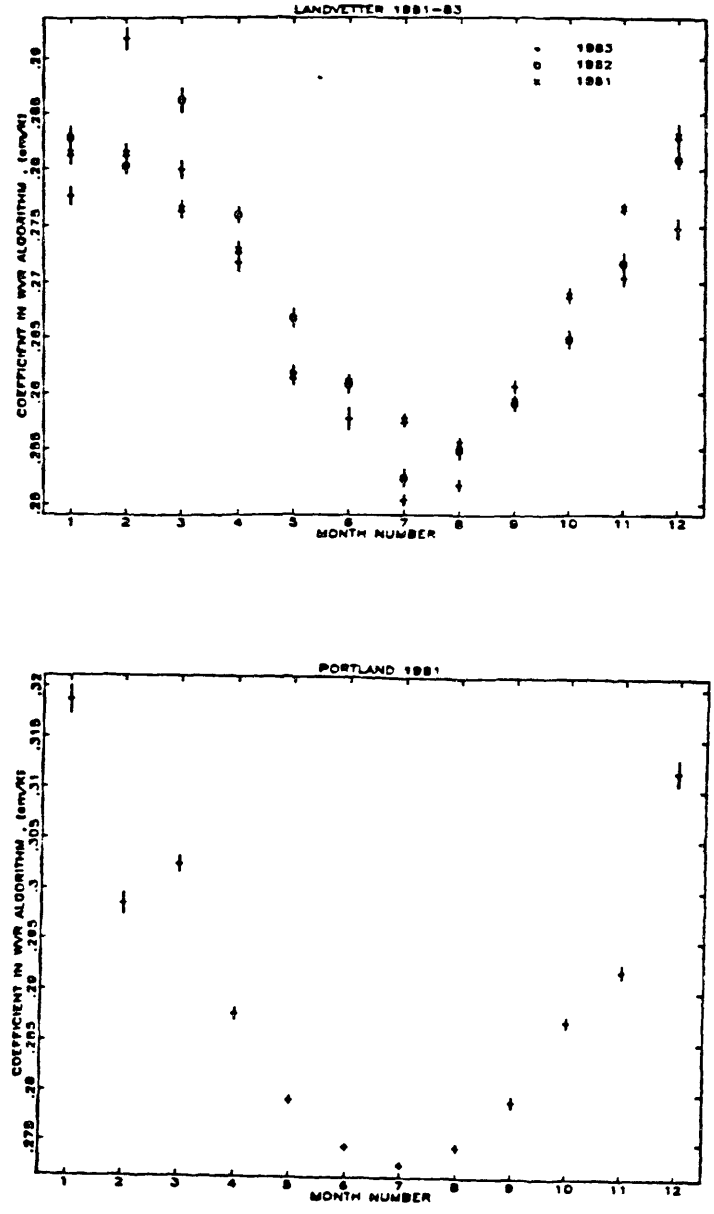


Figure 2.6.2. $\mathcal{W}^{-1}$ *vs.* month for "Method 2" determination of weighting function (see text), from Johansson [1985]. The units are those given for linearized brightness temperatures. For purposes of conversion, note 1 cm $K^{-1} \simeq 531$ cm $neper^{-1}$ (see Section 2.7).

formula was employed in order to form a liquid-independent "observable." However, this algorithm is not unique, and in this section we present two alternative dual-frequency algorithms. The first algorithm is very similar to the algorithm presented in Section 2.3, except it makes use of "linearized brightness temperatures" instead of opacities. The second algorithm is a new algorithm, and very different from the other two; this algorithm is known as the "profile algorithm." In Section 2.8, we will briefly discuss algorithms which utilize more than two frequency channels.

2.7.i Linearized brightness temperatures

The first attempts to use radiometry to obtain the wet path delay or the integrated water vapor content used the sky brightness temperatures as the primary observable [Schaper *et al.*, 1970; Moran and Rosen, 1981]. However, by examining the equation of radiative transfer (2.1.7), one can see that the brightness temperature suffers from effects of saturation: For high opacities, the amount of liquid in the atmosphere can double, say, while the brightness temperature only increases slightly. Therefore, Wu [1979], in the first presentation of the theory behind a dual-frequency WVR, suggested the use of "linearized" brightness temperatures. The linearized atmospheric brightness temperature T_a' is given by

$$T_a' = T_{bg} - (T_{eff}' - T_{bg}) \log\left(1 - \frac{T_a - T_{bg}}{T_{eff} - T_{bg}}\right) \tag{2.7.1}$$

where T_a is the atmospheric brightness temperature, T_{bg} is the cosmic background temperature, T_{eff} is the "effective" atmospheric temperature, defined in (2.1.8), and T_{eff}' is the linearized effective temperature, given by

$$T_{eff}' = \frac{\int_{atm} ds\, T(s)\kappa(s)\rho(s)}{\int_{atm} ds\, \kappa(s)\rho(s)} = \frac{\int_{atm} ds\, T(s)\kappa(s)\rho(s)}{\tau} \tag{2.7.2}$$

The linearized brightness temperature can also be written in terms of the opacity τ:

$$T'_a = T_{bg}(1-\tau) + T'_{eff}\tau \tag{2.7.3}$$

Comparison of (2.1.8) with (2.7.2) and (2.1.7) with (2.7.3) shows that $T'_a \rightarrow T_a$ as $\tau \rightarrow 0$. Because of these similarities, there is probably no discernible difference (at least at our current levels of accuracy: see Chapter 4) between the opacity and linearized brightness temperature dual-frequency algorithms. Using (2.7.3), we can determine a conversion to the weighting function for the linearized brightness temperature from the weighting function for the opacity:

$$\mathcal{W}^{-1}_{T'_a} \simeq \frac{\partial L_w}{\partial T'_a} \simeq \left(\frac{\partial L_w}{\partial \tau}\right)\left(\frac{\partial \tau}{\partial T'_a}\right) \simeq \frac{1}{T'_{eff}}\mathcal{W}^{-1}_{\tau} \tag{2.7.4}$$

A difference in the implementation of the dual-frequency algorithm using linearized brightness temperatures has been the use of "ν_1" to represent the higher frequency (~30 GHz), while "ν_2" represents the lower (~21 GHz). This simply causes an extra factor in the conversion of the weighting functions:

$$\mathcal{W}^{-1}_{T'_a} \simeq \frac{1}{T'_{eff}}\left(\frac{21}{30}\right)^2 \mathcal{W}^{-1}_{\tau} \tag{2.7.4}$$

Using $T'_{eff} \simeq T_{eff} \simeq 260$ K and $\mathcal{W}^{-1}_{\tau} \simeq 170$ cm neper^{-1}, we find that $\mathcal{W}^{-1}_{T'_a} \simeq$ 0.3 cm K^{-1}. If we examine Figure 2.6.2, we find that this estimate is very close to the actual weighting functions determined for the linearized-brightness temperature algorithms.

7.2.ii The profile algorithm

The profile algorithm, developed by Robinson [1985], was introduced very recently, and has not been tested with VLBI data in the way that have the opacity and, by extension, the linearized brightness temperature algorithms. In the profile algorithm, the relative humidity profile has a specific form: The relative humidity varies linearly with altitude from its value at the surface to its value at an altitude of 3 km. The relative humidity at the surface is constrained by measurements there; the relative humidity at 3 km is estimated in a way described below. Above 3 km, the relative humidity is again assumed to vary linearly (although with a different lapse rate), until it reaches the value of zero at an altitude of 10 km. Above 10 km, the relative humidity is assumed to be zero. The density of liquid water is assumed to be proportional to the absolute humidity at every point. The constant of proportionality is another parameter to be estimated. The temperature profile is computed by taking a linearly decreasing temperature profile and adding to this profile an exponentially decreasing term with a scale height of 2 km, such that the temperature at the surface matches the measured value.

The estimation of the two parameters—the relative humidity at an altitude of 3 km and the liquid constant—is performed iteratively. Using the current values of the two parameters to determine the profiles of relative humidity and liquid water, the profiles are integrated as per (2.1.8) to determine the resulting brightness temperatures at the two frequencies of the dual-channel WVR. If the calculated brightness

temperatures do not match the WVR's observed brightness temperatures, the parameters are varied, and the integration is performed again. This process is repeated until the calculated brightness temperatures match the observed brightness temperatures.

Robinson has reported RMS errors of less than 3 mm in testing his algorithm using radiosonde data to simulate WVR observations. These errors are almost a factor of two less than the RMS errors Resch [1984] reported in testing the opacity algorithm, again against radiosonde data only. In the near future, testing should begin with Robinson's algorithm in real situations, and it may be determined if there are site and/or season dependent behaviors of his algorithm (see Chapter 4). However, this algorithm is too new and as yet too untested to support further comment.

In the next section, we will briefly discuss algorithms for WVR's employing more than two frequency channels.

2.8 Multichannel water vapor radiometers

In Section 2.3, we presented an algorithm for estimation of the wet path delay using opacities at two frequencies. The reason we chose to employ observables at only two frequencies is that until very recently, all of the water-vapor radiometers available for geodetic VLBI purposes received radiation at only two frequencies. These WVR's included the R-series, the upgraded R-series, and The Onsala WVR (see Section 2.4). Recently, however, a new design of WVR is becoming available: the J-series, which has three frequency channels. (There are other differences also: see Janssen [1985].)

In this section, we will discuss two ways in which the dual-frequency algorithm may be extended for a three frequency WVR, although only one of these methods is

appropriate for the J-series. We will assume that two of the frequencies are situated in about the same place relative to 22.235 GHz as for the dual-frequency WVR, that is, the first frequency is on this line, while the second is not. (Take 20.7 and 31.4, for example.) The implementation of a three-frequency algorithm then depends on the placement of the third frequency. We will first discuss an algorithm for the third frequency being near about 45 GHz, so that the opacity of oxygen may be determined. However, the third frequency of the J-series WVR's is much closer to 22.235 GHz, and a second algorithm must be used. This algorithm will also be discussed.

2.8.i Elimination of τ_{O_2}

The most obvious use for observations at a third frequency is elimination from the delay algorithm of the opacity due to oxygen. This elimination can only be reliably obtained, however, if the third frequency is close to the oxygen lines near 60 GHz. However, since the form for the oxygen absorption given in Section 2.2.*ii* is only accurate for frequencies less than about 45 GHz (since then the group of oxygen lines can be treated as a single broad line), we will assume that the third frequency is less than, but near to, 45 GHz. In principle this assumption is not necessary, but above 45 GHz a different form for the oxygen absorption must be used (see Rozenkranz [1975], for example).

The three-frequency version of (2.3.5) is

$$\begin{aligned} \tau_1 - \frac{\nu_1^2}{\nu_2^2}\left(\frac{K_3-K_1}{K_3-K_2}\right)\tau_2 + \frac{\nu_1^2}{\nu_3^2}\left(\frac{K_2-K_1}{K_3-K_2}\right)\tau_3 = \\ \tau_{v1} - \frac{\nu_1^2}{\nu_2^2}\left(\frac{K_3-K_1}{K_3-K_2}\right)\tau_{v2} + \frac{\nu_1^2}{\nu_3^2}\left(\frac{K_2-K_1}{K_3-K_2}\right)\tau_{v3} \end{aligned} \tag{2.8.1}$$

In (2.8.1), τ_1, τ_2, and τ_3 are the "observed" opacities; τ_{v1}, τ_{v2}, and τ_{v3} are the corresponding opacities due to water vapor; K_1, K_2, and K_3 are the oxygen shape functions; and ν_1, ν_2, and ν_3 are the three frequencies.

We have not presented an explicit derivation of (2.8.1) because the derivation is long and contains no interesting information. We will, however, outline the procedure. (This procedure was used implicitly to derive the dual-frequency algorithm.) Using (2.3.1), we first write three observation equations corresponding to the three frequency channels:

$$\begin{aligned} \tau_1 &= \tau_{\ell 1} + \tau_{v1} + \tau_{d1} \\ \tau_2 &= \tau_{\ell 2} + \tau_{v2} + \tau_{d2} \\ \tau_3 &= \tau_{\ell 3} + \tau_{v3} + \tau_{d3} \end{aligned} \tag{2.8.2}$$

We have used the subscript d for the oxygen opacities because the subscript O_2 is cumbersome. From Sections 2.2.*ii* and 2.2.*iii*, we know the frequency dependence of the opacities due to liquid water and oxygen. Using the results from these sections, we obtain

$$\begin{aligned} \tau_1 &= \tau_{\ell 1} + \tau_{v1} + \tau_{d1} \\ \tau_2 &= \frac{\nu_2^2}{\nu_1^2}\tau_{\ell 1} + \tau_{v2} + \frac{K_2\nu_2^2}{K_1\nu_1^2}\tau_{d1} \\ \tau_3 &= \frac{\nu_3^2}{\nu_1^2}\tau_{\ell 1} + \tau_{v3} + \frac{K_3\nu_3^2}{K_1\nu_1^2}\tau_{d1} \end{aligned} \tag{2.8.3}$$

We now treat the system of equations (2.8.3) as though there were three unknowns—$\tau_{\ell 1}$, τ_{v1}, and τ_{d1}—and solve for τ_{v1}. This will lead to a solution of τ_{v1} in terms of τ_1, τ_2, and τ_3, as well as τ_{v2} and τ_{v3}. Rearranging leads to (2.8.1).

At present, there are no WVR's available for geodetic use which utilize three frequencies in such a way as to eliminate the oxygen opacity. This situation is unfortunate,

considering the disparity between the theoretical value for the "oxygen correction" and the value determined from radiosonde data, discussed in Section 2.6. As mentioned above, the J-series has three frequency channels; these are arrayed, however, with two near the 22 GHz line of water vapor. In the next section, we outline the use of data from such a WVR.

2.8.ii Three close frequencies

If all three frequencies are below 30 GHz, then it is not clear that reliable information concerning the oxygen opacity can be obtained. The frequencies for the J-series WVR's are 20.7, 22.2, and 31.4 GHz. How, then, can the third frequency be incorporated? If we treat the third frequency as having additional information concerning the strength of the 22 GHz absorption line, then we have an overdetermined problem: we have three observables (the opacities at each of the three frequencies) and two unknowns (the liquid opacity and water-vapor opacity at one of the frequencies). The oxygen opacity is assumed known, as in the dual-frequency problem.

Since we have an overdetermined problem, we can use least-squares to obtain estimates of the unknown parameters. As the problem is stated above, the estimate of the water-vapor opacity at one frequency will be expressed in terms of the three "observed" opacities, the water-vapor opacities at the other two frequencies, and the oxygen opacity. Again, the derivation will be omitted, and we write the three-frequency version of (2.3.5):

$$\tau_1 - \nu_1^2\gamma(\nu_2^2\tau_2 + \nu_3^2\tau_3) = \tau_{v1} - \nu_1^2\gamma(\nu_2^2\tau_{v2} + \nu_3^2\tau_{v3}) - \gamma\left(\frac{K_2}{K_1}\nu_2^4 + \frac{K_3}{K_1}\nu_3^4\right)\tau_{d1} \quad (2.8.4)$$

where

$$\gamma = \left(\nu_2^4 + \nu_3^4 - \nu_1^4\right)^{-1} \tag{2.8.5}$$

Note that regardless of the number or placement of the frequencies available, there is in principle only a single number (albeit time– and season–dependent) which must be estimated from radiosonde data. In the case of a dual-frequency instrument, this "number" is given by (2.3.8). Using (2.8.1) and (2.8.4), we can find an expression for the appropriate three-frequency weighting function.

This ends our discussion of the theory of water-vapor radiometry. In the next chapter, we will return to our discussion of models of the propagation delay, and present a new mapping function for the dry propagation delay. Then, in Chapter 4, we will discuss experimental determination of the accuracy of the dual-frequency WVR algorithm on time scales from several minutes to one year.

Chapter 3

Development of a New Mapping Function for the "Dry" Atmosphere

Introduction

In Chapter 1, we discussed the effects on estimates of site position of errors in the formulas used to model the atmospheric propagation delay. We did not, however, present evidence for any such errors in the formulas commonly used for VLBI data analysis. These formulas include the Saastamoinen zenith delay, the Chao mapping functions, and the Marini "mapping function." (See Section 1.4.*ii*.) In this chapter, we will present evidence of systematic errors in these formulas. We will hypothesize that the primary errors are in the mapping function for the "dry" atmosphere. (See Section 1.3.) We will then describe the development of a new "dry" mapping function, and present the results of a limited test of the new mapping function. In Chapter 4, we will present the results of a more rigorous set of tests of the new mapping function. These latter tests use data from the "Low Elevation" VLBI experiments discussed in that chapter.

3.1 Evidence for mapping function errors

In this section, we will present the evidence which led to the hypothesis of errors in the mapping functions used to analyze VLBI data, although, as we will see, the

evidence is not of itself conclusive. In later sections, we will describe the development of a new mapping function and examine the effects of using this new mapping function in VLBI data analysis.

One useful method for detecting errors in the formulas used to model the propagation delay is to examine estimates of site position for behavior which depends on the distribution in elevation of the observations from which the estimates are derived. From the analyses performed in Section 1.5, we saw that any type of error in these formulas leads to systematic behavior in the estimates of site position, and therefore the presence of this behavior may indicate errors in these formulas. Herring [1983] devised an "elevation–angle–cutoff test" to examine estimates of baseline length for dependence upon the distribution of elevation angles of the observations. In this test, estimates of the baseline length using subsets of the full data set are compared; these subsets are characterized by the minimum elevation angle of the data set. Figure 3.1.1 shows the results of the elevation–angle–cutoff test, as a function of baseline length. Plotted as a function of nominal baseline length are the differences $\hat{b}_0 - \hat{b}_{10}$, where the $\hat{b}_0$ are the estimates of baseline length using all data (no cutoff with elevation angle), and $\hat{b}_{10}$ those for an elevation cutoff of 10°. The error bars represent the standard deviation of the difference between the estimates from the two solutions (see Appendix A). The obvious general trend in this figure is for the differences from longer baselines to be larger than those from shorter baselines. This trend may be explained by nearly constant differences in the estimates of the local vertical component of site position;

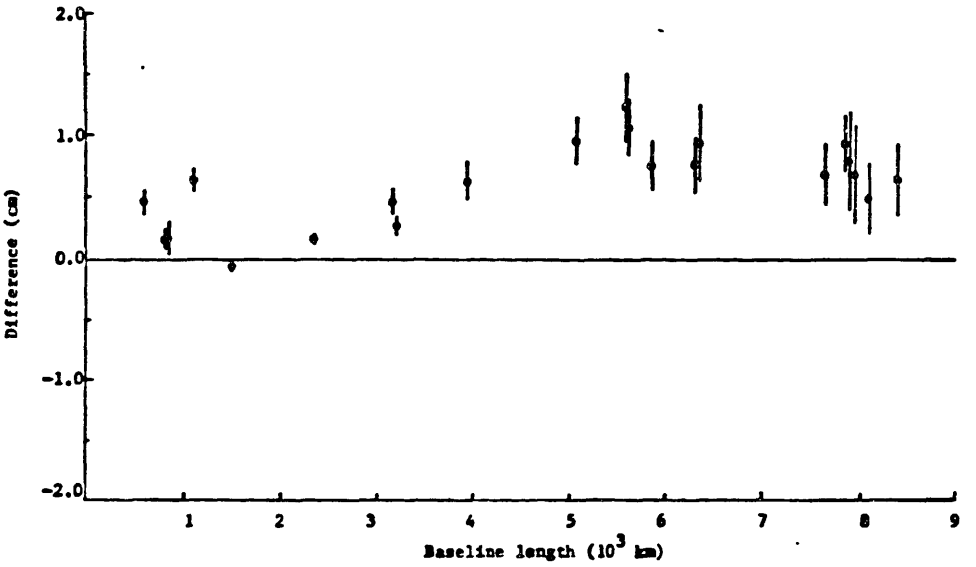


Figure 3.1.1. "Elevation-angle cutoff test" (see text) from Herring [1983]. Shown are the differences of the estimates of baseline lengths using data with a minimum elevation angle of 10° from estimates using data with a minimum elevation angle of 0°, plotted as a function of length of baseline.

simply from geometry we can see that these differences affect the baseline length more for longer baselines, in a way which can be described by

$$\Delta b \simeq \frac{b}{2r_e}\Delta z \tag{3.1.1}$$

where b is the baseline length, r_e is the radius of the earth, Δz is the difference between the solutions in the estimate of the local vertical component of the position of one of the sites forming the interferometer, and Δb is the contribution to the difference in the estimate of baseline length due to the difference in the estimate of the local vertical component of position for that site. (To determine the total effect for the interferometer, add the contributions from each site together.) Thus, if the difference in the estimates of the local vertical for every site were constant, then the differences in the estimates of baseline length would depend linearly on baseline length. Site–dependent variations could account for the variability about a straight line seen in Figure 3.1.1.

What could be causing these errors in the estimates of the vertical coordinate of site position? In arriving at the estimates of baseline length, Herring used approximately 23,000 group–delay measurements (fewer for the 10°–elevation solution) taken from all seasons. An additive adjustment to the *a priori* value of the zenith delay (obtained by using the zenith delay formula (1.4.1) and measurements of pressure, temperature, and relative humidity) was also estimated for each site and for each observing session. Hypothesizing that the effect was due to an error in the mapping function used, Herring estimated a site– and experiment–dependent adjustment to the mapping function, which was equivalent to estimating an adjustment to A in (1.4.6).

The effect on the elevation–angle–cutoff test of estimating this mapping–function parameter can be seen in Figure 3.1.2. This figure does not show the systematic trend of Figure 3.1.1. The estimated adjustments to A were, on the average, only a few per cent of A. We take Herring's results to be very strong evidence for the existence of mapping function errors.

We also found evidence for mapping–function errors in the comparison of estimates made using the adjustment to the zenith delay mentioned above, and corresponding estimates using no such parameter. Figure 3.1.3 shows such a comparison. This figure shows the estimates of the baseline length for the Ft. Davis–Onsala interferometer, for the cases of a zenith delay parameter estimated (solid squares), and for no such parameter estimated (solid circles). These experiments were analyzed as part of an investigation of the accuracy of WVR's (see Chapter 2), and use WVR data to estimate the "wet" delays at both sites. A bias is clearly visible between the estimates of baseline length made with and without simultaneous estimation of an adjustment to the zenith delay. One possible explanation for this result is that the wet propagation delay estimated from the WVR is biased. In order to check this possibility, we performed the same comparison using the Saastamoinen model for the zenith wet propagation delay in (1.4.1), mapped with the Marini mapping function (1.4.5). The results of this comparison are also shown in Figure 3.1.3, with filled circles and squares corresponding to open circles and squares, respectively. We can see that the bias still exists, and that the estimates made using the model for the wet delay are (usually) quite close to the estimates made using the WVR data. We might conclude that the surface model is biased in the same direction and magnitude as the WVR data. This

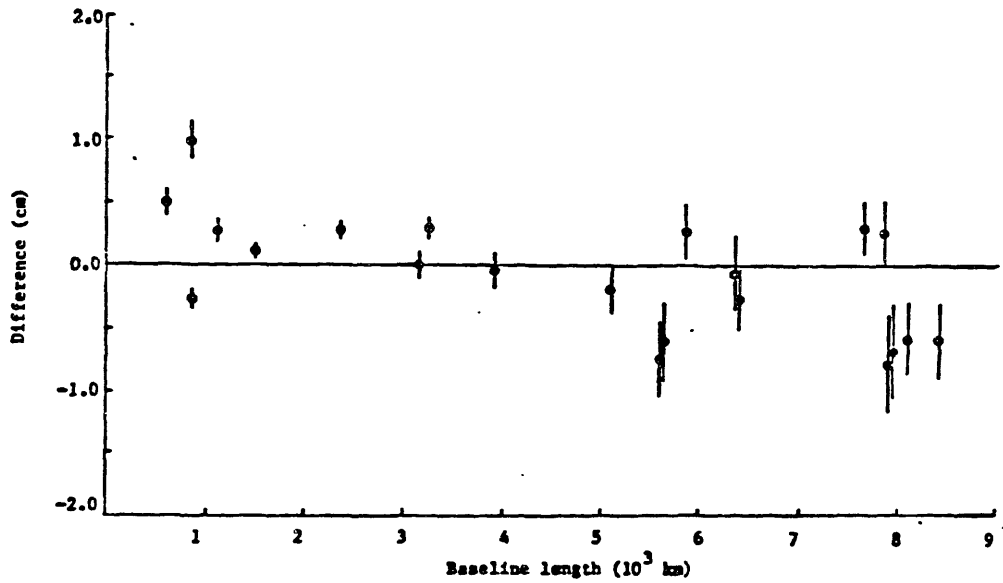


Figure 3.1.2. Elevation–angle cutoff test of Figure 3.1.1, repeated while simultaneously estimating a mapping function correction (see text).

explanation is unappealing, however, because studies have shown that the surface humidity and delay are uncorrelated [Reber and Swope, 1972] whereas this explanation implies that these two quantities are correlated.

Another explanation for the bias could be found in the effect of estimation of a zenith delay parameter if a mapping function error existed. This effect was studied in Section 1.5.*i*, and summarized in Figure 1.5.2. We found that the estimate of the vertical coordinate of site position could change by approximately 4 cm for an elevation–angle lower limit of 10° and a mapping–function parameter error of 10% (see Section 1.5.*i*). As mentioned above, Herring [1983] found a few per cent error in this parameter, so that it is not unreasonable to conclude that the biases in Figure 3.1.3 may be due to an error in the mapping function.

In order to determine conclusively whether the effects described in this section were due to errors in the mapping function, we decided to attempt to develop a more accurate mapping function. Such an attempt could be extremely time consuming, however, if we chose to solve the ray-trace equation (1.2.13) analytically, albeit approximately, as was done by Marini. Instead, we chose an *ad hoc* method whereby we fit some parametrized version of the mapping function to estimates of the mapping function produced by "ray-tracing." This method is described in the following two sections. Then, in Section 3.4, we present the new mapping function, and the initial test of its accuracy.

3.2 Ray–tracing

In this section we will describe the ray–trace calculations which ultimately led to the development of the new mapping function. The description will consist of two

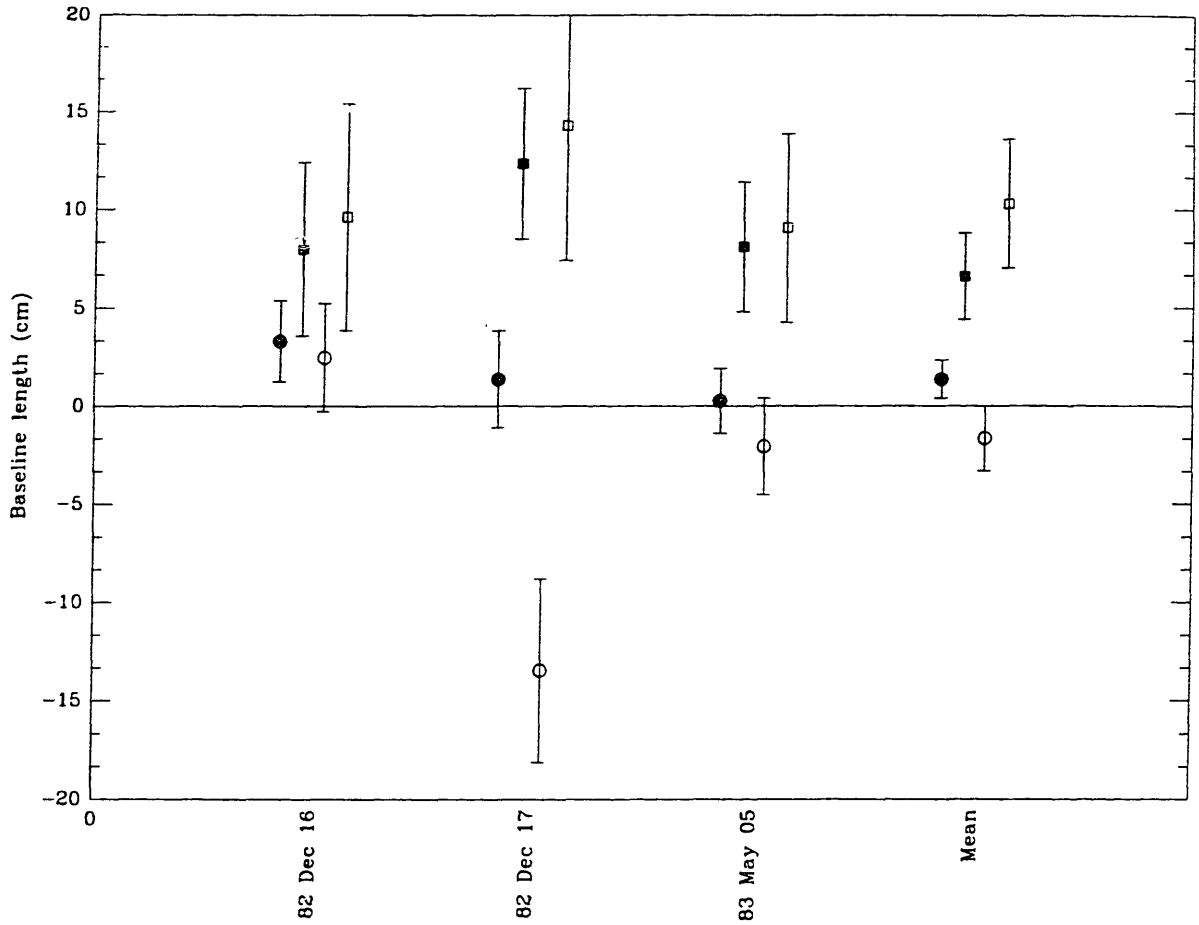


Figure 3.1.3. Comparison of estimates of the Ft. Davis (Texas)–Onsala (Sweden) baseline length, made with (▪,□) and without (•,○) the simultaneous estimation of a zenith delay parameter. The *a priori* estimates of the wet propagation delay were derived from the Saastamoinen model (□,○) or WVR data (▪,•). A convenient but arbitrary value of 794,073,200 cm has been subtracted from the estimates before plotting. The error bars shown are the statistical standard deviations resulting from performing a weighted–least–squares analysis, with the weight taken to be the sum of the signal–to–noise variance given by Clark *et al.* [1985] and a baseline–dependent variance, the value of which was determined by requiring that the χ^2 per degree of freedom be unity.

parts. The first part briefly describes the calculational aspects of ray–tracing, while the second part will describe the physical models actually used in the calculations. Because these models are based on some simple assumptions (gravity is taken to be a constant, for example), in later chapters we will discuss the effects of these assumptions on the mapping function developed.

In principle, the propagation delay can be calculated by performing the integral in (1.2.15). This equation, however, is written as though we knew both the zenith angle at the site (θ_o) as as well as the zenith angle "outside" the atmosphere (θ_f). In practice, we generally know only one of these angles beforehand, and the other angle depends on the profile of the refractive index. Therefore, we use the method of ray–tracing to calculate τ_a, the propagation delay.

We will first define some of the quantities which we will use in these calculations (see also Figure 3.2.1). We will describe ray–tracing for a spherical atmosphere, but the technique can be extended to handle any distribution of refractive index. The calculation will be performed as a series summation of constant radial step Δr. (In principle it is not necessary for Δr to be constant.) The atmosphere is therefore divided into spherical shells of thickness Δr, the boundary for the first layer beginning at $r = r_o$ and ending at $r = r_o + \Delta r$. The boundary for the k^{th} layer begins at $r = r_o + (k-1)\Delta r$ and extends to $r_o + k\Delta r$. We will assume that the refractivity for the k^{th} layer is given by $N((k-1)\Delta r)$, although in principle it is possible to take into account changes in the refractivity within the layer by using, for example, the average refractivity within the layer. The ray forms an angle $\theta(r)$ with the outward normal to the spherical boundary where it intersects the boundary at radius r. Similarly, the

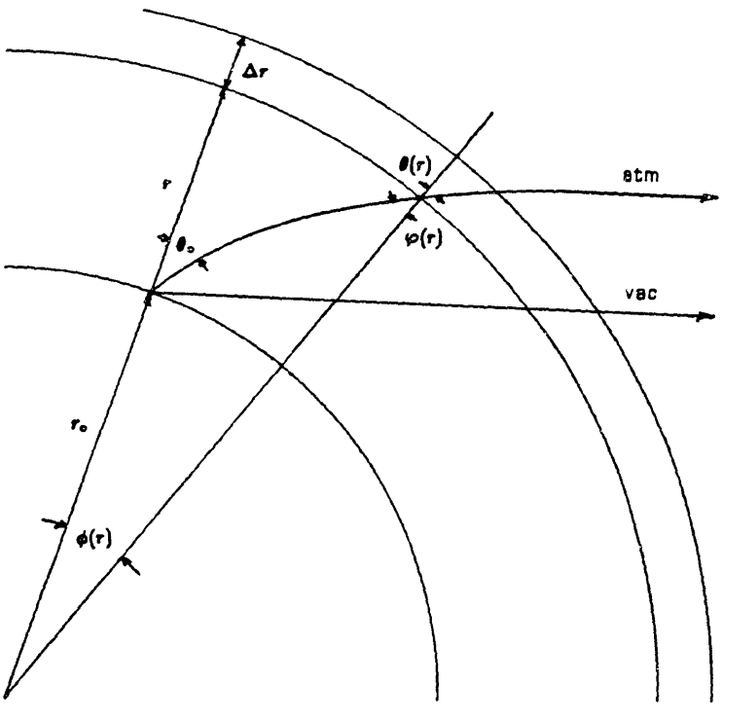


Figure 3.2.1. Geometry and definition of parameters for ray-trace calculations, described in Section 3.2.

ray forms an angle $\vartheta(r+\Delta r)$ with the inward normal to the spherical boundary where it intersects the boundary at $r+\Delta r$. The position angle $\phi(r)$ is the angle between the radius to the origin of the ray, and the intersection of the ray and the boundary at radius r. We will discuss the relationship between the angles below.

We will begin our calculations at $r = r_o$, although it is possible to begin the calculations at $r =$"∞." (Unless we qualify the term "radius," we will mean the distance from the center of the spherical earth to the point in question.) We will first describe the calculations for $r = r_o$, and then for any r. For each layer, we will increment three quantities. The first quantity corresponds to the first integral in (1.2.15), and represents the decreased velocity with which the ray travels through the atmosphere. For the $r = r_o$ calculation we begin with

$$\tau_a^r(r_o) = 10^{-6} N(0)\ \Delta r \sec\theta_o \tag{3.2.1}$$

The term $\tau_a^r(r)$ represents the "reduced velocity" term which includes the layers up to and including the layer whose inner boundary is r. (The superscript r is for "reduced" or "retarded.") The refractivity is parametrized by the altitude $r - r_o$, according to convention. We have retained the use of the subscript for θ_o— rather than writing $\theta(r_o)$—because this term has a special meaning: it is the "refracted" zenith angle. Since beforehand we do not know the refracted zenith angle, at the beginning of our calculation we assume some value for θ_o. Then, at the end of our calculation, if the ray is not directed at the source, we calculate an adjustment to θ_o and repeat the ray–trace. We continue to repeat the ray–trace until we are satisfactorily close to the required direction. However, for the purposes of determining the form for a mapping

function (see Appendix A), we do not need to know the delay in any *specific* direction. It is necessary to know only the delay, and the corresponding elevation.

At the same time the calculation for the "decreased velocity" term is made, the increment to the geometric path is calculated. This increment is

$$\rho_a^g(r_\circ) = \Delta r \sec\theta_\circ \tag{3.2.2}$$

We have used the letter ρ instead of τ to stress that ρ_a^g is a length and not a delay. The function $\rho_a^g(r)$ is the cumulative value of the "decreased velocity" term for all layers up to and including the layer with inner radius r.

We also want to keep track of the position angle $\phi(r)$. Equation (1.2.14) is an integral for this quantity. The increment for the first layer yields

$$\phi(r_\circ + \Delta r) = \frac{\Delta r}{r_\circ}\tan\theta_\circ \tag{3.2.3}$$

Finally, we wish to calculate the angle of incidence $\varphi(r_\circ + \Delta r)$ of the ray at the spherical boundary at $r + \Delta r$. This angle can be calculated using the law of sines. One leg is the radius to the intersection of the ray and the spherical boundary at radius $r_\circ$, and the other is the radius to the intersection of the ray and the spherical boundary at $r_\circ + \Delta r$. The law of sines yields

$$\varphi(r_\circ + \Delta r) = \frac{r}{r + \Delta r}\sin\theta_\circ \tag{3.2.4}$$

We will now perform the ray–trace calculations for the general radius r (refer to Figure 3.2.1). We will assume that we have previously calculated $\phi(r)$, the position angle of the intersection of the ray with the spherical boundary at r, and $\varphi(r)$, the

angle of incidence of the ray at this boundary. We first must calculate $\theta(r)$. For the first layer, this angle was a "given," and so we did not need to calculate it then. Now, in order to calculate this angle, we must use Snell's law of refraction for a ray with angle of incidence $\varphi(r)$, and refractive index $n(r-\Delta r) = 1 + 10^{-6}N(r-\Delta r)$ in the layer of the incident ray and $n(r) = 1 + 10^{-6}N(r)$ in the layer of the refracted ray. This yields for the sine of $\theta(r)$

$$\sin\theta(r) = \frac{n(r)}{n(r-\Delta r)}\sin\varphi(r) \tag{3.2.5}$$

Since $\theta(r) < \frac{\pi}{2}$, $\theta(r)$ can be determined unambiguously from $\sin\theta$.

Having $\theta(r)$, we can now calculate the contributions to the "reduced velocity" delay τ_a^r, the geometric path ρ_a^g, the position angle ϕ, and also the angle of incidence φ for the next layer:

$$\tau_a^r(r) = \tau_a^r(r-\Delta r) + 10^{-6}N(r)\ \Delta r\ \sec\theta(r) \tag{3.2.6}$$

$$\rho_a^g(r) = \rho_a^g(r-\Delta r) + \Delta r\ \sec\theta(r) \tag{3.2.7}$$

$$\phi(r+\Delta r) = \phi(r) + \frac{n_\circ\, r_\circ\, \sin\theta_\circ}{r\sqrt{n^2r^2 - n_\circ^2 r_\circ^2 \sin^2\theta_\circ}}\Delta r \tag{3.2.8}$$

$$\sin\varphi(r+\Delta r) = \frac{r}{r+\Delta r}\sin\theta(r) \tag{3.2.9}$$

The integration proceeds up to some limit r_f. (The choice of r_f must be such that $N(r_f) \simeq 0$ above this radius; we have found that $r_f = 100$ km is adequate.) A final step in the calculation is to calculate the straight–line "*in vacuo*" distance from the origin of the ray to the final position (see Section 1.1). This distance can

be calculated from the law of sines using the position angle ϕ and the original zenith angle θ_o:

$$\rho_v^g(r_f) = \frac{\sin\theta(r_f)}{\sin\left(\theta(r_f) + \phi(r_f)\right)} r_f \tag{3.2.10}$$

The superscript v in (3.2.10) indicates the distance the ray would have traveled in vacuum to reach the point at radius r_f, and the superscript g is to remind us that this quantity is to be combined with ρ_a^g of (3.2.7) to yield the "geometric propagation delay" as in (1.1.2). Combining these delays concludes the calculation of the propagation delay τ_a:

$$\tau_a = \tau_a^r(r_f) + [\rho_a^g(r_f) - \rho_v^g(r_f)] \tag{3.2.11}$$

Before discussing the accuracy of the ray–trace algorithm, we must first present our formulas for the determination of the profile of refractivity. We have already given our formula for the refractivity at any given point in space. The refractivity depends on the pressure, temperature, and humidity at that point, and we repeat its formula here for easy reference:

$$N = k_1 R_d \rho + \left(k_2 - \frac{M_w}{M_d} k_1\right) R_v \rho_v + k_3 R_v \frac{\rho_v}{T} \tag{3.2.12}$$

As in previous chapters, R_d is the specific gas constant for dry air and R_v that for water vapor, M_d is the molar mass for dry air and M_w that for water vapor, T is the absolute temperature, ρ is the *total* density, and ρ_v the density of water vapor. The origin of (3.2.12)—and values for k_1, k_2, and k_3— were discussed in Section 1.2.

How do we determine the refractivity at some height above the surface? One way is to use radiosonde determinations of P, T, and relative humidity, from which N can then be calculated. However, we chose to derive the profile from a model atmosphere,

because radiosonde data were at that time not available to us. This model atmosphere has the following properties:

- The temperature decreases from its surface value T_o with constant lapse rate β ($\beta < 0$). Both T_o and β are user–definable. This decrease with height continues until the temperature reaches T_{min} (also user–definable), at a height of $(T_{min} - T_o)/\beta$. Above this height, the temperature remains constant.
- The relative humidity remains constant at the value ϱ (user–definable) between the surface and a height of 12 km (not user–definable). The partial pressure of water vapor is calculated using the Clausius–Clapeyron equation for the saturation pressure of water vapor at temperature T [Hess, 1959], which can be expressed with an error of less than 1% in the range 240–310 K by [Crane, 1976]

$$\ln\left(\frac{p_{v_s}}{6.11}\right) = \frac{25.2(T-273)}{T} - 5.3\ln\left(\frac{T}{273}\right) \tag{3.2.13}$$

where p_{v_s} is the saturation pressure of water vapor in mbars and T is the temperature in Kelvins. The partial pressure of water vapor p_v is determined from the saturation pressure using

$$p_v = \varrho p_{v_s} \tag{3.2.14}$$

and the constituent density of water vapor ρ_v determined using the ideal gas law. Above 12 km, the relative humidity is zero.

- The atmosphere is in hydrostatic equilibrium. The equation of hydrostatic equilibrium is integrated with respect to height simultaneously with the ray–trace

equations to determine the pressure. Thus we find for the pressure at the radius r

$$P(r) = P(r - \Delta r) - \rho(r - \Delta r)g\Delta r \tag{3.2.15}$$

The acceleration g due to gravity is constant with height and has a value of 9.784 m s^{-2} [Saastamoinen, 1974]. The density is calculated by adding the constituent densities of water vapor and dry air, both calculated individually from the ideal gas law and the appropriate partial pressure. The density of water vapor is calculated using the procedure above, and the partial pressure p_d of dry air is calculated using the law of partial pressures, $p_d + p_v = P$. The value of the surface pressure P_o is user–definable.

Many of the features of our model atmosphere are simplified versions of reality. The easiest way to examine the effects of these simplifications is to increase the complexity of the model, and compare the results to the ray–trace using the simplified model. We will perform this comparison in later chapters, after we have presented the results of ray–tracing with the simplified model. In the following section we will define the new mapping function. Then, in Section 3.4, we will discuss using ray–trace to calculate values for the mapping function, which will be parametrized in terms of the user–definable values used in the atmospheric model.

3.3 Defining the mapping function

In this section we detail the development of a new mapping function for the dry atmosphere. We have already described the ray–trace algorithms used, and so in this

section we will describe how we modeled the ray–trace results and determined the dependence of the mapping function on meteorological variables.

As we discussed above, we did not wish to spend a great deal of effort and time on developing new expressions for the mapping function. For this same reason, we chose to model our ray–trace results using a known form for the mapping function. This form was a continued fraction formula developed by Marini, and modified as per the Chao mapping function so that the value of the mapping function would be unity in the zenith direction (see Section 1.4). In order to obtain accuracy for low elevation angles, we retained one more term than had either Chao or Marini. We therefore arrived at the expression

$$m_d(\epsilon) = \cfrac{1}{\sin\epsilon + \cfrac{a}{\tan\epsilon + \cfrac{b}{\sin\epsilon + c}}} \tag{3.3.1}$$

The parameters a, b, and c were to be determined from the ray–trace data. We chose to develop a mapping function for the dry atmosphere because an error in the mapping function for this component has a much larger effect than a comparable error in the mapping function for the wet component, as the dry delay is on the average a factor of 10–20 times the wet delay. Using our definition for dry delay presented in Section 1.3, we will define the dry mapping function to be

$$m_d(\epsilon) \equiv \frac{10^{-6} k_1 R_d \int_{atm} ds\, \rho(s) + \tau_g^a}{\tau_d^z} \tag{3.3.2}$$

The first term in the numerator of (3.3.2) is the integral along the path of the ray of the first term in the refractivity as presented in (3.2.12). The second term in the

numerator of (3.3.2) represents the "geometric" delay discussed frequently and given by (1.3.19). The denominator of (3.3.2) is the zenith dry delay, defined in (1.3.15). The choice for this definition for the dry mapping function is made with the availability of WVR's in mind. In Chapter 2, we saw that these instruments can be used to estimate the "wet" delay, defined to be

$$\tau_w \equiv 10^{-6} R_v \int_{atm} ds \left(k_2 - \frac{M_w}{M_d} k_1 + \frac{k_3}{T} \right) \rho_v(s) \tag{3.3.3}$$

The wet delay can be recognized to be the integral along the path of the ray of the final terms in the refractivity (3.2.12).

With the definitions presented above, the algorithm for routinely estimating the propagation delay is as follows:

(1) Using WVR data, estimate the wet delay as defined above and as detailed in Chapter 2. We will call this estimate of the wet delay $\hat{\tau}_{WVR}$, to stress the origin of this estimate. If a WVR is not available, use some other estimate $\hat{\tau}_w$, making sure that the definition (3.3.3) is obeyed.

(2) Use surface pressure to estimate the zenith dry delay, as defined in (1.3.15). Since under conditions of hydrostatic equilibrium the dry delay as defined in (1.3.15) is proportional to the surface pressure, and that constant of proportionality can be calculated very accurately, the estimate of the dry zenith delay is extremely reliable. (We will discuss nonequilibrium effects later.) The estimate of the dry zenith delay will be denoted $\hat{\tau}_d^z$.

(3) Estimate the dry mapping function, using (3.3.1), the elevation angle of the source, and the formulas for a, b, and c given in the next section. This estimate will be denoted $\hat{m}_d$.

(4) Combine the above estimates to form the estimate $\hat{\tau}_a$ of the atmospheric propagation delay:

$$\hat{\tau}_a = \hat{\tau}_d^z \hat{m}_d + \hat{\tau}_{WVR} \tag{3.3.3}$$

We have detailed the process of the estimation of the propagation delay explicitly because it is important that the definitions of the individual contributions to the delay (*i.e.*, the two zenith delays and the dry mapping function) be shown to be internally consistent, as indeed they are for the definitions above. Furthermore, from these "instructions," one can see that we have "optimized" the definition of our mapping function for use with a WVR, so that the surface pressure may be used to estimate the dry delay. It is important when presenting a new mapping function to define the zenith delay to be used in conjunction with the mapping function, and also to define all the components of the delay.

The paper which describes the estimation of the mapping function parameters from the ray–trace calculations, and the parametrization of the mapping function parameters in terms of the user–definable atmospheric model parameters, is included in Appendix A. This paper also describes some of the initial testing of the new mapping function (called CfA–2.2). This testing consists of comparing the results of elevation–cutoff tests for the Marini and the CfA–2.2 mapping functions. The estimates of baseline length made using the Marini mapping function exhibit behavior that we have previously described as being indicative of a mapping function error; the CfA–2.2 mapping function, however, seems to be relatively free from these errors. It is important to note, however, that the data used were all obtained within one month of each other. Therefore, the possibility of seasonal effects still exists. In order to test the CfA–2.2 mapping function more rigorously, a series of VLBI experiments involving a number of observations from low elevation angles has been devised. The following chapters describe these experiments and their results.

Chapter 4

Experimental Results

Introduction

Table 3 of Appendix A contains an interesting result. In this table, we summarize the elevation-angle cutoff tests for the Marini and CfA-2.2 mapping functions by presenting the changes in the estimates of the local vertical site positions corresponding to the baseline length differences from each of the elevation-angle cutoff tests. From this table, we can see that the changes in the vertical estimates are nearly equal for the two mapping functions, for sites having no data from below the elevation of 15°. Apparently, if we restrict our VLBI observations to above 15° elevation, we limit our systematic errors due to mapping function errors to the mm-level (assuming that both mapping functions are effectively free from error above this elevation angle). Why, then, did we need a new mapping function at all?

The answer to this question really lies in the answer to the question: Do we need—or simply want—ever to observe at elevation angles lower than 15°? Or even 20°?

There are several reasons why observations in this region are useful, even necessary. Geometric limitations imposed by the locations of the antennas sometimes are such that a large portion of the time some particular source is "mutually visible," that source is at a very low elevation as viewed from one of the sites. Figure 1.5.3

illustrates another reason for having observations in this region. If the *a priori* value of the zenith delay is thought to be incorrect, and a zenith-delay parameter must be estimated, then the uncertainty in the estimate of the vertical coordinate for an elevation cutoff of 20° is threefold its value for an elevation cutoff of 5°. Although with WVR data and accurate knowledge of the weighting function (see Chapter 2) the estimation of a zenith-delay parameter may be obviated, there is still a large number (over 100,000) of VLBI group-delay data alone for which there is no WVR data, and that number is steadily increasing as of this writing. Therefore, it is important to have a mapping function which is accurate over all ranges of elevation angles for which VLBI data exist, which is all elevation angles down to about 5°.

In this chapter, we discuss the results from a series of VLBI experiments intended to test more thoroughly the newly developed CfA-2.2 mapping function (see Chapter 3). For these experiments, as we will see, a large fraction of the data is from observations at low elevation angles; for this reason the experiments are called "low-elevation experiments." We will first describe the design of these experiments, including the site and sources, and the data processing. We will also introduce a set of "standard" solutions to which we can later compare solutions obtained with different processings (*i.e.*, with different atmospheric models or parametrization). These different processings will be used, in Section 4.2, to investigate the accuracy of the dual-frequency algorithm used to reduce the WVR data available during some of these experiments. In Section 4.3, we will test the new mapping function, using elevation-angle cutoff tests (as in Appendix A), and also using alternate models, parametrizations, and subsets of data. The final section of this chapter will be devoted to a

discussion of the implications of these experimental results for VLBI data processing. This final discussion will rely heavily on the material presented in Sections 1.4 through 1.6.

4.1 "Low-elevation" experiments

In this section we will describe a series of VLBI experiments designed to allow us to examine the systematic error arising from the atmospheric delay formulas used. We will describe the sites and sources used, and present a set of solutions which we will characterize as the "standard results."

4.1.i Sites

Two main criteria dictated the choice of sites for the low-elevation experiments: elevation coverage and availability. (Slew speed was also important but did not affect our choice of antennas: see Section 4.1.*ii*.) We selected the Mojave antenna located in the Goldstone antenna complex in California because it is available; for most months of the year it is used for only one day in conjunction with mobile VLBI experiments. Another advantage of using this site is that it is the test site for the retrofit R-series of WVR's (see Chapter 2). Furthermore, prior to the arrival of the first retrofit WVR, this site possessed a WVR belonging to the original R-series. One disadvantage of using the Mojave antenna is that it is unable to observe below about 8° elevation.

The antenna originally used for the other terminal of the interferometer was the Haystack antenna in Westford, Massachusetts. This antenna was chosen for its lower elevation-angle limit of 3.5°. However, its availability was limited since this antenna

is frequently used in conjunction with astronomy experiments. We decided to use this antenna instead of the Westford antenna, located 1.2 km from the Haystack antenna, because the greater sensitivity of the Haystack antenna allowed observations of shorter duration to obtain the same signal-to-noise ratio (SNR). Observations of shorter duration are preferred for two reasons: (*i*) more observations can be obtained in a given amount of time, and (*ii*) there is less possibility of atmospheric variations during the observation to affect the measurement. Several months after the first experiment, the Westford antenna was upgraded with a low-noise receiver, and we began to use this antenna for the low-elevation experiments. The availability of this antenna is greater, and in practice the elevation coverage is the same as the Haystack antenna. In reality there is a region of the sky to the northeast where the Westford antenna cannot observe below 6–8°, but no observations were scheduled in these directions, because sources there are not visible at the Mojave site. Hereafter, if the difference is unimportant, we will refer to both Haystack and Westford as Haystack.

4.1.ii Radio sources

The radio sources were selected on the basis of two criteria. The sources must have been included in the geodetic database comprising, by September 1984, approximately 60,000 group delay observations. These observations yield accurate source positions for the analysis of the low-elevation experiments. The sources must also be strong enough that the integration time necessary for a SNR of at least 7 is less than

about 400 s, for the reasons given in Section 4.1.*i* concerning short integration times. (An SNR of 7 is indicative of a "solid" detection: see Clark *et al.* [1985].)

The choice of the number of sources to use required a somewhat arbitrary decision. The total number of sources meeting the criteria above was over 40 at the time of the designing of the experiments. If all these sources were to be used, then it would be possible to obtain only 3–5 observations per source. In a typical geodetic schedule, only 10–15 sources are used in any single experiment. We chose to use 14 sources, selecting the strong sources while attempting to obtain a uniform distribution in right ascension, so that the time between having a source at low elevations at any one site is minimized. One of these sources (1803+785) is in fact visible at all times at both sites; we believed that observing this source throughout an observing session would help us separate atmospheric and clock noise (see Section 4.1.*iv*).

The sources and their nominal positions are listed in Table 1.4.1. The nominal positions of all sources except 1502+106 were obtained from the global solution described in Herring *et al.* [1986], although these positions were not reported there. The position of 1502+106 was obtained by rotating in right ascension the position of that source reported in Ma *et al.* [1986] by -0.2 ms, which was the amount needed to align the position of the source 37273B reported in that paper with the position determined from the global solution mentioned above.

4.1.iii Observing schedule

In developing an observing schedule, we simply tried to observe a source for a period immediately preceding its setting, or following its rising. (We also attempted to

Table 1.4.1

Radio Sources for "low elevation" experiments

IAU name	Common name	Right Ascension			Declination			Symbol
		hh	mm	ss.ssss	dd	mm	ss.ssss	
0106+013	—	1	8	38.7708	1	35	0.3215	A
0234+285	—	2	37	52.4056	28	48	8.9915	B
0300+470	—	3	3	35.2422	47	16	16.2764	C
0528+134	—	5	30	56.4166	13	31	55.1490	D
0552+398	—	5	55	30.8055	39	48	49.1640	E
0851+202	OJ287	8	54	48.8747	20	6	30.6384	F
0923+392	4C39.25	9	27	3.0136	39	2	20.8495	G
1226+023	3C273B	12	29	6.6995	2	3	8.5962	H
1502+106	—	15	4	24.0860	10	29	47.2147	I
1641+399	3C345	16	42	58.8096	39	48	36.9948	J
1741−038	—	17	43	58.8558	−3	50	4.6132	K
1803+784	—	18	0	45.6828	78	28	4.0199	L
2216−038	—	22	18	52.0375	−3	35	36.8765	M
2251+158	3C454.3	22	53	57.7477	16	8	53.5653	N

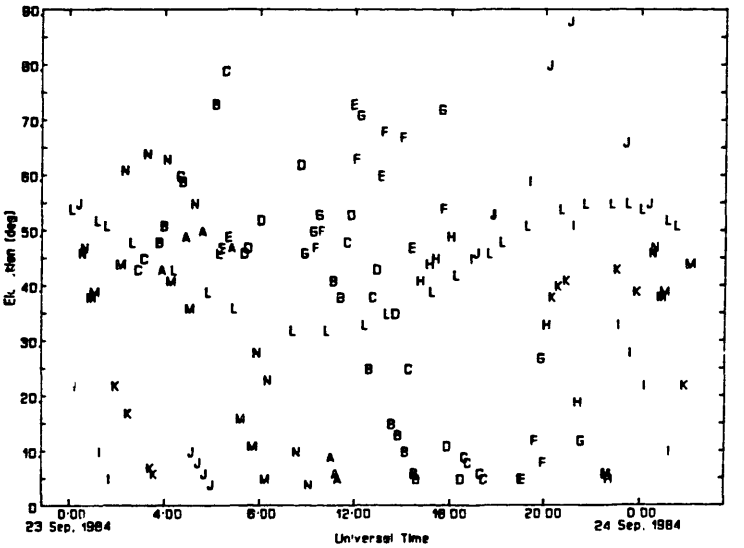
Positions of radio sources are given in J2000 coordinates. These positions were held fixed during processing except for the positions of 1803+784 and 1502+106 (see Section 4.1.*iv*). The symbols are those used to identify the sources in Figure 4.1.1. For references to source positions, see text.

have as many observations in a given amount of time as possible: see below.) A source rising at Haystack is not visible at Mojave, so all observations having low elevations at Haystack involve setting sources. Similarly, a source setting at Mojave is not visible at Haystack, and so all observations having low elevations at Mojave involve rising sources. Thus, developing a schedule for the low-elevation experiments was a matter of observing the sources as they rose at Mojave and as they set at Haystack. Often, we chose to "track" the sources in order to increase the number of observations at low elevation angles. This tracking is contrary to the technique usually used in a geodetic experiment. During a period in which no sources were rising or setting, the visible sources were observed at higher elevation angles.

Figure 4.1.1 shows the elevation angle as a function of the time of the observation for both Haystack and Mojave. The same schedule was used for all the experiments discussed here. The letters in Figure 4.1.1 correspond to the source observed (see Table 4.1.1). Note that the "tracking" of the sources at low-elevation angles is evident. The fraction of observations below 10° elevation at Haystack exceeds 15%.

As we mentioned above, we attempted to have as many observations in a given amount of time as possible. As the time between two observations becomes shorter, the effect of the clock variations decreases, with the amount of decrease depending on the sizes of the clock variations. Ideally, we would like to obtain an instantaneous "snapshot" of the sky. We are prevented from obtaining this "snapshot," however, for several reasons. Primary obstacles are the "integration periods" (see Herring [1983]) required to obtain sufficient signal-to-noise ratios for the various radio sources; these periods range from 100–400 sec for these sources and antennas. Another problem is the

(a)



(b)

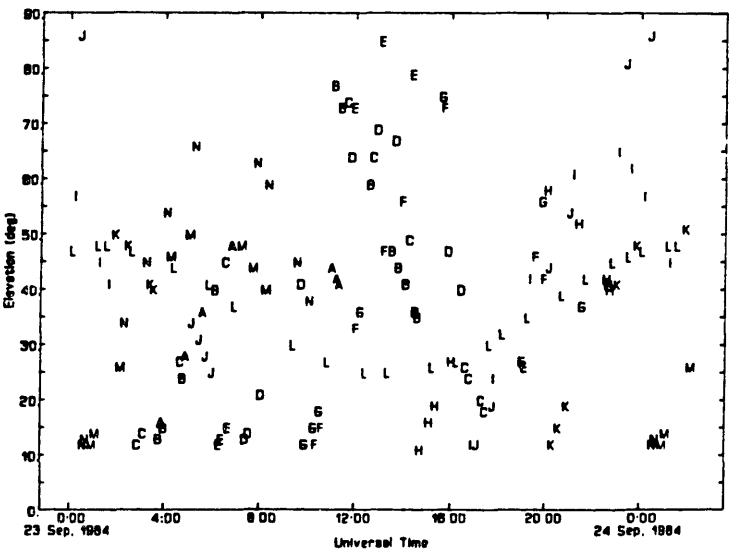


Figure 4.1.1. Elevation angle as a function of time for the Haystack (a) and Mojave (b) antenna. The letters refer to the radio source being observed (see Table 4.1.1).

nonzero time for the antenna to slew between the radio sources, which can be located far apart on the celestial sphere. In order to overcome these problems, we (*i*) used the fast-slewing radio telescopes of Mojave, Haystack, and Westford, and (*ii*) "tracked" sources when they were close to the horizon. Unfortunately, the limited availability of Mark III recording tapes required that we reduce the number of observations acquired using these methods. Therefore, we inserted an amount of idle time between the end and start of successive observations such that the average time between start epochs was about 10 min. Late in 1985, we changed the observing schedule to take advantage of the new high-density tape recorders at Mojave and Westford. We were then able to schedule approximately 245 observations in an experiment and use only $\sim 2\frac{1}{2}$ tapes per site whereas, previously, we had made approximately 160 observations per experiment while using 33 tapes per each site. Schedules spanned 26 hours for both the old and the new experiments. The average time between start epochs for the new schedule is ~6 min. At present, we have not analyzed any data from the experiments with these new schedules, because the schedules were implemented too recently for any of those data to be processed.

In the following section, we describe briefly the technique used to process the data from the low-elevation experiments.

4.1.iv Data processing

To process the data from the low-elevation experiments, we have used a Kalman filter implemented for VLBI data processing. The term "Kalman filter" refers to a specific technique, developed by R.E. Kalman [Kalman, 1960; Kalman and Bucy, 1961]

for performing sequential least-squares parameter estimation, for which some subset of the parameters are the values of a time-varying stochastic process. The statistical properties of this stochastic process are known. Full derivations of the Kalman filter equations are given elsewhere (see, *e.g.*, Liebelt [1967]), and a general discussion of the Kalman filter is given in Section 4.4.*i*. In order for the filter to be completely defined, aside from partial derivatives which are given elsewhere [Whitney, 1974; Robertson, 1976], we must describe the state transition matrix, which relates the expectation of the parameters at some time to the expectation of the parameters at a previous time, and the covariance matrix of stochastic processes which are being modeled.

The parameters estimated in our solutions have been divided into two types: deterministic parameters, whose values are assumed not to change over the duration of an observing session (of approximate duration 26 hours for the low-elevation experiments); and stochastic parameters, the models for which we give below. The deterministic parameters include site coordinates, corrections to the IAU 1980 nutation series [Herring *et al.*, 1986], and source coordinates. The positions of 12 of 14 radio sources are constrained to the values determined by a global least-squares solution involving $\sim 70,000$ data, mentioned above and described in Herring *et al.* [1986]. Corrections to the position of the source 1502+106, taken from Ma *et al.* [1986] and rotated as described above, were estimated because the uncertainties for the position of that source were large due to its being infrequently observed. (The inclusion of this source violates the criterion given in Section 1.4.*ii*, but we did not realize the unreliability of the position determination for this source when designing the low-elevation experiments.) The standard deviations reported in that paper, and used as the *a*

priori uncertainties in the solutions, were 0.02 ms in right ascension and 1.1 mas in declination. We also estimated corrections to the *a priori* position of 1803+785. This was unintentional, but the Kalman-filter software was being tested when we began analyzing the data, and the control file we used indicated that these should be estimated, and we did not change this command. However, the *a priori* standard deviations of 0.09 ms in right ascension and 0.2 mas in declination were obtained from the global solution mentioned above, and these uncertainties were small enough that the corrections were all constrained to $\lesssim 1.5\ \sigma$. Global-solution capability is currently being developed for the Kalman filter processor, so that source positions may be estimated simultaneously using all the low-elevation experiments.

For each stochastic element of the model for the group delay, two parameters are in general estimated. Each stochastic process is modeled as the sum of two independent processes: an integrated random walk process and a random process. A white noise, Gaussian process is assumed to be generating the random walk process in both cases, and these processes are assumed to be independent of each other. The two parameters to be estimated, ϕ and ω, at time t_i, can be respectively interpreted as being an "offset" and a "rate," and can therefore be expressed in terms of their values at time t_{i-1} as

$$\begin{bmatrix} \phi(t_i) \\ \omega(t_i) \end{bmatrix} = \begin{pmatrix} 1 & \Delta t_i \\ 0 & 1 \end{pmatrix} \begin{bmatrix} \phi(t_{i-1}) \\ \omega(t_{i-1}) \end{bmatrix} + \begin{bmatrix} \Delta\phi(t_i) \\ \Delta\omega(t_i) \end{bmatrix} \tag{4.1.1}$$

where $\Delta t_i = t_i - t_{i-1}$, and $\Delta\phi(t_i)$ and $\Delta\omega(t_i)$ are the (unknown) random changes in the respective parameters which are to be estimated. The covariance matrix of the

final vector on the right hand side of (4.1.1), which in effect yields constraints on the solutions for this vector, is

$$\begin{pmatrix} \frac{1}{3}\sigma_\omega^2(\Delta t_i)^3 + \sigma_\phi^2 \Delta t_i & \frac{1}{2}\sigma_\omega^2(\Delta t_i)^2 \\ \frac{1}{2}\sigma_\omega^2(\Delta t_i)^2 & \sigma_\omega^2 \Delta t_i \end{pmatrix} \tag{4.1.2}$$

where σ_ω^2 is the assumed variance associated with the rate term controlling the integrated random walk, and σ_ϕ^2 is the assumed variance associated with the random walk process. Note that by setting $\sigma_\omega^2 = \sigma_\phi^2 = 0$, one can effectively convert that process to a deterministic process.

In the following section, we describe the atmospheric models used for the "standard solutions," and present estimates of baseline length for those solutions.

4.1.v "Standard solutions"

Table 4.1.2 contains a list of the low-elevation experiments conducted to this date. In this section, we will describe a set of "standard solutions" for these experiments. In later sections, we will analyze these experiments using different mapping functions for the dry delay and different *a priori* values for the wet delay, and the standard solutions will be the references with which these later solutions will be compared. Below, we describe the atmospheric models used and parameters estimated for these standard solutions.

- *Surface weather data.* The pressure, temperature, and relative humidity at the sites were monitored during the low-elevation experiments. These data were used in the various formulas as called for below.

Table 4.1.2

Low-elevation experiments

Experiment Date	Interferometer	WVR at Mojave?
84/09/23	Mojave–Haystack	Yes
84/11/16	Mojave–Haystack	Yes
85/01/25	Mojave–Haystack	Yes
85/03/20	Mojave–Westford	Yes
85/05/30	Mojave–Westford	Yes
85/08/13	Mojave–Westford	No
85/09/12	Mojave–Westford	No
85/10/02	Mojave–Westford	No

- *Zenith dry delay.* The Saastamoinen formula for the zenith dry delay was used in the standard solutions: see (1.4.1) and (1.4.2).
- *Zenith wet delay.* If WVR data are available for a particular observing session, then these data are used for the *a priori* values of the wet delay for this experiment (see below). If WVR data are not available, then the *a priori* value for the zenith wet delay is taken to be zero. While a model for the zenith wet delay may be used, we have chosen not to do so since these models are based on the humidity at the surface, which has been shown to be uncorrelated in general with the integrated vapor content [Reber and Swope, 1972]. The extent of this decorrelation seems to depend on the climate. Elgered *et al.* [1985] have examined seven sites distributed globally, and have used radiosonde data to determine the constants c_0 and c_1 for the model $L_w = c_0 + c_1(p_v)_\circ$, where L_w is the wet propagation delay, and $(p_v)_\circ$ is the partial pressure of water vapor at the surface. The regression was done individually for each site, and by season. The smallest RMS difference of the true delay from the resulting model was 0.9 cm, for Barrow, Alaska in the Spring. In general, the RMS difference was 2–3 cm. Interestingly enough, they found one site—Singapore—which exhibited almost no correlation at all (*i.e.* $c_1 \simeq 0$), and for which the regression led to a nearly constant value for the wet delay.
- *Dry mapping function.* The CfA–2.2 mapping function is used for the mapping function of the dry delay in the standard solutions. The development of this mapping function is detailed in Chapter 3 and Appendix A. For the temperature lapse rate and height of the troposphere, the nominal values are used: see Appendix A.

- *Wet mapping function.* See *WVR data*, below.
- *WVR data.* The first five low-elevation experiments are shown as having WVR data available. In fact, it is believed that "good" WVR data are available for a subset of the VLBI observations for each of the remaining three observing sessions, but at this time we cannot determine which, if any, of the data are "good." Therefore, the standard solutions will not use WVR data from these later experiments. For these first five experiments, WVR data were obtained by means of tip-curve calibration (see Section 2.5). The tip curves yield values for the zenith opacity at each of the two frequencies of the WVR. The dual-frequency delay equation (2.3.10) was then used to determine the zenith wet delay. A single value for the weighting function was used for all experiments. This value of 163.9 cm neper^{-1} was determined by Resch [1984], and is based on a "Method 2" study using several radiosonde launch sites distributed across the U.S. (see Section 2.6). The "oxygen correction" was also modified, as discussed in Section 2.6.*iii*. After determining the zenith wet delay, we then "mapped" the delay to the line-of-sight delay by multiplying the zenith delay by the cosecant of the elevation. Although we realize that the cosecant law may be a poor approximation to the "true" wet mapping function, there are no available studies which present a reliable wet mapping function. Note also that a WVR was available only at the Mojave site, from which no observations below 10° elevation were made.
- *Parameter estimation.* The data for each observing session were processed separately from those from the other observing sessions. With the exception noted

below, the solution was parametrized as discussed in Section 4.1.*iv*. However, for those sessions with WVR data, the atmospheric parameters for the Mojave site were limited to a single, constant additive atmospheric parameter. (As with other parameters, one additive atmospheric parameter was estimated independently for each observing session.) The purpose of this parameter was to estimate any bias in the dual-frequency wet path delay algorithm described above.

- *Stochastic process variances.* The values for the variances used to represent the "clock" and atmosphere stochastic processes were chosen to approximate the known statistical behavior of those physical processes on time scales of up to about 20 minutes, representing the maximum time between VLBI observations. For the atmospheres, we used $\sigma_\phi^2 = 0.5$ psec^2 sec^{-1} and $\sigma_\omega^2 = 0$. The "clock" performance is specified differently. We specified that the square root of the Allan variance of the clock variations must be 2×10^{-14} after 20 min, and the actual values for σ_ϕ^2 and σ_ω^2 are determined in software. We are currently investigating the "best" variances to use, and there is some indication that these variances are too large (see below). However, we know from experience that we can change these variance by one order of magnitude in either direction and produce changes in the estimated baseline length of less than 1 σ.

Figure 4.1.2 shows the estimates of the Mojave–Haystack baseline length for the standard solutions. (The estimates of the Mojave–Westford baseline length have been "mapped" onto the Mojave–Haystack baseline using the precisely determined Haystack–Westford baseline vector: see Clark *et al.* [1985].) The error bars shown

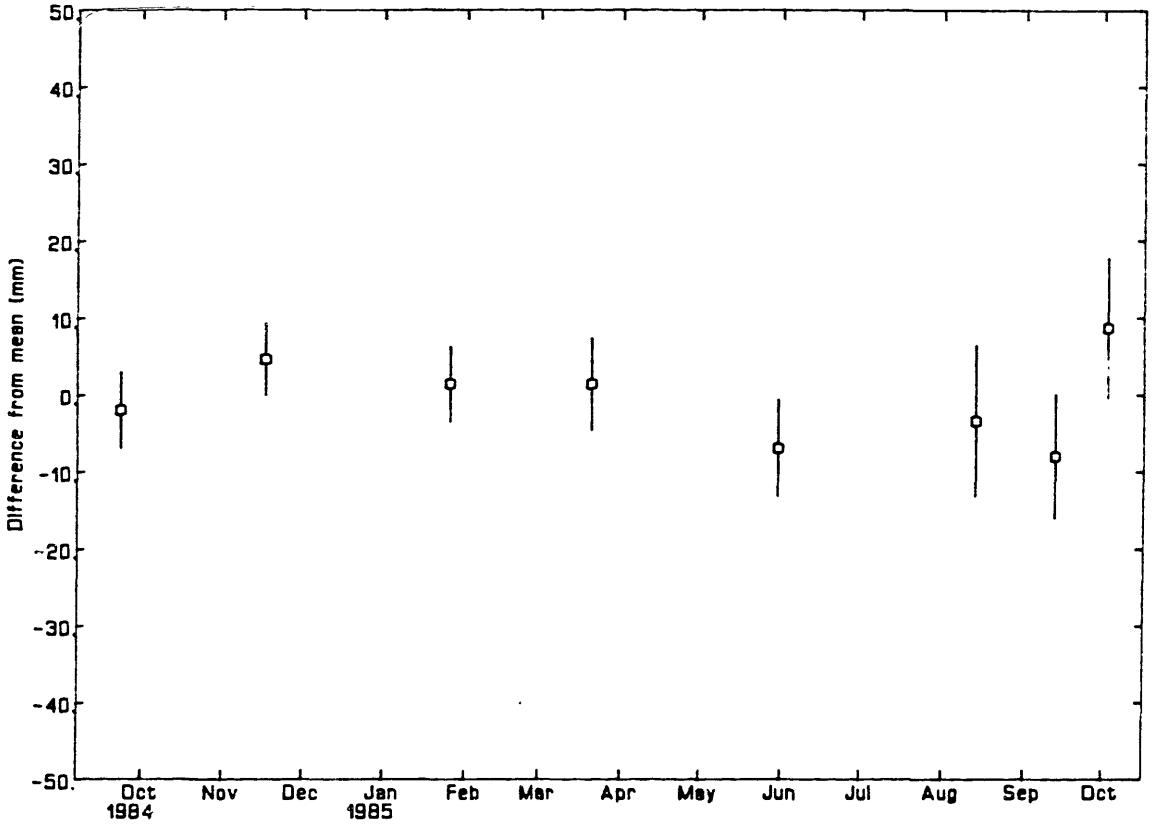


Figure 4.1.2. "Standard results" for the Mojave–Haystack baseline length from the "low-elevation" experiments. The error bars are the statistical standard deviations of the estimates based on the SNR-derived group-delay uncertainties and the values for the stochastic-process variances used to represent the clock and atmosphere behaviors (see text). The estimates of the Mojave–Westford baseline length from the experiments involving those sites have been "mapped" to the Mojave–Haystack baseline using the precisely determined Haystack–Westford baseline vector [Clark *et al.*, 1985].

are the statistical standard deviations of the estimates based upon the noise variances used in the Kalman filter and the SNR-derived uncertainties of the group delay data. (Only group-delay data were used—no phase-delay rate data were used.) The differing sizes of these error bars can be attributed to different numbers of "good" data, to the slightly lower sensitivity of the Westford antenna compared to the Haystack antenna, and to the estimation of stochastic parameters at Mojave for those experiments with no WVR data.

The root-mean-square scatter of the eight baseline determinations about their weighted mean is approximately 4 mm. This scatter is only about 0.8 the value of the expected scatter based on the standard errors of the estimates. This small scatter may imply an overestimation of the variances of the stochastic processes for the clock and atmospheres (see above). However, it is not true in general that the scatter of the baseline length is less than, or even consistent with, the scatter predicted by the standard errors. Typical values for the scatters are approximately 1.1–1.3 times the predicted values. Furthermore, the purpose of the low-elevation experiments is to expose systematic errors, none of which are evident in the standard solutions.

It is often useful to examine the postfit residuals of the group-delay observations for individual experiments. We will in fact look at these residuals later. Now we will test various features of the atmospheric models used to generate the standard solutions. We will begin with the "WVR algorithm."

4.2 Accuracy of the dual-frequency WVR algorithm

In this section, we will discuss the use of data from the low-elevation experiments to test the dual-frequency WVR algorithm described in Section 4.1. Several previous

studies have placed the accuracy of the dual-frequency WVR algorithm at between 0.5 and 1 cm (for example, Schaper *et al.* [1970]; Guiraud *et al.* [1979]; Snider *et al.* [1980]; Moran and Rosen [1981]; Elgered [1983]; Resch [1984]). None of these studies has used radio interferometry data to establish these accuracies. An important study by Resch *et al.* [1984] demonstrated that phase fluctuations of the VLA agreed well with estimates of the two-antenna wet path delay obtained from dual-frequency WVR data. There were periods, however, where for unexplained reasons these did not agree. There also seemed to be drifts in the WVR data that were probably calibration errors. In Section 4.2.*i*, we will compare estimates of the single-antenna wet propagation delay obtained from VLBI group-delay data processed with the Kalman filter to estimates of the wet delay obtained from WVR data, examining the temporal variations predicted by the two methods for time scales less than about one day. In Section 4.2.*ii*, we will discuss the variations in the differences between the two methods on seasonal time scales.

4.2.i Short-term accuracy

In Section 4.2, we discussed the Kalman-filter processor for VLBI data. In principle, the Kalman filter is able to estimate the values of a time-varying zenith delay and "clock" offset. Numerical simulations have established the ability of the Kalman filter to separate the variations of these individual contributions [T.A. Herring, Harvard-Smithsonian Center for Astrophysics, private communication, 1984]. We should therefore be able to compare estimates of the wet delay obtained from processing VLBI

data with the Kalman filter to estimates of the wet delay obtained from water-vapor radiometry.

In order to estimate the wet delay with the Kalman filter, we in effect use the dry delay as the *a priori* value of the total delay. Then, if this value for the dry delay is correct, the remaining delay, which is estimated by the Kalman filter, is the wet delay.

In Figures 4.2.1–4.2.5, we show the comparison of the estimates of the zenith wet propagation delay. The estimates of the zenith wet propagation delay obtained from WVR data are shown as squares with error bars. These error bars are the formal standard deviations from the tip curve analysis based on the sample standard deviation of the WVR output for the duration of the sampling interval. The estimates of the zenith wet propagation delay from the Kalman filter are shown as a solid connected line. Error bars have not been placed on all these estimates to avoid cluttering the figures. In general, the standard deviations of these estimates are nearly constant. Near the boundaries of data "gaps," the standard deviations tend to increase due to this lack of continuity. For observations at, or close in time to, other observations at low elevations, the standard deviations are smaller. For illustrative purposes, the error bars are shown for several Kalman-filter points on each figure.

It is obvious from these figures that the two series of estimates agree well at the level of about 2 cm, even for rapidly varying atmospheres such as that of the 23–24 September 1984 observing session. In this session, the estimates of zenith wet path delay change by approximately 5 cm in a few hours, and the Kalman filter estimates track this variation extremely well. However, there are also features which appear in

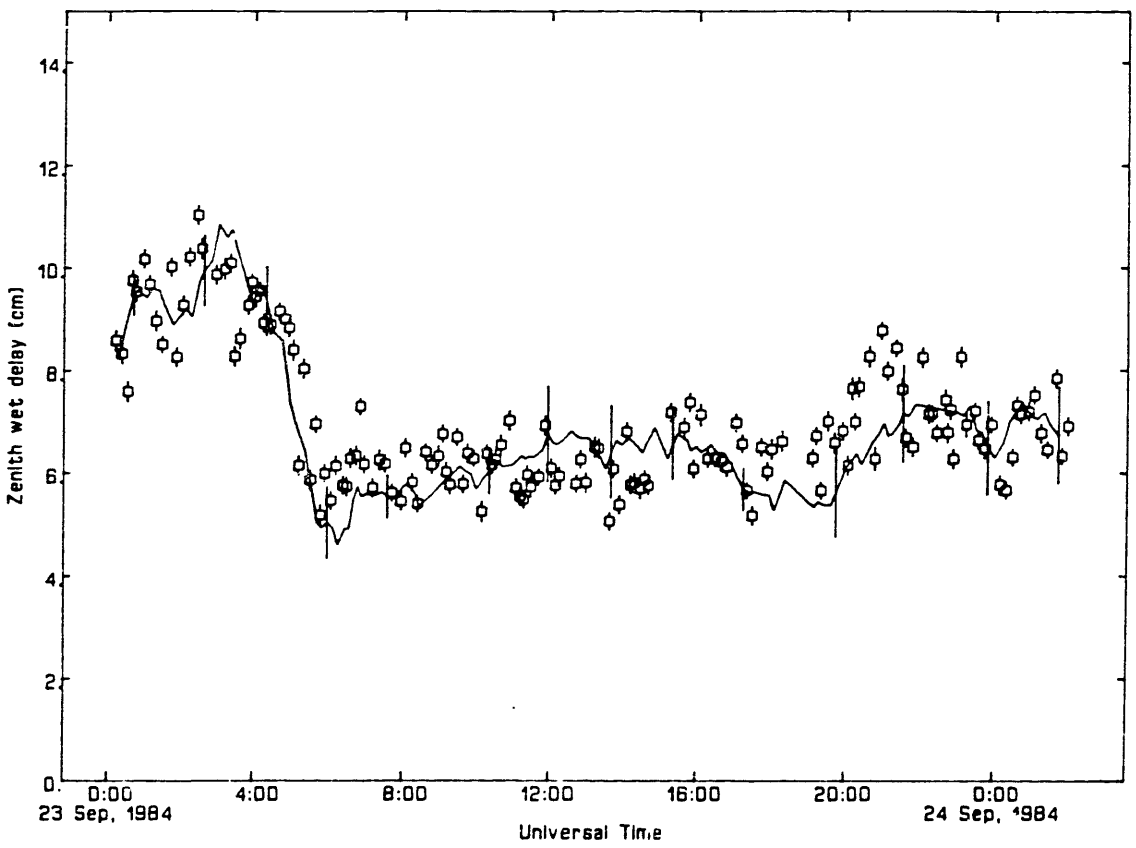


Figure 4.2.1. Comparison of estimates of zenith wet propagation delay obtained from WVR data (squares) and Kalman-filter processing of VLBI data (solid line), for "low elevation" experiment beginning 1984 September 23. For meaning of error bars, see text.

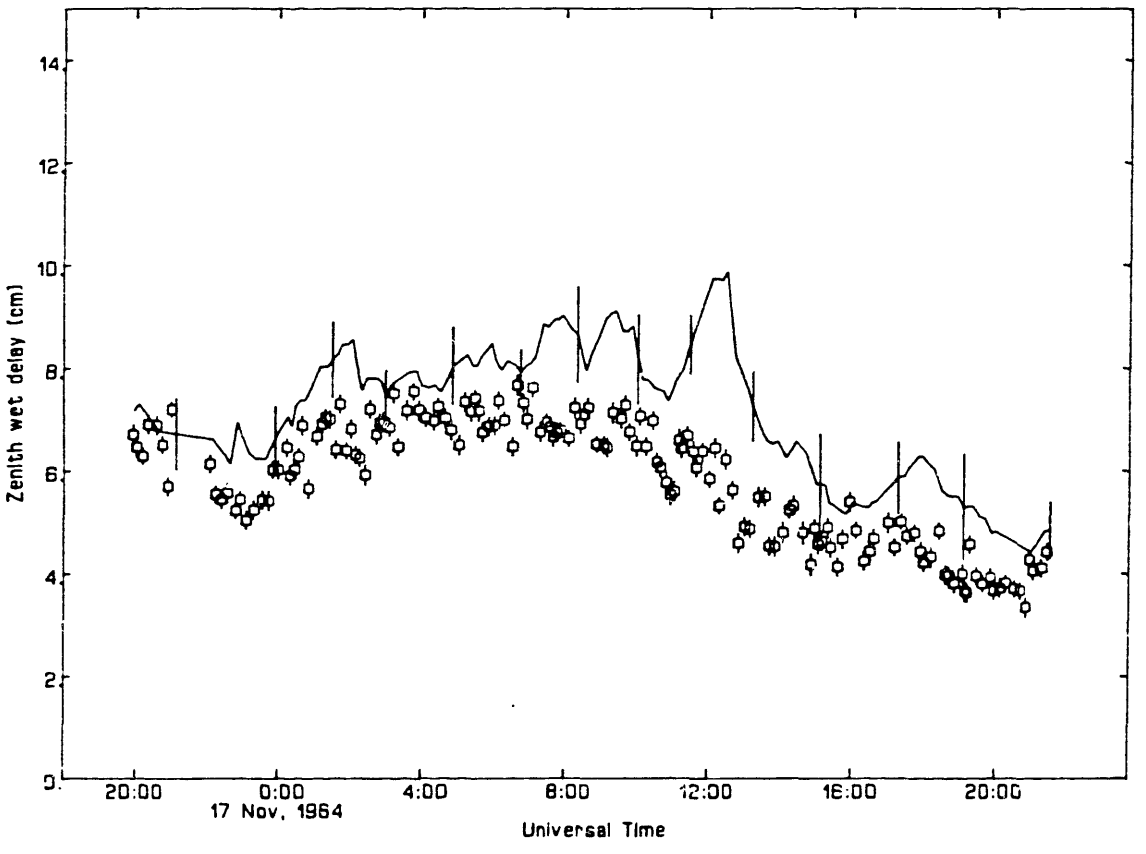


Figure 4.2.2. Comparison of estimates of zenith wet propagation delay obtained from WVR data (squares) and Kalman-filter processing of VLBI data (solid line), for "low elevation" experiment beginning 1984 November 16. For meaning of error bars, see text.

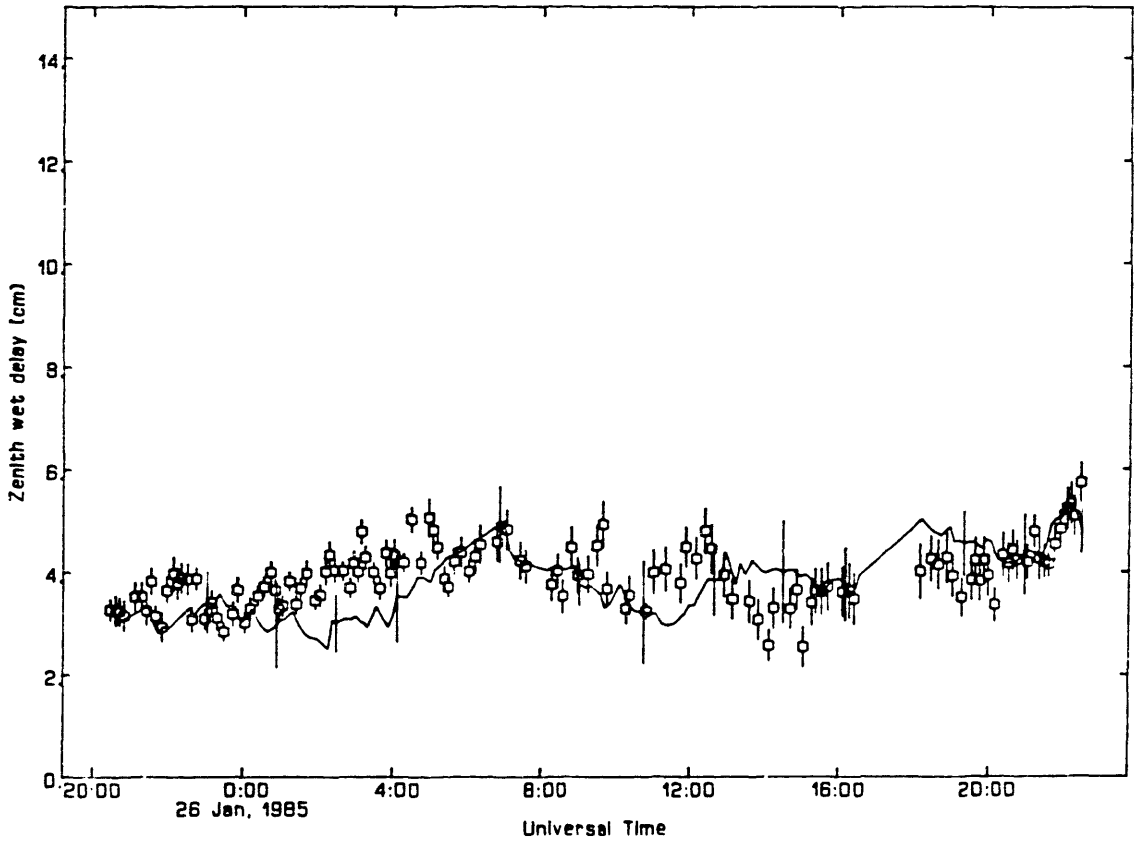


Figure 4.2.3. Comparison of estimates of zenith wet propagation delay obtained from WVR data (squares) and Kalman-filter processing of VLBI data (solid line), for "low elevation" experiment beginning 1985 January 25. For meaning of error bars, see text.

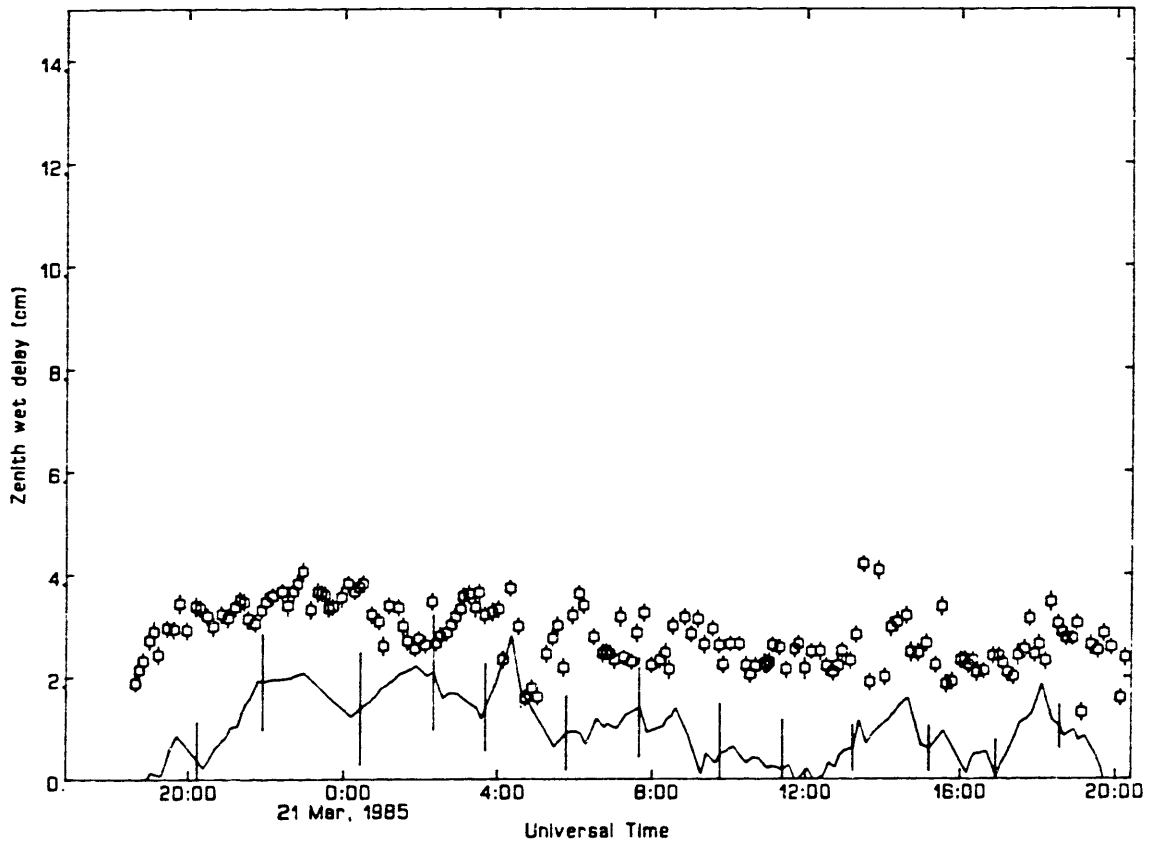


Figure 4.2.4. Comparison of estimates of zenith wet propagation delay obtained from WVR data (squares) and Kalman-filter processing of VLBI data (solid line), for "low elevation" experiment beginning 1985 March 20. For meaning of error bars, see text.

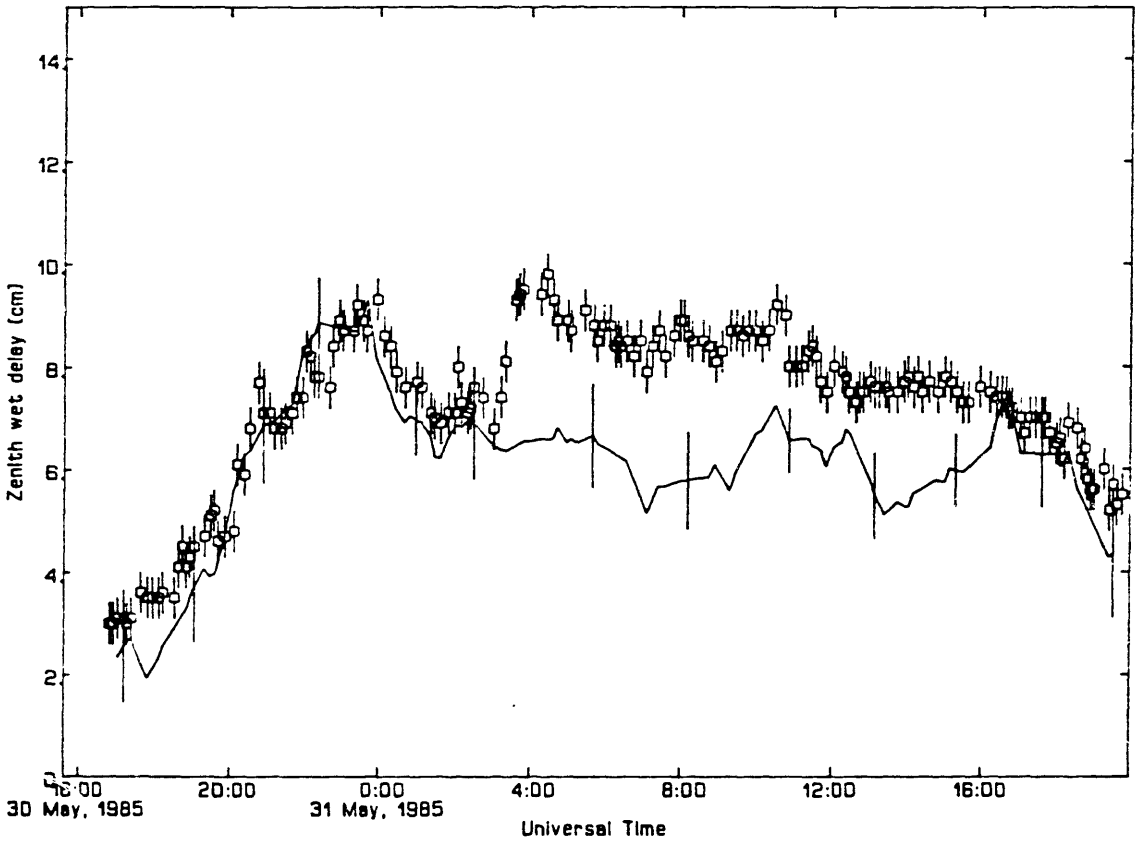


Figure 4.2.5. Comparison of estimates of zenith wet propagation delay obtained from WVR data (squares) and Kalman-filter processing of VLBI data (solid line), for "low elevation" experiment beginning 1985 May 30. For meaning of error bars, see text.

one series but not in the other, as in the "bumps" in the Kalman filter estimates for the 16–17 November 1984 observing session. Most intriguing is the comparison for the 30–31 May 1985 experiment. For this observing session, the WVR and Kalman-filter estimates agree for the first several hours while the zenith wet propagation delay undergoes large variations. Then, rather abruptly, the WVR estimates increase several centimeters while the Kalman-filter estimates remain at about the same value. The two sets of estimates remain apart but slowly converge near the end of the observing session.

What could cause differences between these independent estimates of the wet delay? In Section 2.6, we discussed sources of error in the estimates of the wet propagation delay from WVR data. These sources were e. ors in (*i*) T_{bg}, (*ii*) T_{eff}, (*iii*) τ_{O_2}, and (*iv*) $\mathcal{W}$. We shall for the moment discount (*i*) as being capable of producing the differences we observe in Figures 4.2.1–4.2.5, since from the discussion in Section 2.6 we expect these errors to be much smaller than observed. Errors in the effective temperature can also be eliminated, since from the discussion in Section 2.6 we would expect these errors to be larger for days on which the wet propagation delay was larger. In fact, we can observe some of the greatest differences on days with very small wet delays.

The two remaining candidates deserve attention. We have already pointed out that the size of the "oxygen correction" in our particular wet-delay algorithm is about half the size of the theoretical value for this correction. This error would amount to about 3 mm in the zenith wet propagation delay. However, this error would be roughly constant over one day, due to the weak dependence on the surface temperature and

surface pressure expected for the oxygen correction: see (2.6.8). This error would also be independent of the size of the delay, and for small delays could therefore account for a sizeable fractional error.

We also expect to incur errors in the value for the weighting function. As we mentioned earlier in this chapter, a constant value for the weighting function was used. However, we have seen that we may expect seasonal and site dependence of the value for the dual-frequency weighting function. We will examine this possibility further in the following section.

Another possible source of error in the estimates of the wet propagation delay from WVR data is a calibration error in one or both channels of the WVR. This source of error was not considered in Section 2.6, but had been previously discussed in Sections 2.4 and 2.5. Calibration used to obtain the estimates of the wet delay for these experiments was obtained from tip curves. Nothing in the tip-curve data suggests that any large calibration errors may have been incurred. However, we continue to examine these data in an attempt to discern some evidence for the cause of these errors.

Errors in the estimates of the wet-delay from the VLBI data processed with the Kalman filter would of course also contribute to the observed differences. One possible source of error in these estimates is an error in the dry propagation delay, since the estimation technique cannot separate the individual contributions to the propagation delay. (In principle this separation is possible if the wet and dry propagation delays possess different mapping functions. In practice, these mapping functions have so similar a dependence on elevation angle as to make this separation impossible.) Thus, estimates of the wet delay also reflect corrections to the dry delay. The error in the

dry delay could have two contributions: an error in (*i*) the zenith dry delay, and (*ii*) the dry mapping function.

In Appendix A, we discuss the precision of the zenith dry propagation delay. In Equation (A14) we find that the uncertainty in the zenith dry delay is approximately $(0.0005 \text{ mm mbar}^{-1})P_\circ$, where $P_\circ$ is the surface pressure. This uncertainty takes into account the uncertainties in the dry refractivity constant, the mean acceleration of gravity, the universal gas constant, and the variability of the mean molar mass of dry air. The formula for the zenith dry delay and this uncertainty were derived, however, under the assumption of hydrostatic equilibrium. An extreme upper limit for the effect of nonequilibrium conditions is about 20 mm/1000 mbar, or about 20 mm at a pressure of 1 atm. Certainly, errors in the dry zenith delay of this size could account for the observed differences. Furthermore, we might expect that errors due to this source would vary on time scales shorter than one day, since nonequilibrium conditions may be caused by storms or possibly passages of fronts. We hope to study the effects of nonequilibrium conditions in the future, in order to obtain a more accurate estimate of these effects.

The possibility that the differences are due to errors in the dry mapping function is attractive, for it gives us a method for detecting such errors. If the errors were due to this source, we would expect that differences are elevation-dependent. Unfortunately, we can find no evidence of such dependence, possibly because the Mojave antenna does not observe below 10° elevation. Perhaps we will obtain more evidence when the Haystack site receives a WVR.

In the next section, we return to the idea that the differences may be due to errors in the value for the weighting function used.

4.2.ii Seasonal variations

As mentioned in Section 4.1, when WVR data were available, only a single atmospheric parameter was estimated for Mojave. This estimation was to allow for the fact that the WVR algorithm may be biased. In the previous section, we indeed saw that the estimates of the zenith wet delay from the Kalman filter were often offset from the estimates derived from WVR data. Figure 4.2.6 shows the estimated bias parameters for the five observing sessions with WVR data. These bias parameters can be roughly described by an offset and a sinusoid with a period of one year. An offset indicates a constant error in the value for the weighting function used. This error might be due to the site-dependence of the weighting function, for example. The annual variation can be understood if we examine Figure 2.6.2. This figure shows the annual variation in the value for the dual-frequency weighting function for the radiosonde launch site at Landvetter airport near Göteborg, Sweden. From this study, we may assume that the value for the weighting function at the Mojave site also undergoes some annual variation, and this assumption is borne out by the results of the low-elevation experiments. The bias parameters of Figure 4.2.6 are "in phase" with the sinusoid evident in Figure 2.6.2, as we might expect. However, the amplitudes of the variations are different. In Figure 4.2.7 we have plotted the corrections to the weighting function implied by the corrections in Figure 4.2.6, in terms of the original value of the weighting function. We can see that in one case the implied correction of the weighting function

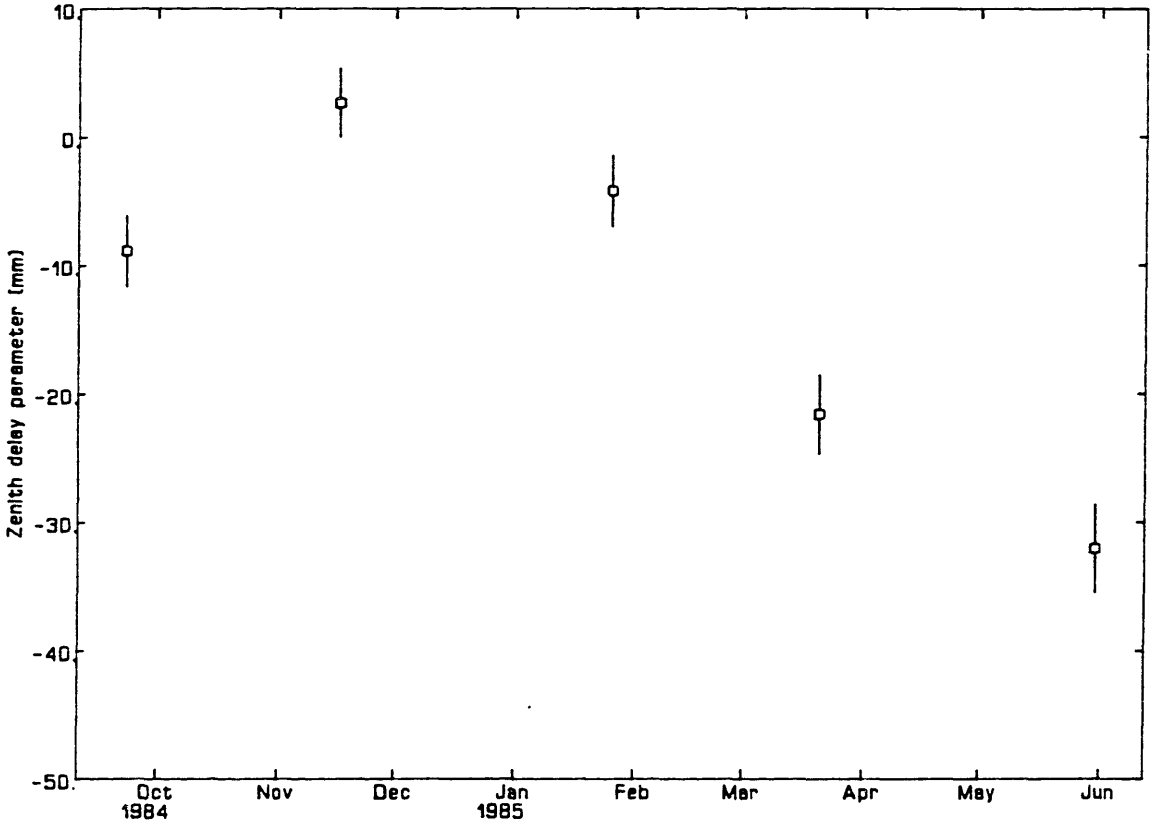


Figure 4.2.6. Estimated bias parameters for the zenith atmospheric propagation delay at the Mojave site. The error bars shown are the statistical standard deviations based on the SNR-derived group-delay uncertainties and the values for the stochastic-process variances used to represent the clock and atmosphere behaviors (see text).

has a value of ~50% of the value for the weighting function used. Errors in the original value for the weighting function of this size are difficult to understand, and it may be that some other (additive) error is contributing to the corrections in Figure 4.2.6. For small values of the zenith wet delay, this additive error would correspond to a large fraction of the total delay. Clearly, studies of the weighting function using radiosonde data for (*i.e.*, near) Mojave are needed. Simultaneously, an effort should be made to examine the small uncertainties in the dual-frequency algorithm, such as those in the oxygen opacity. Errors in this term (see Section 2.6) may contribute to the corrections in Figure 4.2.6 and 4.2.7.

The examination of the accuracy of wet-path-delay algorithms for WVR's is an ongoing research topic. The techniques for this examination presented in this and the previous sections have only recently been developed: the results presented here may be considered as intermediate results in that they indicate the direction of research to be taken, but we cannot say anything final concerning the accuracy of the WVR algorithms at the level of 2–4 cm or less. Nevertheless, these results do indicate the potential for accuracies of less than 1 cm, depending on our ability to model the variations of the weighting function. Currently, the modeling of these variations as a function of surface temperature and pressure by Elgered [Onsala Space Observatory, Sweden, private communication, 1985] has yielded encouraging if preliminary results. It will quite likely be several years before a "final" site– and season–dependent algorithm will have been tested and proven satisfactory. It is also possible that as our ability to model the variations in the weighting function improves, our ideas concerning a "satisfactory" accuracy will change.

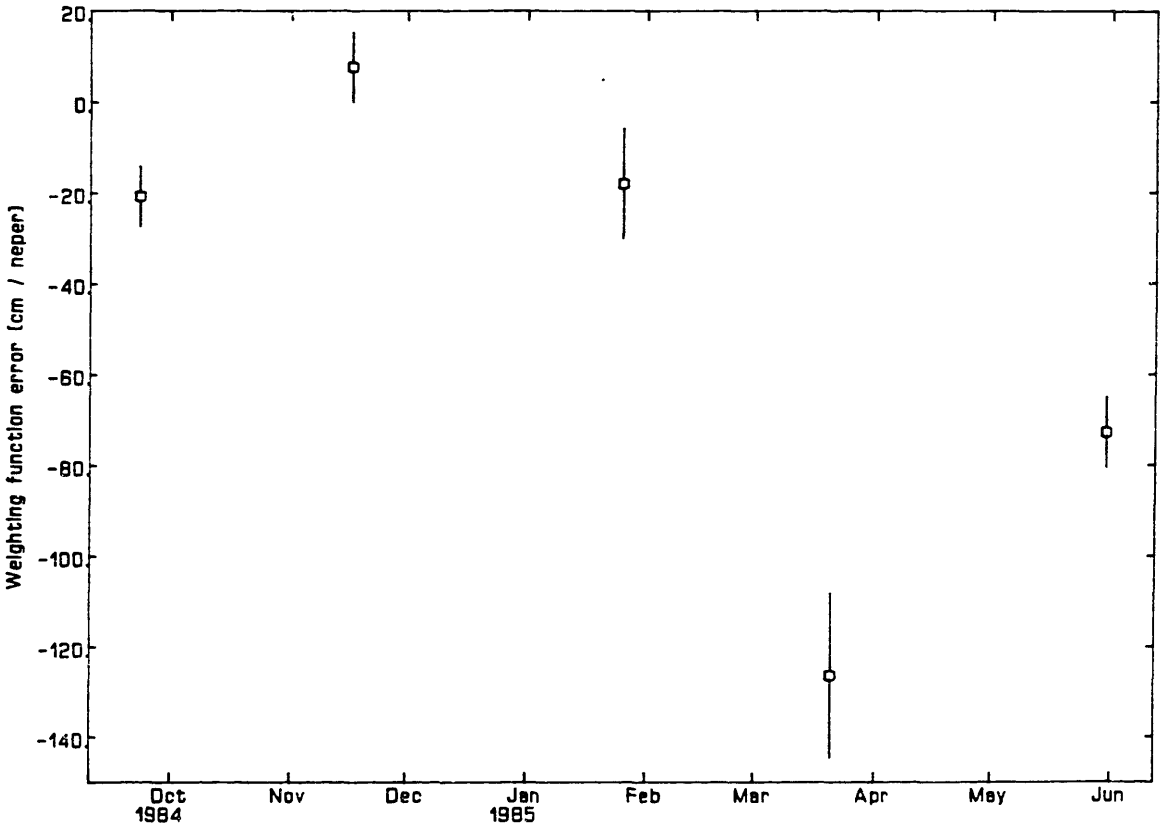


Figure 4.2.7. Errors in $\mathcal{W}^{-1}$ implied by the estimated bias parameters in Figure 4.2.6.

In the following section, we will describe the use of data from the low-elevation experiments to test the new mapping function for the dry path delay, the development of which was described in Chapter 3 and Appendix A.

4.3 Testing the mapping function

As discussed in the introduction to this chapter, the primary purpose of conducting the low-elevation experiments was, at the time of their designing, the testing of the new mapping function for the dry path delay, CfA-2.2. In this section, we will test the mapping function in two ways. First, we will perform elevation-angle cutoff tests using the CfA-2.2 mapping function, and compare the results of these tests to similar tests using the Marini mapping function. We will then discuss seasonal variations in the estimates of baseline length, which will be interpreted as seasonal errors in the mapping function.

4.3.i Elevation-angle cutoff tests

Figure 4.3.1 contains comparisons of the results from the elevation-angle-cutoff tests for the CfA-2.2 and Marini mapping functions, for the eight low-elevation experiments. Shown are the differences $\hat{b}_\epsilon - \hat{b}_0$, where $\hat{b}_0$ is the estimate of the Mojave–Haystack baseline length from all the data, and $\hat{b}_\epsilon$, $\epsilon = 5°, 7.5°, 10°, 15°, 20°, 25°, 30°$, is the estimate of that baseline length using data from above the elevation ϵ. The differences obtained using the CfA-2.2 mapping function for the dry delay are indicated by a square, those using the Marini mapping function by a triangle. The error bars are the statistical standard deviations of the differences based on the statistical standard

deviations of the estimates of baseline length, obtained using the SNR-derived group delay uncertainties, and depending on the assumed standard deviations of the stochastic processes representing the clock and atmosphere variations. These error bars are shown only for the CfA-2.2 points, as they are the same for the Marini points.

We will first make a few general statements concerning the results of the elevation-angle-cutoff tests using the CfA-2.2 mapping function. If this mapping function described the mapping of the dry delay with no error, then we would expect the differences shown in Figure 4.3.1 to display no systematic dependence on elevation-angle cutoff, or on date of experiment, and to have a statistical expectation of zero. If we limit our discussion only to the differences using cutoffs of 5°, 7.5° and 10°, we could probably say with confidence that the previous statement describes the results: the differences are all within 1–2 standard deviations of zero. Since, from the discussion in Chapter 1, these differences (*i.e.*, from the lower elevation-angle minima) are most greatly affected by errors in the mapping function, the results in Figure 4.3.1 seem to imply that the CfA-2.2 mapping function induces "negligible" errors for minimum elevation angles below about 10°, for these experiments. We will discuss this conclusion further later in this and the following section.

For elevation-angle minima above 15–20°, an elevation-angle cutoff test practically gives no information about the CfA-2.2 mapping function. Since above these elevation angles this mapping function is nearly equal to the cosecant of elevation, and, based on ray-tracing studies, this law describes the dry delay quite accurately for these elevation angles, we can assume that the error in the estimates of baseline length attributable to mapping function error is very nearly zero. The differences

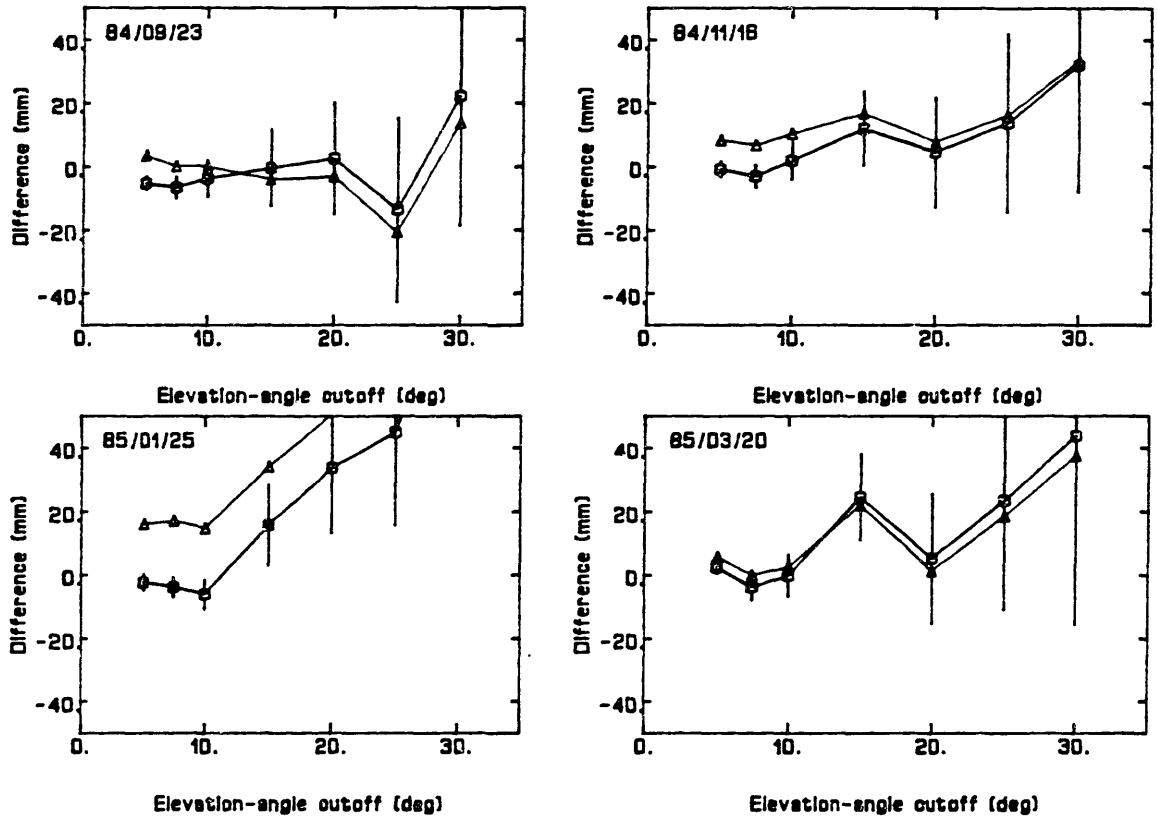


Figure 4.3.1. Elevation-angle-cutoff tests for the low-elevation experiments, for the CfA–2.2 (squares) and Marini (triangles) mapping functions. Shown are the differences of the estimates of the Mojave–Haystack baseline length obtained using all the data from the estimate of that length obtained using data from elevation angles greater than the indicated angle. The error bars are the statistical standard deviation of the differences, based on the standard deviations of the original estimates of baseline length (see Appendix A). These error bars have been placed only on the CfA–2.2 differences, but are the same as the Marini differences for the corresponding minimum elevation angle.

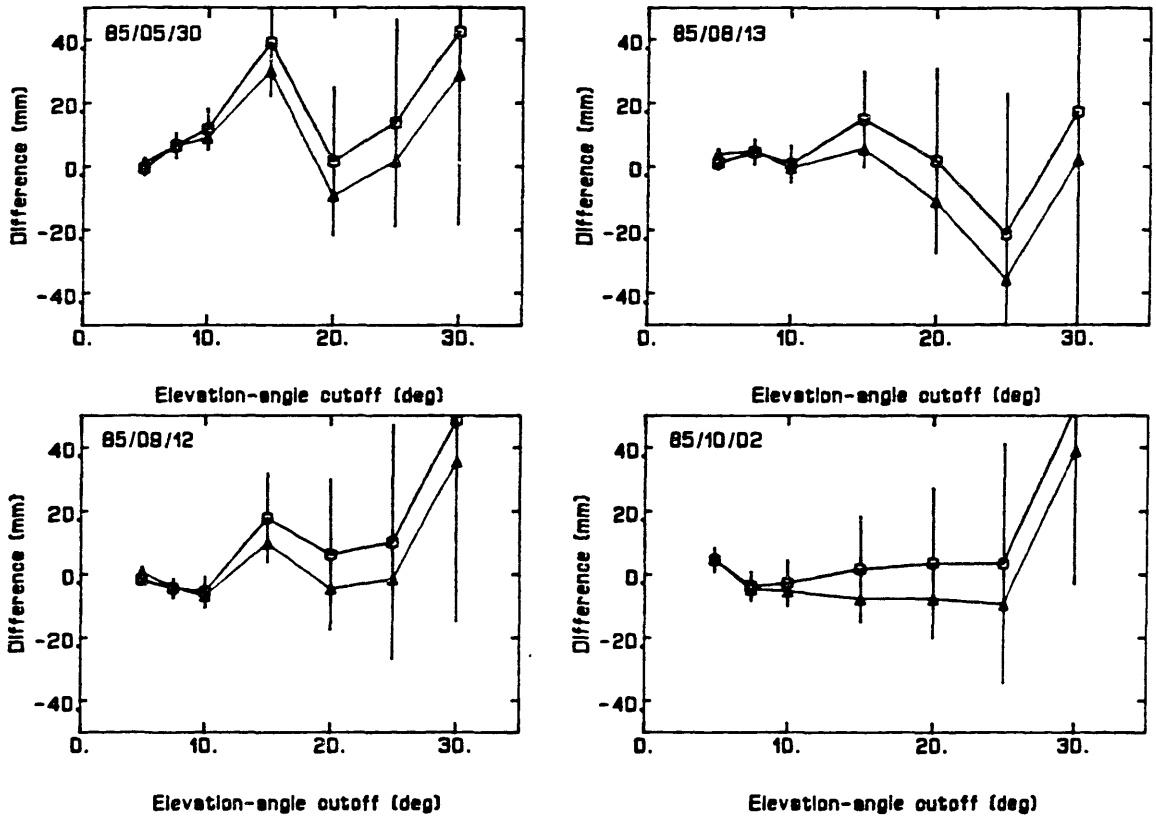


Figure 4.3.1. (continued)

shown in Figure 4.3.1 for above about 15° should be constant, and that constant value represents the difference between the "true" baseline length, and the baseline length estimated using all the data. Of course, as we discussed in Chapter 1, the random errors in the estimate of baseline length become very large as the elevation-angle minimum increases, and so the differences in Figure 4.3.1 above the 15° point are not constant, and, unfortunately, we cannot discern the "true" baseline length with any precision.

An interesting phenomenon which seems to be present in the results in Figure 4.3.1 is this systematic positive difference for the $\hat{b}_{15} - \hat{b}_0$ points. For the majority of the experiments, this point seems to "pop up" from the surrounding points. A possible explanation for this behavior may lie in the fact that the Mojave antenna cannot observe below 10° elevation. Thus, only low elevation-angle observations at Haystack are deleted when raising the elevation-angle minimum to 10°. There are a large number of observations between 10° and 15° elevation at Mojave, however, and deletion of these points may affect the results shown in Figure 4.3.1 if there is a mapping function error at Mojave. Thus, this "bump" may be indicative of a site-dependent variation not accounted for in the implementation of the CfA-2.2 mapping function. However, we will need to perform more low-elevation experiments if such a result is to be believed. If we continue to obtain evidence pointing in this direction, then we could attempt to "climatize" the mapping function. A first step would be to use radiosonde or other data to determine site- and season- dependent values for the temperature lapse rate and tropopause height, instead of using the nominal values

(see Section 4.1.v). Other sites should also participate in these or other low-elevation experiments to test this hypothesis on a larger sample of climates.

If we examine Figure 4.3.1 for the results for the Marini mapping function, we first notice that the Marini differences seem to "parallel" the CfA-2.2 differences, especially for the higher minima. This is due to a fact which we have already discussed. Since the Marini and CfA-2.2 mapping functions agree above the elevation of $\sim 15°$, the estimates of baseline length made with the two different mapping functions will agree if the minimum elevation angle is above this elevation. (Note that there is no random effects involved: the estimates will be equal if the model is the same.) The offsets seen in Figure 4.3.1 between the CfA-2.2 and Marini curves for elevation-angle minima above 15° are due only to the differences between the estimates $\hat{b}_0$ for the different mapping functions, and are included to indicate that difference.

In fact, the curves for the CfA-2.2 and Marini mapping functions agree quite well for the lower elevation-angle minima as well. It is only for the first three experiments that there is evidence of nonzero differences. The largest differences from zero can be seen in the results for the experiment dated January 25, 1985. For this experiment, the estimates of the baseline length obtained using elevation-angle minima of 5°, 7.5°, and 10° are all approximately 20 mm greater than the estimate of the baseline length using all the data. These differences indicate that the Marini mapping function exhibited severe errors below an elevation angle of 5°. Note that no such behavior is evident from the CfA-2.2 mapping function. The behaviors for both mapping functions for minimum elevation angles above 10°, with only one experiment from this time of year,

are difficult to interpret. Again, we must wait for more experiments to be conducted before we can decide on the meaning of such results.

Throughout this section, we have been comparing *differences* of estimates of baseline length. In the following section, we will compare the actual estimates of baseline length, and attempt to interpret the results in light of the comparisons of this section.

4.3.ii Seasonal variations

In Figure 4.3.2, we again present the "standard results" for the series of low-elevation experiments conducted before November, 1985. As described in Section 4.1, we used the CfA-2.2 mapping function to map the dry delay for the solutions from which the data in Figure 4.3.2 are generated. Since we are attempting to determine the effect of errors in this mapping function, we would like to know the errors in the estimates shown in Figure 4.3.2. Unfortunately, we have no estimates of this length of comparable precision with which to assess the accuracy of our estimates.

Figure 4.3.2 shows estimates of the Mojave–Haystack baseline length obtained using a processing identical to that for the standard results, except that the Marini mapping function has been used for the dry delay. (We also show the standard results in this figure.) Since we do not know the true values of the baseline length (which may be changing with time), we cannot say for sure that these estimates are any "better" or any "worse" than the standard results. However, we make the following observations:

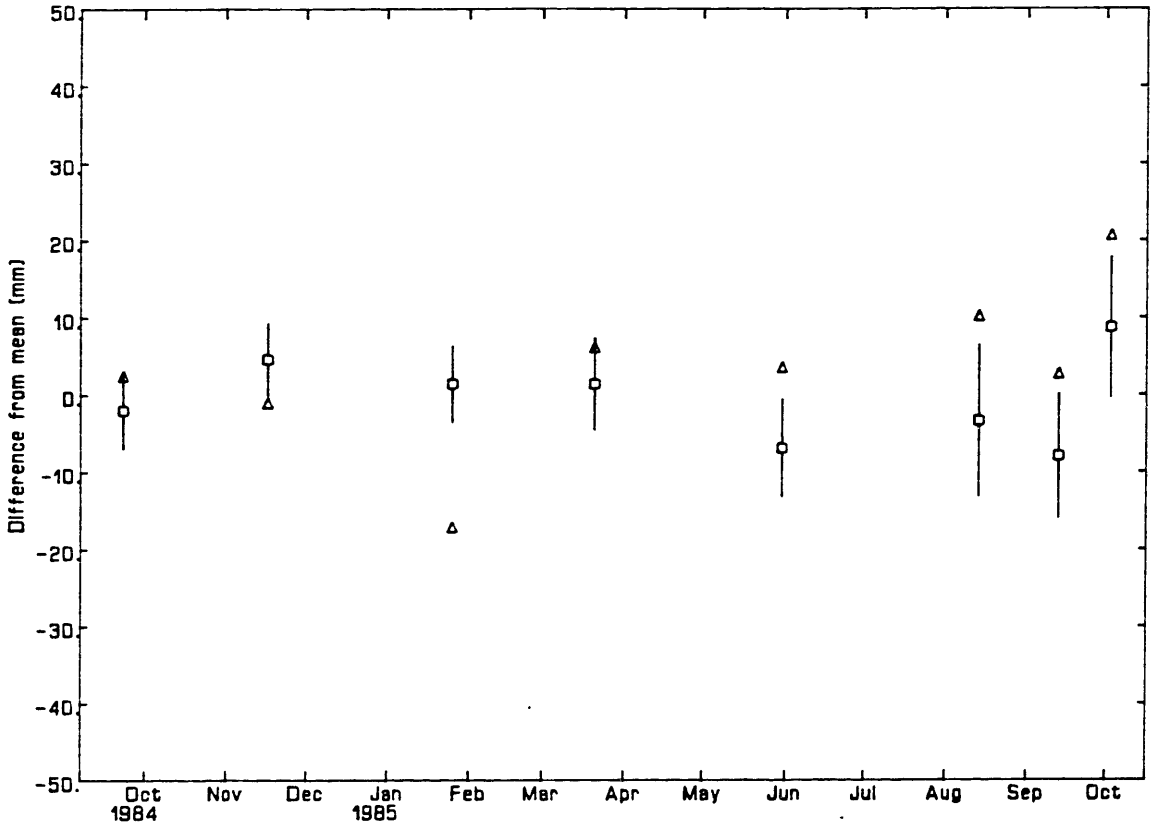


Figure 4.3.2. Standard results from the low-elevation experiments (squares). Also shown are the "standard Marini" results (triangles; see text).

- The weighted mean of the estimates of the Mojave–Haystack baseline length made using the Marini mapping function is 3,904,144,247.3 mm, while the standard results have a mean of 3,904,144,246.9 mm. The difference between these estimates is only 0.4 mm.
- The weighted root-mean-square (WRMS) deviation of the Marini results about their weighted mean is 9.6 mm, for a χ^2 per degree of freedom of 2.5. The standard results have a WRMS deviation of 4.5 mm and a χ^2 per degree of freedom of 0.6.

Can we interpret the above observations in terms of the accuracy of our standard results and therefore, by extension, in terms of the accuracy of the CfA-2.2 mapping function? Our inclination is to say that based on the RMS about the weighted means alone, the reproducibility of the standard results implies that these results are far more accurate than the Marini results. Indeed, in terms of the "fractional stability" (*i.e.*, the ratio of the WRMS deviation to the baseline length), the standard results are among the most stable estimates for long (> 1000 km) baselines ever reported. The small reduced χ^2 may be due to the overestimate of the variability of the stochastic processes describing the behavior of the clocks and atmospheres (see Section 4.4.*i*). On the other hand, the probability that the χ^2 per degree of freedom is less than 0.6 for seven degrees of freedom is about 0.25. Thus, the small WRMS alone is probably not a good basis for the judgement of accuracy (although this line of reasoning assumes

that the errors in the estimates of baseline length are statistical in nature, *i.e.*, that there are no significant systematic errors due to, say, mapping function errors).

Nevertheless, it is still true that the estimates using the Marini mapping function exhibit more scatter than the standard results. There is some evidence that the variation in the Marini estimates depends on the weather. In Figure 4.3.3, we plot the difference between the Marini and standard results plotted as a function of average temperature at Haystack for the respective experiment. The error bars shown are the standard deviation for the baseline length estimates, which are independent of mapping function, and do not represent statistical uncertainty in the differences. (Since the estimates are obtained using identical data, but different models, the concept of statistical uncertainty does not apply here.) There is a hint of correlation in this figure, due mostly to the difference between the estimates for the January 25, 1985 experiment. The difference of almost 20 mm from the mean for the estimate for this experiment is alone responsible for a large fraction of the WRMS for the Marini results. Thus, we might base a conclusion of temperature-dependent errors in the Marini estimates upon this single point, except that we expect such errors based on the expression for the mapping functions. Figure 4.3.4 shows the propagation delay at 5° elevation for the Marini mapping function and for the CfA-2.2 mapping function as a function of surface temperature. The meteorological parameters used for this figure are $P_o = 1013.25$ mbar, $e_o = 0$ mbar, $\beta = -6.5$ K km^{-1}, $h_t = 11.231$ km. For the site position parameters, we used values representative for the Haystack antenna of $H = 0.1$ km and $\phi = 43°$. (See (1.4.1) and (1.4.2) for the zenith delay formulas, (1.4.3)–(1.4.5) for the Marini mapping function, and (9)–(12) of Appendix A for the CfA-2.2 mapping function.)

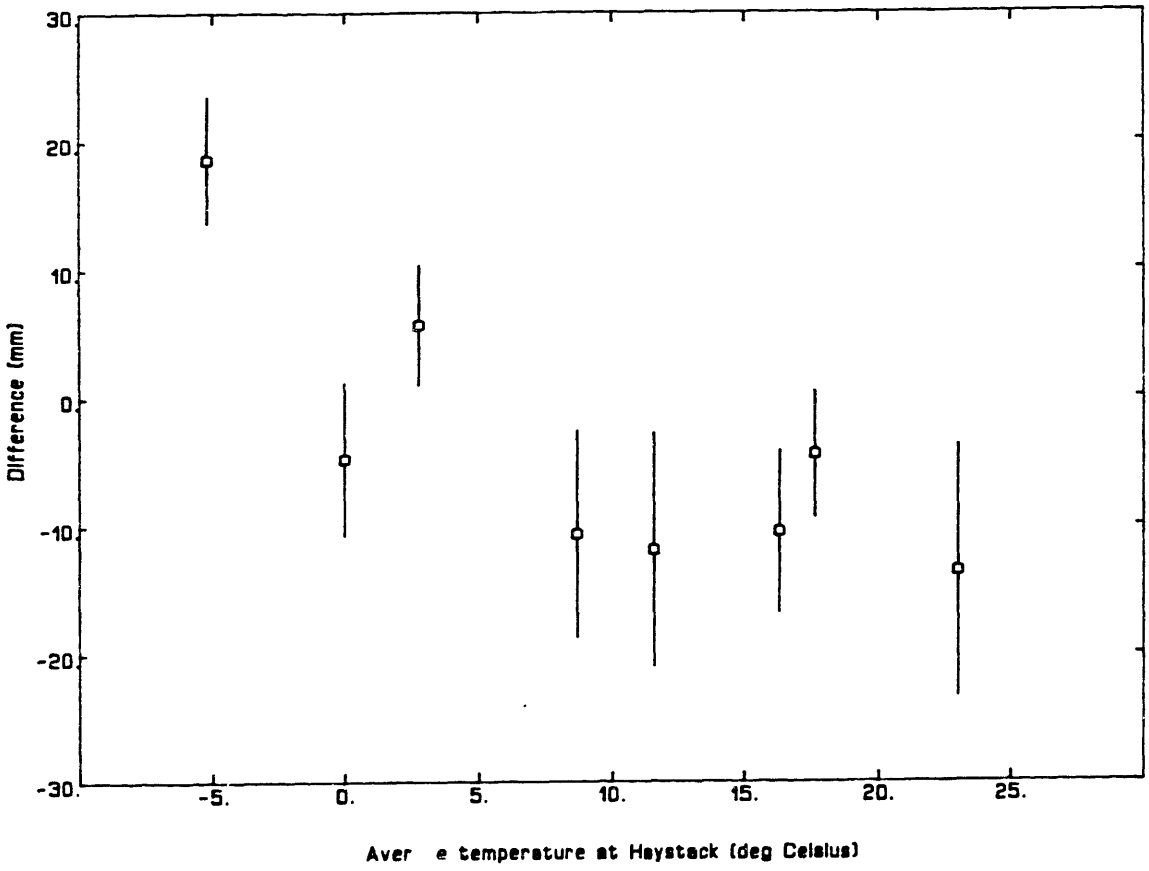


Figure 4.3.3. Differences of the baseline-length estimates from the "standard" and "Marini standard" results, as a function of average surface temperature at Haystack. The error bars shown are the standard deviations for the baseline-length estimates, which are independent of mapping function, and do not represent statistical uncertainty in the differences (see text).

We can see that the propagation delay predicted by the Marini mapping function has no dependence upon temperature, while the delay predicted by the CfA-2.2 mapping function has a strong temperature dependence. The CfA-2.2 mapping function, in fact, agrees with ray-trace predictions of the temperature-dependence of the propagation delay, since it was derived from ray-trace calculations (see Chapter 3).

The above discussion leads us to conclude that at least some of the differences between the estimates of baseline length obtained from the CfA-2.2 and Marini mapping functions are due to the improper temperature dependence of the Marini mapping function. Moreover, the stability of the CfA-2.2 estimates under varying temperature conditions might lead us to believe that the the differences in these estimates are due mainly to errors in the Marini estimates. This line of argument does not prove that the CfA-2.2 mapping function yields unbiased estimates, because it assumes this result based upon the small scatter of the CfA-2.2 results. The most that we can say at this time is that the conclusions drawn above are consistent with (*i*) the CfA-2.2 yielding an unbiased estimate of the baseline length, and (*ii*) the known erroneous temperature dependence of the Marini mapping function yielding biased estimates of the baseline length.

The program of monthly low-elevation experiments continues as of this writing. From September 1984 through August 1985, the experiments were conducted bi-monthly at best. Beginning in August 1985, the experiments have been conducted monthly. They will continue to be conducted monthly at least through to the end of 1986 (with a gap for the mobile VLBI summer campaigns). It is not clear that after this time we will be able to say anything more definite concerning the CfA-2.2

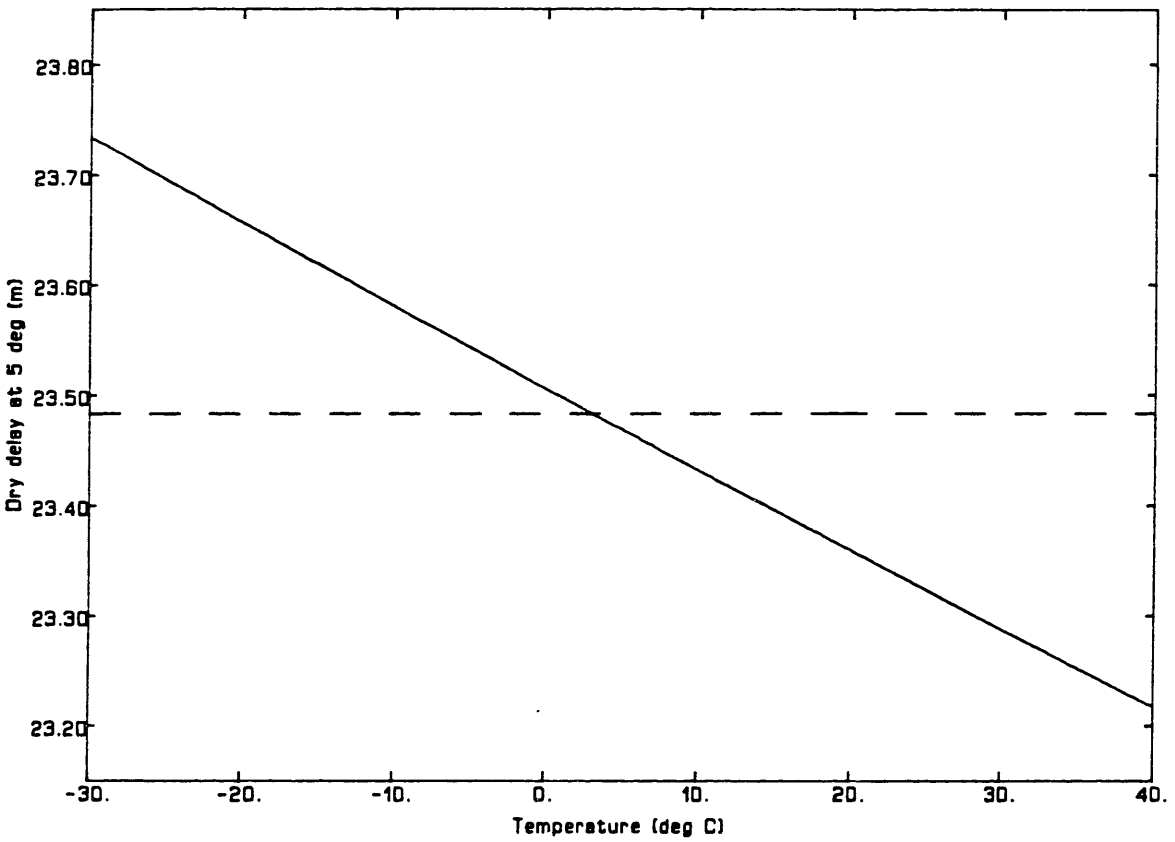


Figure 4.3.4. Dry delay at an elevation of 5° predicted by the CfA-2.2 mapping function (solid line) and Marini mapping function (dashed line). For values of parameters used in the formulas, see text.

mapping function, since we will still have a relatively small number of experiments. In a sense, it is unfortunate (and certainly unexpected) that the standard results exhibit so little scatter. The purpose of the low-elevation experiments is to test and improve our mapping function; by observing at low elevation angles, the errors in the CfA-2.2 mapping function should "stand out." Yet the standard results give no reason to believe that the CfA-2.2 mapping function induces large errors in the estimates of baseline length. Nevertheless, we do expect observable errors which are both site– and season–dependent (see Appendix A). One possible improvement in the current "standard" processing technique may occur with the advent of a WVR at Haystack, and with improved site– and season–dependent water-vapor weighting functions. Then the necessity of estimating the zenith delay will be eliminated, and another test of systematic error will be available: the zenith delay/ no zenith delay comparison (see Chapter 3).

In the final section of this chapter, we again discuss the treatment of the propagation delay during data analysis.

4.4 Treatment of the atmosphere in VLBI data analysis. III

In this section, we will again discuss the treatment of the propagation delay in VLBI data analysis, this time in view of the results from the previous sections in this chapter. We will assume that these results imply that (*i*) the dual-frequency weighting function can be modeled and the "oxygen correction" determined so that the estimate of wet path delay is unbiased, and (*ii*) the mapping function for the dry path delay can be determined such that the error in the dry delay is "negligible." Although these

conditions do not follow directly from the results in the previous sections, these results do imply that the conditions are possible to attain. However, we will first discuss the Kalman filter processor in more detail than in Section 4.1.*v*.

4.4.i The Kalman filter

A difference from the previous sections dealing with the treatment of the atmosphere in VLBI data analysis will be our use of a Kalman filter for processing VLBI data. The Kalman filter was briefly introduced in this chapter, and estimates using this processor presented. It is still not necessary in this section to present the Kalman filter equations [Liebelt, 1967], but we will outline the theory of the Kalman filter as a sequential least squares estimator.

We first review least-squares estimation with *a priori* information. Suppose that $\hat{\mathbf{x}}$ is our "*a priori* estimate" of the parameter vector $\mathbf{x}$, with "*a priori* covariance matrix" P_x. We wish to "update" our values of the parameters using data. This "update" either represents a "better" estimate of an unchanging set of parameters or the new values of a changing set of parameters, or a combination. If the parameters or some subset are changing in a random way, then the vector $\hat{\mathbf{x}}$ can be interpreted in a Bayesian sense as being the expected value of $\mathbf{x}$, with that statistic having the covariance matrix P_x. The data are assumed to be related to the parameter vector by

$$\mathbf{y} = A\mathbf{x} + \vec{\epsilon} \tag{4.4.1}$$

where $\vec{\epsilon}$ represents experimental error, and A is a matrix of partial derivatives. The "updated" estimate of $\mathbf{x}$ can be written

$$\hat{\mathbf{x}}' = P_x' A^T G_y^{-1} (\mathbf{y} - A\hat{\mathbf{x}}) \tag{4.4.2}$$

where $G_y = \langle \vec{\varepsilon}\vec{\varepsilon}^T \rangle$, and the covariance matrix P_x' of $\hat{\mathbf{x}}'$ is given by

$$P_x' = (P_x^{-1} + A^T G_y^{-1} A)^{-1} \tag{4.4.3}$$

One observation which we wish to make for later referral is the dependence of $\hat{\mathbf{x}}$ in (4.4.2) on the difference between the observed value and the prediction for that observed value based on the *a priori* parameter values. This difference has two contributions: experimental error and the error in the *a priori* estimates. It is assumed that both of these errors are zero mean and Gaussian and that their covariance matrices are given by G_y (experimental error) and P_x (error in *a priori* estimates).

We are now in a position to review the estimation of parameters by a Kalman filter. The first step for a Kalman filter is to take the "best estimates" for a set of parameters—some of which are stochastic, and some of which are deterministic—at the time t_{i-1}, and obtain the expected values of those parameters at time t_i. Obviously, if the parameters do not change with time, the expected values at time t_i equal the values at time t_{i-1}. For parameters, such as those that represent the clocks and atmospheres, which change randomly with time, the expected values at time t_i are given by the expectation of (4.1.1). In taking the expectation of this expression, we use $\langle \Delta\phi(t_i) \rangle = \langle \Delta\omega(t_i) \rangle = 0$. (The random variations are assumed to be zero mean Gaussian, as mentioned previously.) Thus, the expected values of the parameters for the clocks and atmospheres are determined simply by *projecting* those values forward in time. For example, suppose that at some time, the "offset" value for a particular clock is 10 psec, and the "rate" value is 0.1 psec minute^{-1}. Then, using

our particular model given in (4.1.1), the "expected" value for the offset 20 minutes later is $10 + 0.1 \times 20 = 12$ psec. (The "rate" in this model remains the same.)

What is the covariance matrix of these "projected" estimates. In determining this matrix we must take into account (*i*) the covariance matrix of the estimates at the previous time, and (*ii*) the covariances due to the unknown random variation during the time over which we projected the previous estimates. That is, because we assumed that the variations were zero when we projected the stochastic parameters forward, the values of those parameters are more uncertain. It is quite easy to show that the covariance matrix at time t_i is simply the sum of the covariance matrix at time t_{i-1} and the covariance of the stochastic parameters, which for our particular model for those parameters is given by (4.1.2).

We now perform a constrained least-squares solution for the "updated" values of the parameters at time t_i. The data at time t_i form our data vector, the "projected" parameter estimates form our "*a priori* estimates," and their covariance matrix forms the "*a priori* covariance matrix." A Kalman filter does not use (4.4.2) and (4.4.3) to make these estimates, but instead uses a very efficient algorithm with a much smaller number of matrix inversions.

4.4.ii Estimating atmospheric parameters

Suppose that for a given experiment, the estimate of the line-of-sight wet path delay—inferred from WVR data—for the VLBI observation at time t_i is $\hat{L}_w(t_i)$. The true wet path delay for this observation is $L_w(t_i)$ and the error in the estimate is

$$\Delta L_w(t_i) = \hat{L}_w(t_i) - L_w(t_i) \tag{4.4.4}$$

If the $\Delta L_w(t_i)$ are zero mean, Gaussian, with corresponding standard deviation σ_i (which represents WVR measurement error), how would we utilize our information? Within the framework of a Kalman-filter processor, we effect this utilization in a manner consistent with the assumptions from which the Kalman filter equations are derived: stochastic modeling of the wet path delay is implemented, and the observations at each epoch t_i include the "observation" of the wet path delay with value $\hat{L}_w(t_i)$ and "experimental uncertainty" σ_i.

Now suppose that in our study and evaluation of a season-dependent model for the dual-frequency weighting function, we find that we cannot model the weighting function perfectly, so that on any given day our estimate of the weighting function has some error which for illustrative purposes we will assume as being constant for the observing session. Then one method which takes this error into account would be to use as the wet delay "observation equation"

$$L_w^{obs}(t_i) = [1 + \Lambda]\hat{L}_w^{\circ}(t_i) + \varepsilon_i \tag{4.4.5}$$

where $\hat{L}_w^{\circ}(t_i)$ is the estimate of the wet delay obtained from WVR data using the nominal value for the weighting function; ε_i is the experimental error from the WVR, and is zero mean with standard deviation σ_i; and Λ is a parameter to be estimated. (We have omitted a term which is the product of Λ and ε_i.) Note that for $\Lambda = 0$, we have $\Delta L_w(t_i) = \varepsilon_i$, where $\Delta L_w(t_i)$ was defined above. For consistency with our Bayesian approach, an "observation" giving zero for the value of Λ, and using the standard deviation for Λ determined from the radiosonde studies should be included. If the *a priori* value of Λ is not zero (if there are remaining seasonal effects, for instance), then

this value should be used instead of zero in the "observation," because, as we saw in the previous section, the *a priori* uncertainty used refers to the standard deviation of the difference between the *a priori* value and the true value. If the *a priori* uncertainty does not reflect this difference, then the expectation of the "updated" estimates will not necessarily be the true values of the parameters, and the corresponding covariance matrix will not necessarily reflect accurately the statistical differences between the estimates and the true values.

It should be clear that any number of potential errors can be treated—at least formally—by interpreting the errors in the parameters within a general Bayesian framework in which all *a priori* information can be treated as observations. The usefulness of this approach depends on the "precision" of the *a priori* information. Poor "precision" of the *a priori* information can be equivalent to no *a priori* information at all; the ability to estimate the parameters then becomes simply a matter of "separating" the parameters in the familiar least-squares sense. For example, we have seen that with no *a priori* information, the estimate of a single zenith delay parameter is highly correlated with the estimate of the vertical coordinate of site position (see Section 1.5). This increased correlation causes the statistical standard deviation of this coordinate of site position to inflate. It is therefore important to use all the information available to attempt to decrease correlations among the parameters. At the same time, it is important that no systematic errors be introduced, so that the statistics derived from the "data" uncertainties (*i.e.*, the group-delay uncertainties as well as the "pseudo-observation" uncertainties) truly reflect the errors in the parameter estimates.

Conclusions

Errors in the formulas for the atmospheric propagation delay used in conjunction with the analysis of VLBI data will induce errors in the estimates of parameters. In Chapter 1, we examined the induced errors in the estimates of site position due to several types of errors in the formulas for the propagation delay. The coordinate of site position whose estimate is most affected is the vertical, because the estimate of this coordinate is highly correlated with estimates of atmospheric parameters of almost any type. These correlations can be reduced—but not eliminated—by having observations at low elevation angles. Unfortunately, systematic errors of atmospheric origin are generally larger at low elevation angles.

A solution to this dilemma would be to eliminate errors in the atmospheric formulas and to estimate no atmospheric parameters, or else to estimate these parameters only if strongly constrained with *a priori* information. Such a solution could be obtained by using observations only from high elevation angles, thereby approximately eliminating mapping-function errors, using measurements of the surface pressure to estimate the "dry" propagation delay, and having accurate WVR estimates of the "wet" propagation delay. Up until the present, we had formulas for mapping functions believed to be adequate above the elevation of 20°, we routinely and continuously measured and recorded the surface pressure during VLBI observing sessions, but the accuracy of the WVR delay algorithm had never been not well studied.

In studying the accuracy of the WVR delay algorithm, described in Chapter 2, we became aware of a possible error at low elevation angles in the mapping function we

were then using. The use of observations at low elevation angles was necessary to this study to reduce the correlations mentioned above. The evidence for this mapping-function error was presented in Chapter 3. We then attempted to develop a new mapping function for use at elevation angles down to about 5°. The development of this new mapping function was also described in Chapter 3. The initial tests of this mapping function, using VLBI data obtained in the MERIT intensive campaign of 1980, seemed to indicate that systematic errors in the estimates of the baseline length, which had been ~5 cm for 8000-km long baselines with use of the Marini mapping function, were less than 1 cm with use of the new mapping function, dubbed CfA–2.2.

In order to test possible site- and season-dependence of the mapping function, a series of special VLBI experiments was undertaken. These experiments had a large number of observations from very low elevation angles, and were intended to increase the effects of errors in the mapping functions to cause the systematic errors to be distinguishable from random errors. These experiments, described in detail in Chapter 4, presented us with an unexpected result. Instead of the estimates of baseline length displaying the effects of systematic mapping-function errors with seasonal dependence, these estimates proved to be some of the most repeatable for that length of baseline (~3,900 km) ever obtained, with a weighted-root-mean square deviation from their weighted mean of ~4 mm. These estimates were obtained using the CfA–2.2 mapping function. The estimates obtained using the Marini mapping function have approximately twice this scatter, and display what may be a seasonal variation. One possible explanation is that the primary mapping function errors were caused by erroneous temperature dependence of the Marini mapping function, and that the CfA–2.2 mapping function is a better approximation in this regard.

We were also able to compare estimates of the wet delay obtained from WVR data at the Mojave site to estimates of the wet delay at that site obtained from the VLBI data processed with a Kalman filter. These comparisons demonstrate that the rapid variations observed in the WVR-derived estimates also appear in the VLBI-derived estimates—most of the time. There are several instances in which large variations appear in one of the series but not in the other. These differences are as yet unexplained, although possible causes are discussed in Chapter 4.

Using the WVR-derived delays as *a priori* estimates of the wet delay, we can also use the VLBI data to estimate constant daily corrections to those estimates. These corrections display a possible annual variation. Such an annual variation is expected, based on recent detailed studies of the variation of the dual-frequency weighting function (Chapter 2), which show that quantity also to have an annual variation. The variation of the corrections is "in phase" with the expected variation based on these studies, but the amplitude of the variations is too large. This discrepancy may reflect an error in the "oxygen correction" (Chapter 2). Several groups are currently working to improve algorithms for the estimation of the wet path delay from WVR data [Johansson, 1985; Robinson, 1985; Gary *et al.*, 1985].

The "low-elevation" experiments are continuing at present. They will continue to furnish information on the accuracy of our atmospheric models. We will also use radiosonde data to investigate independently the annual variations in the weighting function at Mojave. In the near future, a second WVR will be placed near the other terminal of the interferometer, at the Haystack Observatory, allowing us to assess both the site- and season-dependence of the weighting function.

Appendix A

The CfA–2.2 Mapping Function

In this appendix, we include the paper *Geodesy by radio interferometry: Effects of atmospheric modeling errors on estimates of baseline length.* In this paper, we supplement the evidence for errors in the Marini mapping function given in Chapter 3. We then use the ray-trace procedure described in Chapter 3 to develop a new mapping function. We also describe the initial testing of the new mapping function.

Radio Science, Volume 20, Number 6, Pages 1593–1607, November–December 1985

Geodesy by radio interferometry: Effects of atmospheric modeling errors on estimates of baseline length

J. L. Davis,[1] *T. A. Herring,*[2] *I. I. Shapiro,*[2] *A. E. E. Rogers,*[3] *and G. Elgered*[4]

(Received March 14, 1985; revised June 27, 1985; accepted June 27, 1985.)

Analysis of very long baseline interferometry data indicates that systematic errors in prior estimates of baseline length, of order 5 cm for ~8000-km baselines, were due primarily to mismodeling of the electrical path length of the troposphere and mesosphere ("atmospheric delay") Here we discuss observational evidence for the existence of such errors in the previously used models for the atmospheric delay and develop a new "mapping" function for the elevation angle dependence of this delay. The delay predicted by this new mapping function differs from ray trace results by less than ~5 mm, at all elevations down to 5° elevation, and introduces errors into the estimates of baseline length of ≲ 1 cm, for the multistation intercontinental experiment analyzed here

1. INTRODUCTION

A signal from a distant radio source received by an antenna located on the surface of the earth will have been refracted by the terrestrial atmosphere. The corresponding delay introduced by the atmosphere depends on the refractive index along the actual path traveled by the received signal. For an atmosphere which is azimuthally symmetric about the receiving antenna, this delay depends only on the vertical profile of the atmosphere and the elevation angle of the radio source. The function which describes the dependence upon elevation angle of the atmospheric delay has become known as the mapping function. This mapping function is used, along with some model for the zenith delay, to account for the atmospheric delay in models for the interferometric observables. Historically, analyses of very long baseline interferometry (VLBI) data for geodesy [e.g., *Robertson*, 1975] have made use of the "Chao" mapping function and, more recently [*Clark et al.*, 1985], of the "Marini" mapping function. The Chao mapping function [*Chao*, 1972] is based on ray tracing studies in which refractivity profiles, averaged over all seasons and various sites, were used. This mapping function therefore contains no parametrization based on surface weather conditions at the site. However, for the refractivity profiles used, the mapping function describes the elevation angle behavior of the atmospheric delay to better than 1% for elevation angles above 1°. On the other hand, the Marini mapping function [*Marini*, 1972; J. W. Marini, unpublished manuscript, 1974] contains terms which depend on surface meteorological conditions, but it is based on approximations that degrade its accuracy below about 10°. Mapping functions other than these two have appeared in the literature [e.g., *Hopfield*, 1969; *Saastamoinen*, 1972; *Black*, 1978; *Black and Eisner*, 1984], but these other mapping functions have not undergone testing with VLBI data.

In the following, we review the manner in which the atmospheric delay is modeled, and the effects of these models on estimates of baseline length made from radio interferometric data. We present evidence of systematic errors in estimates of baseline length made from VLBI data, and we demonstrate that these errors are caused by the mismodeling of the atmosphere. Finally, we describe the development and testing of a new mapping function.

2. MODELING THE ATMOSPHERIC DELAY

The models for the interferometric observables of group and phase delay, and of phase delay rate, must account for the atmospheric delay:

$$\tau_a = \int_{\text{atm}} ds\, n(s) - \int_{\text{vac}} ds \qquad (1)$$

[1]Department of Earth, Atmospheric, and Planetary Sciences, Massachusetts Institute of Technology, Cambridge.

[2]Harvard-Smithsonian Center for Astrophysics, Cambridge, Massachusetts.

[3]Haystack Observatory, Westford, Massachusetts.

[4]Onsala Space Observatory, Chalmers University of Technology, Sweden.

Paper number 5S0522.
0048-6604/85/005S-0522$08.00

where the first integral is evaluated along the path of a hypothetical ray originating from the direction of the radio source and passing through the atmosphere to a receiving antenna, and $n(s)$ is the index of refraction at the point s along the path; the second integral is evaluated along the path the ray would take were the atmosphere replaced by vacuum. For simplicity, we have chosen units in which the speed of light is unity. (Delay will therefore be expressed in units of equivalent length.) The difference, $\tau_a^{(i)} - \tau_a^{(j)}$, for the two antennas i and j, of an interferometer gives the contribution of the atmosphere to the model of the interferometric delay. (Using the term "tropospheric delay" here would be inaccurate, since about 25% of the atmospheric delay occurs above the troposphere.)

We can find the point at which the integration in both parts of (1) is terminated at the earth by visualizing the path of the (hypothetical) paraxial ray. This ray would strike the vertex of the paraboloid of the antenna normal to the surface of the antenna and be reflected back along the axis. For a prime focus antenna, this paraxial ray would continue to travel until it enters the antenna feed at the focus. For a Cassegrain focus antenna, the ray would be reflected once more at the subreflector and then enter the feed located at or above the vertex along the axis. The path(s) after the initial reflection can be ignored in evaluating (1), because the delay is very nearly constant. The daily variation is usually less than 0.5 mm per 10 m of travel. (The largest diurnal variation in this delay, calculated from data taken by meteorological sensors located at the sites, was recorded for the Westford antenna site; the value was 0.8 mm per 10 m of travel and was associated with a rapid decrease in the humidity.) A constant delay of any type at one of the sites is indistinguishable from a constant clock offset or instrumental delay for that site.

For most antennas, the vertex of the primary reflector moves when pointing is changed; the size of the movement is usually a few meters. This movement can usually be ignored, with consequent negligible error, and a fixed reference point used in the evaluation of (1). For example, the intersection of axes of rotation of the Haystack antenna (one of the antennas used in VLBI experiments; see below) is located 4.3 m from the vertex along the axis of the parabola in a direction opposite to that of the prime focus. In this case, if we use the axis intersection as the fixed reference point for the evaluation of (1), the errors introduced will be equal to the neglected delay from vertex to the subreflector and back to the secondary focus (a total distance traveled of 25.2 m), minus the erroneously added path from the vertex to the intersection. These paths should contribute less than ~ 1 mm amplitude of diurnal variation (due to diurnal variations of temperature and humidity) and less than 0.01 mm variation with antenna pointing angle.

Evaluation of the second integral on the right-hand side of (1) requires only knowledge of the source and antenna coordinates. However, evaluation of the first integral requires, as well, knowledge of the index of refraction in the neighborhood of the correct ray path, which is necessary in order to obtain the path itself via Fermat's principle [*Born and Wolf*, 1970]. Since in practice it is not possible to obtain this knowledge, one usually relies on models of the structure of the atmosphere. For example, one often assumes that the index of refraction of the atmosphere is constant from the surface of the earth up to an altitude H; for altitudes above H, the index of refraction is assumed to be unity, and the bending of the ray at the atmosphere/vacuum boundary is ignored. Then for a plane parallel model of the earth and the atmosphere, (1) reduces to

$$\tau_a = \csc \varepsilon \int_0^H dz\, (n_0 - 1) \tag{2}$$

where ε is the elevation angle of the radio source and n_0 is the index of refraction at the surface of the earth.

It is possible to write (1) in a form which is motivated by the simple form of (2). Quite generally, we can write

$$\tau_a = m(\varepsilon, \mathbf{P}) \int_0^\infty dz\, [n(z) - 1] \tag{3}$$

The function $m(\varepsilon, \mathbf{P})$, which is defined by (1) and (3), depends on the elevation angle ε as well as on the parameter vector $\mathbf{P}$, which is a parametrized representation of the behavior of the index of refraction in the atmosphere. The number of elements (parameters) in $\mathbf{P}$ depends on the assumptions made about "regular" atmospheric structure. For example, if one assumes that no discernible atmospheric structure exists, then $\mathbf{P}$ will be an infinite-dimensional vector containing the index of refraction at all points. Since, as previously discussed, the refraction at all points is not known, this assumption would void (3) of any possible advantages. Instead, one usually makes some assumptions and approximations concerning the structure of the atmosphere and its effects on the ray path. A simple set of assumptions and approxi-

mations is, for example, that which led to the cosecant law. In this case, the dependence of the function m on parameters **P** other than elevation is completely absent. (See below for a discussion of the assumptions used to develop the new mapping function.)

The integral in (3) is the atmospheric delay for a radio source at zenith. This integral will be denoted τ_a^z, the "zenith delay." Its dependence on atmospheric conditions directly above the antenna is discussed in the Appendix A. The function $m(\varepsilon, \mathbf{P})$ is known also as the "mapping function." For simplicity, the dependence on the parameters **P** will be suppressed, and we will write simply $m(\varepsilon)$.

Often separate mapping functions are used for the "wet" and "dry" components of the delay:

$$\tau_a = \tau_d^z m_d(\varepsilon) + \tau_w^z m_w(\varepsilon) \tag{4}$$

where the subscript d on the zenith delays and mapping functions refers to "dry" and w to "wet." Such a form is used when, for instance, water vapor radiometer (WVR) data are used to estimate the "wet" component directly [*Resch*, 1984]. Then (4) is replaced by

$$\tau_a = \tau_d^z m_d(\varepsilon) + \tau_{\mathrm{WVR}} \tag{5}$$

where τ_{WVR} is obtained from the WVR data. The user of such formulas, however, must be extremely careful to understand exactly what is meant by the terms "dry" and "wet," because the path the radio signal travels through the atmosphere is dependent on the contributions to the index of refraction from all atmospheric constituents. Furthermore, the so-called "dry" zenith delay also contains contributions from water vapor (see Appendix A).

3. SYSTEMATIC ERRORS IN ESTIMATES OF BASELINE LENGTH

The manner in which estimates of baseline length are affected by errors in the mapping function used to model the atmospheric delay can be understood by first examining the approximate expression for the "geometric" term τ_{geom} of the group delay model

$$\tau_{\mathrm{geom}} = -\mathbf{b} \cdot \hat{\mathbf{e}} = -(r_2 \sin \varepsilon_2 - r_1 \sin \varepsilon_1) \tag{6}$$

where $\mathbf{b}$ is the baseline vector (directed from site 1 to site 2), $\hat{\mathbf{e}}$ is a unit vector in the direction of the source, r_i is the distance from the center of the earth to the ith site, and ε_i is the elevation of the source at the ith site ($i = 1, 2$), and where the total group delay (of which τ_{geom} is but one term) is defined as the time of arrival of the signal at site 2 minus the time of arrival of the signal at site 1. (See *Robertson* [1975] for a more complete discussion of the group delay model.) The feature of importance in (6) is the dependence of the group delay on the elevation of the source at each site. Any elevation-dependent error (such as a mapping function error) in the group delay model which correlates with the sine of the elevation will corrupt the least squares adjustment of the radial component of the site position. An error Δr introduced into the estimate of the radial component of the position of either site introduces a corresponding error Δb in the estimate of the length of the baseline between the sites:

$$\Delta b \cong \frac{b}{2r_e} \Delta r \tag{7}$$

where b is the baseline length and r_e is the radius of the earth. The above equation is accurate to order $(\Delta r)^2/b$.

Mapping function errors introduce systematic errors into the estimates of other parameters as well. In fact, all estimated parameters will be systematically affected, although the magnitude with which the mapping function error manifests itself in the estimate of a particular parameter depends on the functional dependence of the group delay on that parameter. Thus one can expect systematic errors in estimates of source position, earth orientation, nutation, and any and all other estimated parameters; however, for illustration, this paper will confine itself to studying only errors in baseline length estimates.

Do we have evidence of mapping function errors, and, if so, how large are they? A useful method which can be used to indicate the presence and size of elevation-dependent systematic errors, such as mapping function errors, is the "elevation angle cutoff test." In this test, all baselines are estimated simultaneously using all the data available. (Of course, other parameters are estimated along with baselines, but here and in the following, as stated above, the discussion will be confined for illustration to the effects on the estimates of baseline length.) The baselines are then reestimated with the data limited to observations above some minimum elevation angle. More estimates can be made with different elevation angle minima. If there are no elevation-dependent systematic errors, the mean of the differences between the corresponding baseline-length estimates should tend toward zero. Significant biases indicate mapping function errors. Figure 1 contains the results from such a test. Plotted are the differences in baseline-length estimates for 5° and 15° minimum

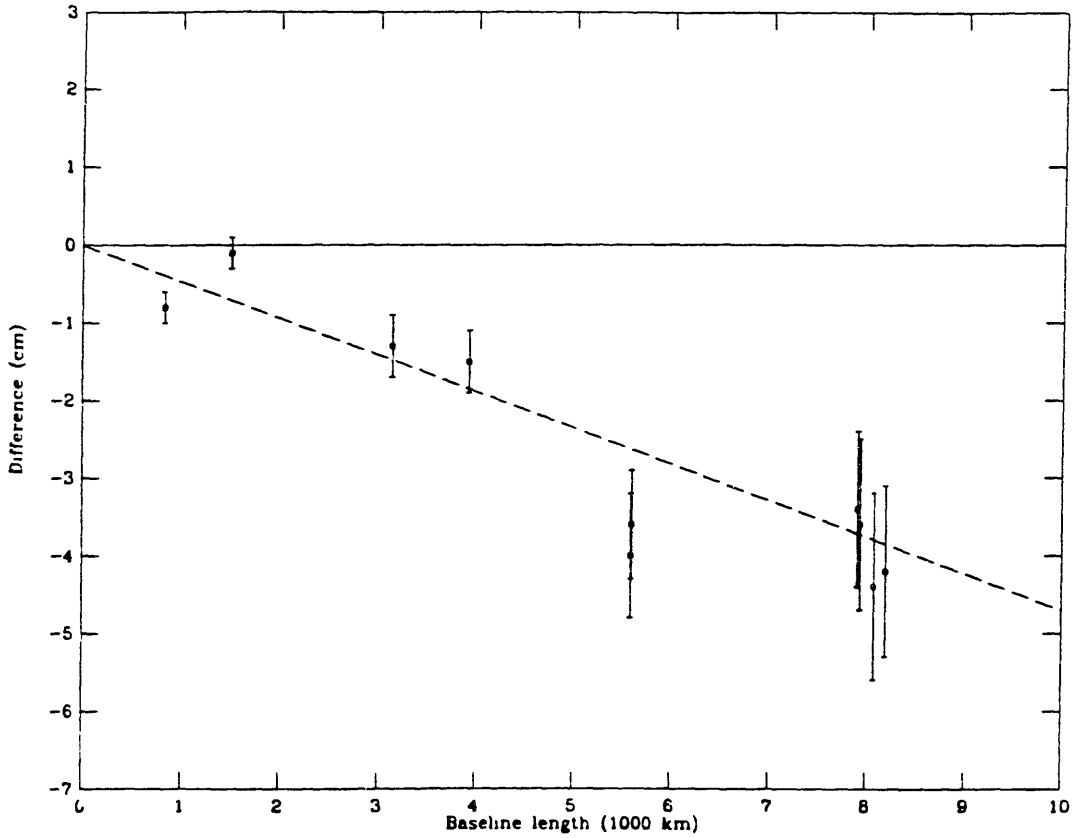


Fig. 1. Difference in baseline length estimates for the 15°–5° elevation angle cutoff test of the Marini mapping function. The error bars are the statistical standard deviations of the differences (see Appendix B). The straight line represents the effects of a change in the local vertical component of site position of 3 cm at each of the sites.

elevation angles. These differences are plotted against the length of the baseline. The error bars shown are the standard deviations of the differences, obtained from the statistical standard deviations of the individual estimates. (It can be quite easily shown that the variances of the differences of the baseline-length estimates are the differences of the corresponding variances resulting from the two least squares solutions; see Appendix B.) The group delay data used to generate these differences are the entire yield of VLBI group delay data from the project MERIT short campaign of September and October 1980 [*Robertson and Carter*, 1982], with the exception noted below. These data were processed as described by *Clark et al.* [1985]. The atmospheric delay was modeled by using the Marini formula, which requires surface weather data. The group delay data involving the site at Chilbolton, United Kingdom, were deleted, since surface weather data were not available for this site for some periods of the campaign. The phase delay rate data were not included.

From Figure 1 it can be seen that the differences in the estimates of baseline length seem to be nearly proportional to baseline length. Recalling (7), we can interpret these differences as due to corresponding differences in the estimates of radial positions of the individual sites, if these latter differences are nearly equal. For reference, Figure 1 contains a line representing the effect of a 3-cm radial difference at each site. (The sense of the radial difference is, from (7), such that the estimates of the radial positions from the 0° cutoff solution were greater than those from the 15° cutoff solution.) It can be seen that this nearly represents the actual situation. We thus conclude that the differences evident in Figure 1 are due to mapping function errors, on the assumption that there do not exist any other elevation angle–dependent errors of this magnitude.

4. THE NEW MAPPING FUNCTION

Marini [1972] showed that the continued fraction form of the mapping function

$$m(\varepsilon) = \cfrac{1}{\sin\varepsilon + \cfrac{a}{\sin\varepsilon + \cfrac{b}{\sin\varepsilon + \cfrac{c}{\sin\varepsilon + \cdots}}}} \tag{8}$$

where a, b, c, $\cdots$ are constants, can be used to approximate the elevation angle dependence of the atmospheric delay. Only two terms are used in the Marini mapping function (J. W. Marini, unpublished manuscript, 1974). *Chao* [1972] uses two terms as well, except he replaces the second sin ε with tan ε, thereby ensuring that $m(90°) = 1$. We have attempted to develop a mapping function for the "dry" or "hydrostatic" component of the atmosphere (see Appendix A) based on the Chao model, but with improved accuracy at low elevation angles. In order to achieve subcentimeter accuracy at 5° elevation, we have "continued the fraction" by adding one more term but keeping the tangent:

$$m(\varepsilon) = \cfrac{1}{\sin\varepsilon + \cfrac{a}{\tan\varepsilon + \cfrac{b}{\sin\varepsilon + c}}} \tag{9}$$

The advantage of using this form is its simplicity, both in calculating the mapping function itself and in calculating partial derivatives of the mapping function with respect to the parameters to be estimated. The disadvantage of this form is that for higher elevation angles (20°–60°), tan ε does not approach sin ε quickly enough. As a result, one can expect 1- to 2-mm errors in representing the atmospheric delay with (9) for these elevations.

In order to determine the mapping function parameters a, b, and c, we performed ray trace analyses for various values of a limited number of atmospheric conditions. The ray trace algorithm we used was based on a spherically symmetric, layered atmosphere. The temperature profile was taken to have a linear dependence with height up to the tropopause, above which the temperature was assumed constant. The total pressure was assumed to result from hydrostatic equilibrium, and the relative humidity was assumed to be constant up to 11 km and zero above that height. The acceleration due to gravity was assumed to be constant with height. This simple set of assumptions concerning the structure of the atmosphere allowed us to examine the dependence of the mapping function on variations about the nominal values of the following parameters: surface pressure, surface relative humidity, surface partial pressure of water vapor, temperature of the tropopause, and height of the tropopause. However, the sampling of parameter space was not done in a systematic manner due to the large number of ray trace analyses which this would entail. For example, if just three values for each parameter were used, there would be $3^5 = 243$ different combinations of parameters. Instead, 57 analyses were performed, and there are resulting gaps in the sampling of the parameter space.

For each set of atmospheric conditions, then, we determined the ray trace values for the mapping function, in steps of 1° for elevations from 5° to 90°. We then used least squares to estimate a, b, and c. However, c could be fixed at some nominal value and not appreciably degrade the solution; the nominal value ultimately decided upon (see below) for c was taken to be the approximate average of the values for the first several ray traces performed. The mapping function form given in (9) was, for each set of atmospheric conditions, able to model the elevation angle dependence of the delay to within 3 mm for all elevation angles down to 5°, and with an rms deviation of less than 1.5 mm.

The ray trace analyses thus provided a set of estimates of each of the mapping function parameters, a and b, covering a variety of atmospheric conditions. We then represented a and b each as a linear function of the various atmospheric parameters that were varied and used least squares to determine the coefficients. Such a linear model fits the mapping function parameter a within 0.2% (corresponding to ~5 mm at 5° elevation) and the parameter b to within 0.5% (~2 mm at 5° elevation) in all cases; the rms fit for a is 0.08% (~2 mm at 5° elevation) and for b is 0.15% (~0.6 mm at 5° elevation). In particular, we have

$$\begin{aligned} a = 0.001185\,[1 &+ 0.6071 \times 10^{-4}(P_0 - 1000) \\ &- 0.1471 \times 10^{-3} e_0 \\ &+ 0.3072 \times 10^{-2}(T_0 - 20) \\ &+ 0.1965 \times 10^{-1}(\beta + 6.5) \\ &- 0.5645 \times 10^{-2}(h_t - 11.231)] \end{aligned} \tag{10}$$

$$b = 0.001144[1 + 0.1164 \times 10^{-4}(P_0 - 1000) + 0.2795 \times 10^{-3}e_0 + 0.3109 \times 10^{-2}(T_0 - 20) + 0.3038 \times 10^{-1}(\beta + 6.5) - 0.1217 \times 10^{-1}(h_t - 11.231)] \quad (11)$$

$$c = -0.0090 \quad (12)$$

where P_0 is the total surface pressure in millibars, e_0 is the partial pressure of water vapor at the surface in millibars, T_0 is the surface temperature in degrees Celsius, β is the tropospheric temperature lapse rate in K km^{-1}, and h_t is the height of the tropopause in kilometers. This version of the "dry" mapping function has been dubbed CfA-2.2. The sensitivities of the CfA-2.2 mapping function to changes in these atmospheric parameters are summarized in Table 1. For example, a 10-mbar change in the partial pressure of water vapor produces a change of approximately 5 mm in the predicted delay at 5° elevation. (It is fortunate, in fact, that the mapping function is not very sensitive to the amount of water vapor in the atmosphere, since this quantity is spatially highly variable and not well predicted by surface measurements.) Figures 2 and 3 contain the differences between ray

TABLE 1. Sensitivities of the Path Delay From Model CfA-2.2 to Changes in Atmospheric Parameters for $\Delta L_{zen} = 240$ cm and for Different Elevations ε

ε	P, cm/mbar	T, cm/°C	e, cm/mbar	β, cm/(K km)	h_t, cm/km
15	-9.1×10^{-4}	−0.046	0.002	−0.29	0.082
10	-2.8×10^{-3}	−0.14	0.007	−0.88	0.25
5	−0.017	−0.75	0.053	−4.4	1.1

See text for explanation of model CfA-2.2 P, pressure; T, temperature; e, partial pressure of water vapor; β, temperature lapse rate; h_t, height of tropopause.

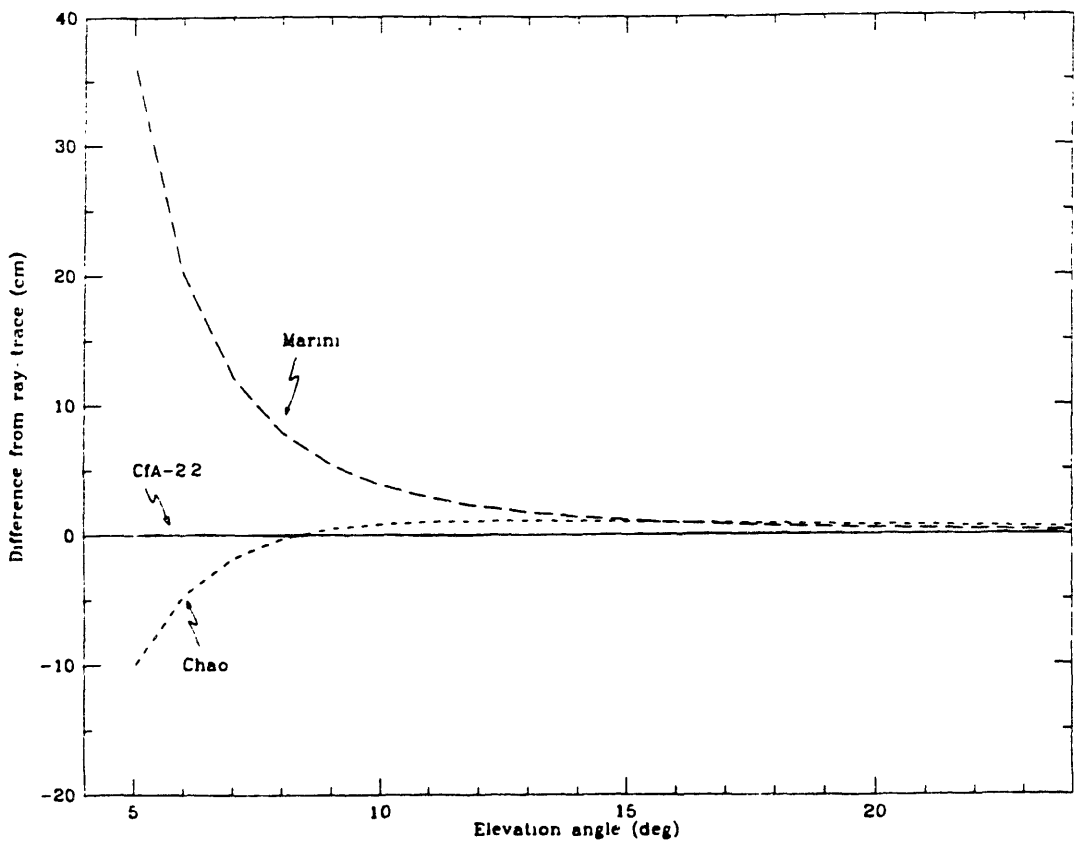


Fig 2. Differences from ray tracing of the new mapping function, the Chao mapping function, and the Marini mapping function for $P_0 = 850$ mbar and $T_0 = 15$°C. The partial pressure of water vapor, temperature lapse rate, and tropopause height are all at their nominal values of 0 mbar, −6.5 K/km, and 11.231 km, respectively. The corresponding value of the zenith delay is 1.935 m of equivalent length (6.5 ns).

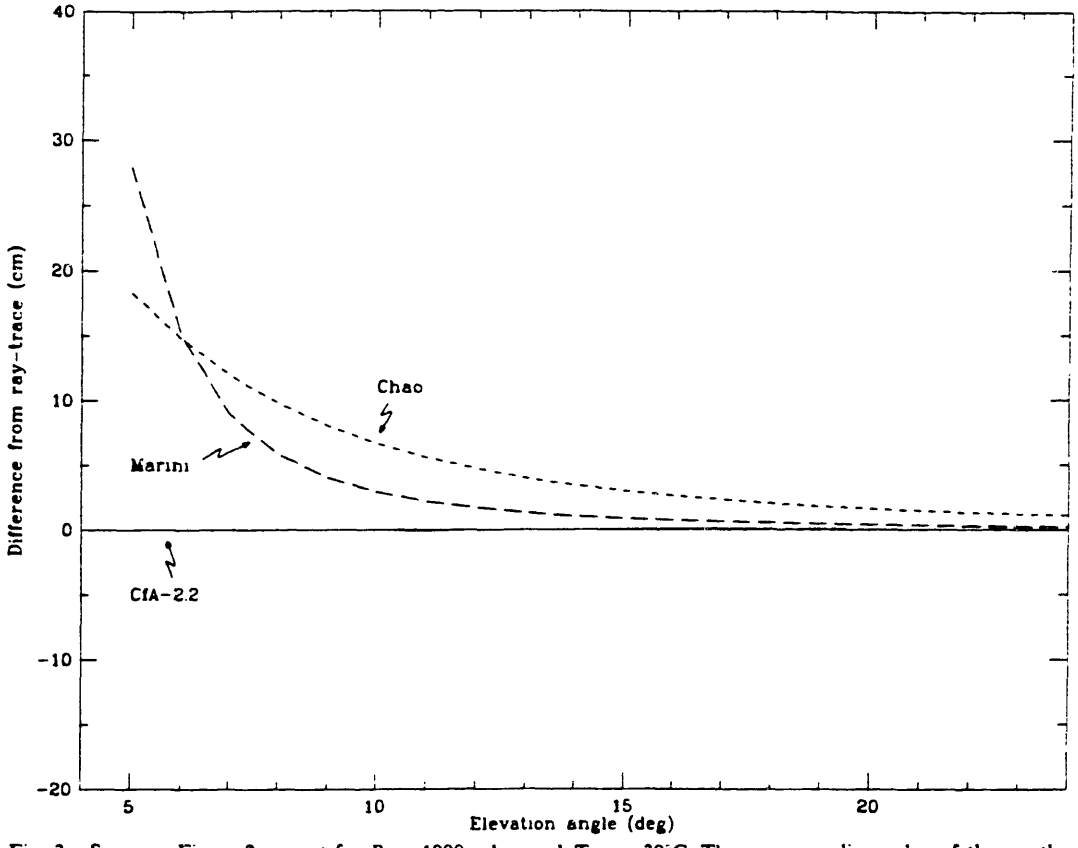


Fig. 3. Same as Figure 2, except for P_0 = 1000 mbar and $T_0 = -30$°C. The corresponding value of the zenith delay is 2.277 m.

tracing and CfA-2.2 for the different atmospheric conditions indicated. These conditions, which represent the nominal conditions of humidity, lapse rate, and tropopause height, were chosen because they represent locations in the pressure-temperature plane near which no ray trace analyses were performed. Their agreement with ray tracing is therefore an indication of the robustness of the method used in the plane of temperature and pressure. Also shown are the differences from ray tracing for the Chao and Marini models. These models are the most commonly used mapping functions in VLBI data analysis [*Fanselow*, 1983; *Clark et al.*, 1985].

The accuracy of the CfA-2.2 mapping function model seems higher near latitudes of 45°N, for which the nominal values of tropopause height and lapse rate used in CfA-2.2 are representative. For example, for conditions representative of a latitude of 30°N (h_t = 16 km, $\beta = -4.7$ K/km to -5.9 K/km), the difference between CfA-2.2 and ray trace values reaches ~4 cm at 5° elevation. Relatively large differences have also been noted for higher latitudes in the extreme of winter: For a latitude of 60°N (h_t = 8 km, $\beta = -3.9$ K/km), the differences from the ray trace values reach ~2.5 cm at 5°. These (comparatively) large differences from the ray trace values seem to be due to the simultaneous departures from the nominal values of lapse rate and tropopause height. Although our choice of -6.5 K/km is the standard one for the lapse rate in the troposphere (U.S. Standard Atmosphere, 1976), it seems to be somewhat large (in magnitude) when one considers compilations of temperature profiles found, for example, in the work of *Smith et al.* [1963]. However, even with a better choice of nominal value, a site-dependent model of some type will have to be developed: lapse rate and tropopause height do not truly vary independently, since the temperature of the tropopause varies less than the surface temperature. Thus those climates with a very low tropopause height (high latitudes) can be expected to have correspondingly small (in magnitude) lapse rates. Those climates with a high

tropopause height (equatorial latitudes) will have correspondingly large (in magnitude) lapse rates.

The lack of dependence of the new mapping function on azimuth results directly from the assumption of azimuthal symmetry. *Gardner* [1977] expressed the index of refraction in (1) in cylindrical coordinates (with the z coordinate aligned along the local vertical) and expanded it in powers of horizontal distance from the z axis. He showed that the zeroth-order term represents the spherically symmetric term. Thus our new mapping function represents this zeroth-order term. The first-order term in Gardner's expansion arises from horizontal refractivity gradients; Gardner showed that this term can be as large as 5 cm at 10° elevation angle. However, this term has never been included in our VLBI data analysis because of the lack of a network of meteorological sensors in the near vicinity of our sites from which to determine the refractivity gradients. In principle, though, there is no reason that this gradient term could not be introduced into our atmospheric models; its utility would depend on (1) a dense enough network of meteorological sensors being in place around each site, (2) models for the gradient being developed that depend on the meteorological conditions at the site only (such as wind direction and speed), and possibly on climate and/or season, or (3) the ability to estimate accurately the gradient term being demonstrated for data from a network of distant (>100 km) meteorological sensors such as exist at airports and other weather stations.

5. PROCESSING VLBI DATA WITH CfA-2.2

We have performed the elevation-angle-cutoff test on the CfA-2.2 mapping function. For this test, the Saastamoinen formula for the zenith delay [*Saastamoinen*, 1972] was used to be consistent with the zenith delay values used for the Marini formula. The "wet" part of the delay (see Appendix A) was mapped by using (9)–(12) as well, even though this use introduces a small error which is, from (4), the "wet" delay multiplied by the difference between CfA-2.2 and the "true" wet mapping function. The values listed in Table 2 were used for tropopause height and lapse rate. These values are based on tables of mean temperature profiles near the 80th meridian west [*Smith et al.*, 1963]. No attempt was made to obtain the exact profiles of temperature that prevailed at the sites, since for this elevation angle cutoff test we were attempting only to remove the gross effects of differences from the nominal values of β and h_t. The procedure used was first to obtain estimates of β and h_t at the latitudes of 30°, 40°, and 50° by fitting a linear function of height to the values given in the tables; possible variations of these parameters with longitude were ignored. The three estimates for each of the parameters β and h_t were then expressed via least squares as second-order polynomials in latitude. For each North American site, the latitude of the antenna was then substituted to determine β and h_t. Each European site was treated as though it were 5° south of its true position to account approximately for the warmer climate at European longitudes in the choice of β and h_t. (The value of 5° was based upon visual inspection of world maps of tropopause height found by *Bean et al.* [1966].)

TABLE 2. Values for Tropopause Height h_t and Temperature Lapse Rate β Used in Elevation Angle Cutoff Test of CfA-2.2 Mapping Function

Site Name	Geographic Location	North Latitude	h_t,* km	β, K/km
Onsala	south Sweden	57°	10.5	−5.7
Effelsberg	West Germany	51°	9.6	−5.7
Haystack	east Massachusetts	43°	13.6	−5.6
Owens Valley	south California	37°	12.8	−5.6
Fort Davis	southwest Texas	31°	13.4	−6.3

*Height of tropopause given as height above station for direct use in CfA-2.2 mapping function formula; see text.

The results of this elevation angle cutoff test are shown in Figure 4. Any systematic trend that may be present in this figure is clearly much smaller than that seen in Figure 1. Table 3 allows us to compare the results from the two tests more quantitatively. The second column contains least squares estimates of the differences in the radial site positions which, from (7), would yield the baseline-length differences evident in Figure 1. The third column contains the same information, except for Figure 4. The fourth column contains the differences of the second two columns. The numbers in this fourth column, then, represent the changes in the inferred differences of the radial position we obtained in performing the 15°–5° elevation angle cutoff tests. It can be seen that for the sites at Haystack, Onsala, and Effelsberg, these changes were ~4 cm, over 10 times the changes at Ft. Davis and Owens Valley. This difference can be explained by the entries in the fifth column. This column contains the fraction of data obtained below 15° elevation at these sites. That these last two sites had no data from these lower elevation angles

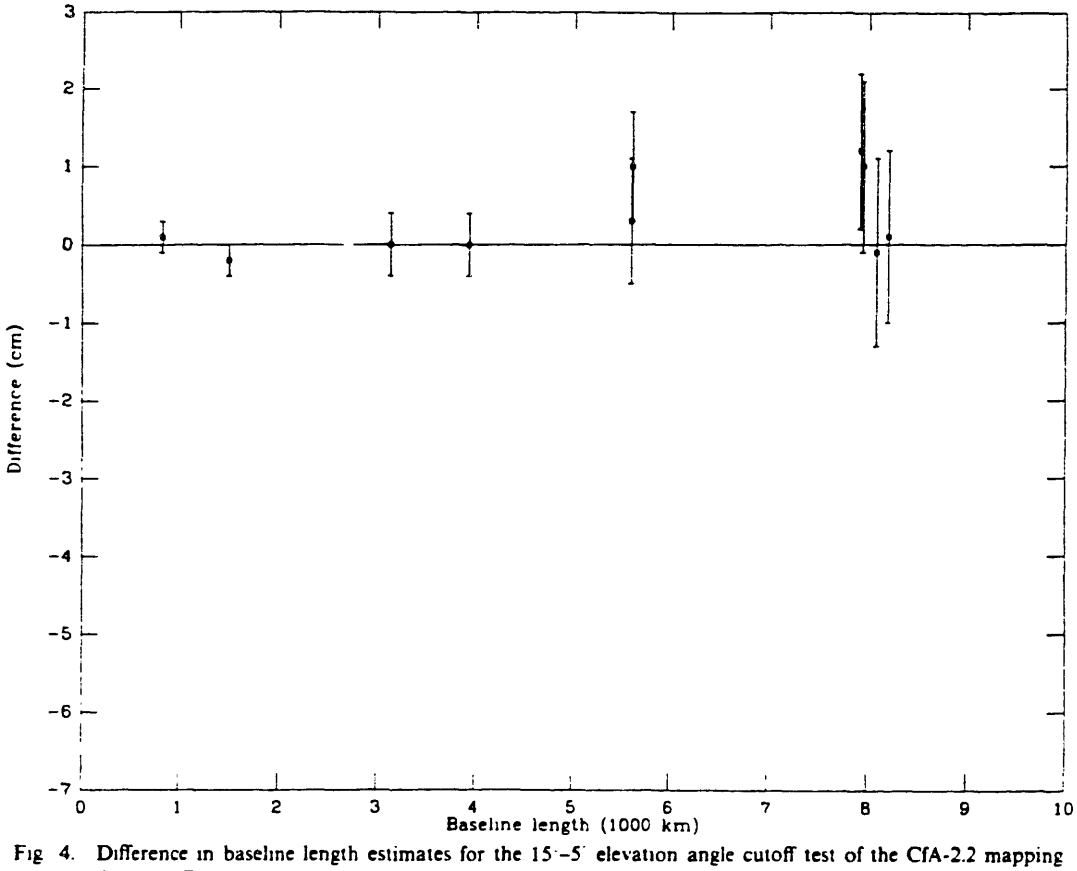


Fig 4. Difference in baseline length estimates for the 15°–5° elevation angle cutoff test of the CfA-2.2 mapping function. The error bars are the statistical standard deviations of the differences (see Appendix B)

implies that there should be little difference between the results for either of these sites from using different mapping functions.

Further testing of the CfA-2.2 mapping function is underway. Single-baseline experiments are now being carried out in which a large fraction of the observations from one site are obtained for elevation angles below 5° elevation, and for a very large fraction below 10° elevation, while observations from the other site remain at relatively high (>10°) elevation angles. This procedure should enable us to isolate mapping function errors for the site at which the low elevation observations are taken, since we would be relatively insensitive to mapping function errors for the other site.

Plans are also being made to optimize the coefficients in (10)–(12) for site location, and to develop seasonal atmospheric structure parameters. For this purpose, radiosonde data obtained from the National Climatic Data Center for U.S. sites, and from various European centers, will be used. Simultaneously, an effort will be made to attempt to increase the accuracy of the mapping function at all elevation angles (but with emphasis at extending the mapping function for use at elevation angles below 5°) and to in-

TABLE 3 Comparison of Elevation Angle Cutoff Tests

Site Name	Δr, cm (Marini)	Δr, cm (CfA-2.2)	Difference, cm	Fraction of Data Below 15° Elevation
Onsala	−3.7 ± 1.0	0.9	−4.6	10.4%
Effelsberg	−4.9 ± 1.3	−0.7	−4.2	7.7%
Haystack	−3.9 ± 0.8	0.1	−4.0	3.1%
Owens Valley	−1.0 ± 0.3	−0.8	−0.2	0%
Fort Davis	−1.4 ± 0.5	−1.1	−0.3	0%

The entries in the columns headed by Δr are the changes in the estimates of the local vertical positions of the sites corresponding to the baseline length differences from each of the elevation angle cutoff tests, shown in Figure 1 (Marini) and Figure 4 (CfA-2.2). The column labeled "difference" is the difference between the changes in the radial estimates. The uncertainties for the values of Δr (CfA-2.2) are the same as for the values for Δr (Marini)

vestigate possible means for modeling of horizontal gradients.

6. SUMMARY

Errors in modeling the elevation angle dependence of the atmospheric delay can cause systematic errors in the estimated radial positions of the antenna sites. These radial errors will "map" into the estimates of baseline length by an amount approximately proportional to the baseline length. An elevation angle cutoff test performed with the Marini mapping function indicated that the errors in the estimates of baseline length introduced by this mapping function were of the order of ~ 5 cm for a baseline length of ~ 8000 km and that these errors display a systematic dependence on baseline length indicative of a mapping function error. A new mapping function has been developed which is based on ray-tracing through model atmospheres. Repetition of the elevation angle cutoff test with this new mapping function yields apparent errors in baseline-length estimates of $\lesssim 1$ cm, with the differences showing little or no dependence on baseline length.

APPENDIX A: ZENITH DELAY FORMULAS

The purpose of this appendix is to derive an accurate expression for the zenith delay from the wet and dry refractivity formulas. We pay particular attention to the treatment of the wet/dry mixing ratio. We also obtain an estimate for the accuracy of the hydrostatic (i.e., "dry") delay formula, and derive an expression for the "wet" zenith delay which is consistent with the "dry" zenith delay formula. This "wet" delay formula makes use of the most recent expression for the wet refractivity and can be used to establish the relationship between the observables of instruments which measure the radiative emission of atmospheric water vapor (e.g., water vapor radiometers) and the line-of-sight delay due to water vapor.

Derivation of the zenith delay from the refractive index

The three-term formula for the total refractivity of moist air, as given by *Thayer* [1974], is

$$N = k_1 \frac{p_d}{T} Z_d^{-1} + k_2 \frac{p_w}{T} Z_w^{-1} + k_3 \frac{p_w}{T^2} Z_w^{-1} \qquad \text{(A1)}$$

Here T is the temperature, p_d is the partial pressure of the "dry" constituents ("dry" is defined below), p_w the partial pressure of water vapor, and Z_d^{-1} and Z_w^{-1} are the respective inverse compressibilities, with the subscripts having the same meaning as for the pressures. The symbol e is usually used in place of p_w. Thayer's values for the constants k_1, k_2, and k_3 are summarized in Table A1. The uncertainties of these values limit the accuracy with which the refractivity can be calculated to about 0.02%.

The first term in (A1) represents the effect of the induced dipole moment ("displacement polarization") of the dry constituents. The second term represents the same effect for water vapor, whereas the third term represents the dipole orientation effects of the permanent dipole moment of the water molecule. None of the primary constituents of dry air (shown in Table A2) possesses a permanent dipole moment.

The values for k_2 and k_3 listed in Table A1 have been disputed by *Hill et al.* [1982]. They point out that Thayer's extrapolation of the value of k_2 from its value at optical wavelengths ignores the effect of the rotational and vibrational resonances in the infrared [*Van Vleck*, 1965]. Hill et al. calculate a theoretical value for k_2 and k_3 and find $k_2 = 98 \pm 1$ K/mbar and $k_3 = (3.583 \pm 0.003) \times 10^5$ K^2/mbar. However, these results are so greatly in disagreement with published values of k_2 and k_3, which have been obtained by measurements in the microwave region [*Boudouris*, 1963; *Birnbaum and Chatterjee*, 1952]. that Hill recommends using the measured values instead of either his or Thayer's. Birnbaum and Chatterjee find $k_2 = 71.4 \pm 5.8$ K/mbar and $k_3 = (3.747 \pm 0.029) \times 10^5$ K^2/mbar, while Boudouris finds $k_2 = 72 \pm 11$ K mbar and $k_3 = (3.75 \pm 0.03) \times 10^5$ K^2/mbar. As a compromise, we keep Thayer's values for k_2 and k_3 (which differ from the experimental values by less than the uncertainties of the latter), but choose the (rounded) experimental uncertainties, as shown in Table A1.

The grouping together of all the dry constituents into one refractivity term is possible because the relative mixing ratios of these gasses remain nearly constant in time and over the surface of the earth [*Glueckauf*, 1951]. The eight main constituents of the dry atmosphere are listed in Table A2, along with their molar weight and fractional volume, and a standard deviation representing the variability of that constituent in the atmosphere. Using these numbers, we find the mean molar weight M_d of dry air to be $M_d = 28.9644 \pm 0.0014$ kg/kmol, where the standard deviation is an upper bound on the variability of M_d based on the values in Table A1 and on the assumption that these constituents vary independently.

TABLE A1. Constants Used in the Appendix

Constant	Value*	Uncertainty*	Uncertainty Used (See Text)	Units
k_1	77.604	0.014	0.014	K mbar^{-1}
k_2	64.79	0.08	10	K mbar^{-1}
k_3	377600	400	3000	K^2 mbar^{-1}

Derived Constant	Equivalent	Units
k_2'	$k_2 - k_1 \frac{M_w}{M_d} = 17 \pm 10$	K mbar^{-1}
k_3'	$k_3 + k_2' T_m = (3.82 \pm 0.04) \times 10^5$	K^2 mbar^{-1}

Here $M_d = 28.9644 \pm 0.0014$ kg kmol^{-1}, molar mass of dry air, $M_w = 18.0152$ kg kmol^{-1}, molar mass of H_2O; $T_m = 260 \pm 20$ K, "mean temperature" (see text).
*From *Thayer* [1974].

The inverse compressibilities in (A1) represent the nonideal behavior of their respective atmospheric constituents. This behavior is described by the equation of state for the ith constituent $p_i = Z_i \rho_i R_i T$, where p_i is the partial pressure, Z_i is the compressibility, ρ_i is the mass density, and R_i is the specific gas constant for that constituent ($R_i = R/M_i$, where R is the universal gas constant and M_i is the molar mass), and T is the absolute temperature. For an ideal gas $Z = 1$; Z differs from unity by a few parts per thousand for the atmosphere. The expressions for the inverse compressibility Z_d^{-1} for dry air and Z_w^{-1} for water vapor were determined by *Owens* [1967] by least squares fitting to thermodynamic data. These expressions are

$$Z_d^{-1} = 1 + p_d[57.97 \times 10^{-8}(1 + 0.52/T) - 9.4611 \times 10^{-4} t/T^2] \quad \text{(A2)}$$

TABLE A2. Primary Constituents of Dry Air and Their Variability

Constituent	Molar Weight,* kg/kmol	Fractional Volume,† (Unitless)	σ†
N_2	28.0134	0.78084	0.00004
O_2	31.9988	0.209476	0.00002
Ar	39.948	0.00934	0.00001
CO_2	44.00995	0.000314	0.000010
Ne	20.183	0.00001818	0.0000004
He	4.0026	0.00000524	0.00000004
Kr	83.30	0.00000114	0.0000001
Xe	131.30	0.000000087	0.000000001

*U.S Standard Atmosphere (1976)
†*Gleuckauf* [1951]

and

$$Z_w^{-1} = 1 + 1650(p_w/T^3)(1 - 0.01317t + 1.75 \times 10^{-4}t^2 + 1.44 \times 10^{-6}t^3) \quad \text{(A3)}$$

where t is the temperature in degrees Celsius, p_d and p_w are in millibars, and T is in Kelvins. Owens found that (A2) and (A3) model the compressibility to within a few parts per million.

The total zenith delay L_z is

$$L_z = 10^{-6} \int_0^\infty dz\, N(z) \quad \text{(A4)}$$

Integration of the refractivity in the form given in (A1) requires knowledge of the profiles of both the wet and dry constituents, the mixing ratio of which is highly variable. However, it is possible to create a term nearly independent of this mixing ratio. We can rewrite the first two terms in (A1) by using the equation of state as

$$k_1 \frac{p_d}{T} Z_d^{-1} + k_2 \frac{p_w}{T} Z_w^{-1} = k_1 R_d \rho_d + k_2 R_w \rho_w = k_1 R_d \rho + k_2' \frac{p_w}{T} Z_w^{-1} \quad \text{(A5)}$$

where the total mass density $\rho = \rho_d + \rho_w$ is indicated by the absence of subscripts, and the new constant k_2' is given by

$$k_2' = k_2 - k_1 \frac{R_d}{R_w} = k_2 - k_1 \frac{M_w}{M_d} \quad \text{(A6)}$$

If a value for the molar weight of water $M_w = 18.0152$ kg/kmol is used [*CRC Press*, 1974], and if

independent errors in k_1, k_2, and M_d are assumed, then we find $k_2' = (17 \pm 10)$ K mbar^{-1}. By using (A5), we find the expression for the total refractivity to be

$$N = k_1 R_d \rho + k_2' \frac{P_w}{T} Z_w^{-1} + k_3 \frac{P_w}{T^2} Z_w^{-1} \tag{A7}$$

It is important to note that the first term in (A7) is dependent only on the total density and not on the wet/dry mixing ratio. This term can be integrated by applying the condition that hydrostatic equilibrium is satisfied:

$$\frac{dP}{dz} = -\rho(z)g(z) \tag{A8}$$

where $g(z)$ is the acceleration due to gravity at the vertical coordinate z, $P(z)$ is the total pressure, and, as above, $\rho(z)$ is the total mass density. Denoting the result of the integration of the first term in (A7) as L_1, we find that

$$L_1 = [10^{-6} k_1 R_d g_m^{-1}] P_0 \tag{A9}$$

where P_0 is the total atmospheric pressure at the intersection of rotation axes of the radio antenna (not the surface pressure, since the antenna is located some height above the ground; see text), and where g_m is given by

$$g_m = \frac{\int_0^\infty dz\, \rho(z) g(z)}{\int_0^\infty dz\, \rho(z)} \tag{A10}$$

By expanding $g(z)$ to first order in z, it can be seen that (A10) very nearly represents the acceleration due to gravity at the center of mass of the vertical column. The value of g_m at this point is [*Saastamoinen*, 1972]

$$g_m = 9.8062 \text{ m s}^{-2}(1 - 0.00265 \cos 2\lambda - 0.00031 H_c) \tag{A11}$$

where λ is the geodetic site latitude and H_c is the height in kilometers of the center of mass of the vertical column of air. The quantity H_c and therefore g_m is dependent upon the atmospheric total density profile, but *Saastamoinen* [1972] used his simple model atmosphere and "average" conditions to generate the expression

$$H_c = 0.9H + 7.3 \text{ km} \tag{A12}$$

where H is the height in kilometers of the station above the geoid. Saastamoinen claims that this expression is accurate to within 0.4 km for all latitudes and for all seasons. Substituting for H_c into (A11) yields

$$g_m = 9.784 \text{ m s}^{-2}(1 - 0.00266 \cos 2\lambda - 0.00028H) \pm 0.001 \text{ m s}^{-2} = g_m^0[f(\lambda, H) \pm 0.0001] \tag{A13}$$

where $g_m^0 = 9.784$ m s^{-2}. Combining all the constants in (A9), along with their uncertainties (assumed uncorrelated), gives

$$L_1 = [(0.0022768 \pm 0.0000005) \text{ m mbar}^{-1}] \frac{P_0}{f(\lambda, H)} \tag{A14}$$

where a value of $R = 8314.34 \pm 0.35$ J kmol^{-1} K^{-1} has been used for the universal gas constant [*CRC Press*, 1974]. The uncertainty of the constant in (A14) takes into account the uncertainty of k_1, the uncertainty in g_m, the uncertainty in R, and the variability of the dry mean molar mass. It does not include the effect due to nonequilibrium conditions. It is, in fact, difficult to assess this effect without actually integrating vertical profiles of vertical wind acceleration (which, in general, are not available); no attempt to assess this effect will be made here. *Fleagle and Businger* [1980] state that only under extreme weather conditions (thunderstorm or heavy turbulence) do these vertical accelerations reach 1% of gravity, corresponding to an error in L_1 of about 20 mm/1000 mbar. Exactly where the true uncertainty lies between these values of 0.5 and 20 mm/1000 mbar must be left to future investigation.

Because the uncertainty associated with L_1 in (A14) is so small, and because variability is associated with water vapor, L_1 is usually (and inaccurately) termed the "dry delay." Something like the "hydrostatic delay" would be more descriptive, for in principle the uncertainty of the dry density at any point is no less than the uncertainty of the wet density, whereas the total density is very predictable.

The remaining two terms in the expression for the refractivity are wet terms

$$N_w = \left[k_2' \frac{P_w}{T} + k_3 \frac{P_w}{T^2}\right] Z_w^{-1} \tag{A15}$$

The partial pressure of water is not by itself in equilibrium, and water vapor can remain relatively unmixed, making the wet delay very unpredictable. Water vapor radiometers (WVR's) will, we hope, obviate this problem. However, there are large amounts of VLBI and other data for which no WVR calibration is available, and more such data are being con-

tinually generated. Thus there is still a need for models of the zenith wet delay. No attempt will be made to develop one here. All such models in current use [e.g., *Chao*, 1972; *Berman* 1976; *Saastamoinen*, 1972] use obsolete values for the refractivity constants k_2 and k_3. However, these old values induce errors on the submillimeter level, much less than the inherent error in the prediction of the wet delay. On the other hand, these models also tend to be based on empirical models for the wet atmosphere, averaged over location and season. However, we believe that site and season dependence of the atmospheric profile could cause seasonal and site-dependent biases in these wet models of up to 10–20%, based on a comparison of expressions for "average" profiles reported throughout the literature.

Water vapor radiometers

A water vapor radiometer is a multichannel radiometer which uses the sky brightness temperature near the 22-GHz rotational absorption line of atmospheric water vapor to obtain an estimate of the integral of the wet refractivity in (A15). The WVR's now coming into use should have their "retrieval coefficients" [see *Resch*, 1984] "optimized" for site and seasonal dependence of the atmospheric profiles. For this optimization, one uses radiosonde estimates of the wet delay ΔL_w to determine the retrieval coefficients a_1 and a_2 defined in the equation

$$\Delta L_w = a_1 f(\mathrm{WVR}) + a_2 g(P_0, T_0) \tag{A16}$$

where f(WVR) is some function of the WVR observables, and $g(p_0, T_0)$ is some function of the surface temperature and pressure [*Resch*, 1983]. Both f(WVR) and $g(P_0, T_0)$ are determined by theory. By "radiosonde estimates of the wet delay" we mean that ΔL_w is determined by numerical integration of the wet refractivity given in (A15) using radiosonde profiles of p_w and T. In practice, most investigators use a one-term expression for the wet delay:

$$\Delta L_w = 10^{-6} k_3' \int dz \, \frac{p_w}{T^2} \tag{A17}$$

where k_3' is the k_3 in (A15) modified for the effect of k_2'. This modification is made possible by using the mean value theorem to introduce a "mean temperature" via

$$\int dz \, \frac{p_w}{T} = T_m \int dz \, \frac{p_w}{T^2} \tag{A18}$$

whence the (nearly) constant k_3' is given by

$$k_3' = k_3 + k_2' \, T_m \tag{A19}$$

Most investigators choose a constant value for T_m for all sites and seasons. For example, for $T_m = 260 \pm 20$ K, we find, assuming independent errors in k_2' and k_3, $k_3' = (3.82 \pm 0.04) \times 10^5$ K^2 mbar^{-1}. This approach is adequate, since the $k_2' T_m$ term is only about 1% of k_3, and based on seasonal temperature profiles, seasonal variations in $k_2' T_m$ are one order of magnitude smaller, or <0.2 mm for a zenith delay. However, it is fairly common in the literature to use an incorrect value for k_2'. This usage arises from implicitly assuming that $M_d = M_w$ in (A6), which actually changes the sign of k_2'. The value then found for k_3' is approximately 0.373, or about $\sim 2.5\%$ smaller than the 0.382 number derived here. This (incorrect) value results in an underestimate by ~ 5 mm for a zenith wet path delay of ~ 20 cm.

APPENDIX B: COVARIANCE OF DIFFERENCED PARAMETERS

In this appendix we derive the expression for the covariance matrix for the difference of two (different) least squares estimates of the same parameter vector. We assume that one of the estimates is based on a subset of the data used to make the estimate of the other. We begin by writing the linearized equation relating the observations to the parameters.

$$\mathbf{y}_1 = A_1 \mathbf{x} + \boldsymbol{\varepsilon}_1 \tag{B1}$$

where $\mathbf{x}$ is a vector of parameters to be estimated, $\boldsymbol{\varepsilon}_1$ is an unknown, Gaussian, zero-mean random vector whose covariance matrix is $G_{y,1}$, and whose mean square is to be minimized, and where $\mathbf{y}_1$ is a vector of observations. (The subscript y was chosen for the covariance matrix of $\boldsymbol{\varepsilon}_1$ to emphasize that it represents the experimental errors of the observations $\mathbf{y}_1$.) The least squares estimate $\hat{\mathbf{x}}_1$ of $\mathbf{x}$ based on $\mathbf{y}_1$ is

$$\hat{\mathbf{x}}_1 = [A_1^T G_{y,1}^{-1} A_1]^{-1} A_1^T G_{y,1}^{-1} \mathbf{y}_1 \tag{B2}$$

The covariance matrix $G_{x,1}$ of the parameter estimate $\hat{\mathbf{x}}_1$ is

$$G_{x,1} = [A_1^T G_{y,1}^{-1} A_1]^{-1} \tag{B3}$$

Let us now consider the least squares estimate of $\hat{\mathbf{x}}_t$ given a set of observations $\mathbf{y}_t$ which are composed of the previous observations $\mathbf{y}_1$ as well as a distinct set of observations $\mathbf{y}_2$:

$$\mathbf{y}_t = \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} \tag{B4}$$

We will assume that $\mathbf{y}_1$ and $\mathbf{y}_2$ are uncorrelated, so that

$$G_{y,t} = E[\mathbf{y}_t \mathbf{y}_t^T] = \begin{bmatrix} G_{y,1} & 0 \\ 0 & G_{y,2} \end{bmatrix} \tag{B5}$$

where $E[\ ]$ indicates expectation, and $G_{y,2}$ is the covariance matrix of $\boldsymbol{\varepsilon}_2$, the observational errors associated with $\mathbf{y}_2$. The least squares estimate $\hat{\mathbf{x}}_t$ based on $\mathbf{y}_t$ is therefore given by

$$\begin{aligned}\hat{\mathbf{x}}_t &= [A_t^T G_{y,t}^{-1} A_t]^{-1} A_t^T G_{y,t}^{-1} \mathbf{y}_t \\ &= [A_1^T G_{y,1}^{-1} A_1 + A_2^T G_{y,2}^{-1} A_2]^{-1} (A_1^T G_{y,1}^{-1} \mathbf{y}_1 + A_2^T G_{y,2}^{-1} \mathbf{y}_2) \\ &= G_{x,t} G_{x,1}^{-1} \hat{\mathbf{x}}_1 + G_{x,t} A_2^T G_{y,2}^{-1} \mathbf{y}_2 \end{aligned} \tag{B6}$$

where

$$A_t = \begin{bmatrix} A_1 \\ A_2 \end{bmatrix} \tag{B7}$$

and

$$G_{x,t} = [A_1^T G_{y,1}^{-1} A_1 + A_2^T G_{y,2}^{-1} A_2]^{-1} \tag{B8}$$

The difference between the parameter estimates $\hat{\mathbf{x}}_1$ and $\hat{\mathbf{x}}_t$ will be denoted $\Delta\hat{\mathbf{x}}$. The covariance matrix $G_{\Delta x}$ of $\Delta\hat{\mathbf{x}}$ is

$$\begin{aligned} G_{\Delta x} &= E[\Delta\hat{\mathbf{x}}\ \Delta\hat{\mathbf{x}}^T] \\ &= E[((G_{x,t} G_{x,1}^{-1} - I)\hat{\mathbf{x}}_1 + G_{x,t} A_2^T G_{y,2}^{-1} \mathbf{y}_2) \times \text{transpose}] \end{aligned} \tag{B9}$$

where I is the identity matrix.

Since $\mathbf{y}_1$ and $\mathbf{y}_2$ are uncorrelated, we have

$$E[\hat{\mathbf{x}}_1 \mathbf{y}_2^T] = E[\mathbf{y}_2 \hat{\mathbf{x}}_1^T] = 0 \tag{B10}$$

and therefore

$$\begin{aligned} G_{\Delta x} &= (G_{x,t} G_{x,1}^{-1} - I) G_{x,1} (G_{x,1}^{-1} G_{x,t} - I) \\ &\quad + G_{x,t} A_2^T G_{y,2}^{-1} G_{y,2} G_{y,2}^{-1} A_2 G_{x,t} \end{aligned} \tag{B11}$$

where we have used the fact that a covariance matrix is symmetric. Algebraic manipulation of (B11) yields

$$\begin{aligned} G_{\Delta x} &= G_{x,t} G_{x,1}^{-1} G_{x,t} + G_{x,1} - 2G_{x,t} \\ &\quad + G_{x,t} A_2^T G_{y,2}^{-1} A_2 G_{x,t} \end{aligned} \tag{B12}$$

From (B3) and (B8) it can be seen that

$$A_2^T G_{y,2}^{-1} A_2 = G_{x,t}^{-1} - G_{x,1}^{-1} \tag{B13}$$

Substitution of (B13) into (B12) and cancellation yield

$$G_{\Delta x} = G_{x,1} - G_{x,t} \tag{B14}$$

In terms of this paper, $\mathbf{y}_1$ would be composed of the VLBI observations from elevations above the elevation angle cutoff, while $\mathbf{y}_2$ would be composed of observations from below this cutoff in elevation. The vector $\hat{\mathbf{x}}_1$ is the least squares estimate of $\mathbf{x}$ resulting from the observations $\mathbf{y}_1$, while $\hat{\mathbf{x}}_t$ results from using all the data (both $\mathbf{y}_1$ and $\mathbf{y}_2$). From (B14) it can be seen that the covariance matrix of the difference between $\hat{\mathbf{x}}_1$ and $\hat{\mathbf{x}}_t$ is the difference of their respective covariance matrices.

Acknowledgments. This work was supported by Air Force Geophysics Laboratory contract F19628-83-K-0031, NASA contract NAS5-27571, and NSF grants EAR-83-02221 and EAR-83-06380.

REFERENCES

Bean, B. R., B. A. Cahoon, C. A. Samson, and G. D. Thayer, *A World Atlas of Atmospheric Radio Refractivity*, *ESSA Monogr.*, vol. 1, U.S. Government Printing Office, Washington, D. C., 1966.

Berman, A. L., The prediction of zenith refraction from surface measurements of meteorological parameters, *Rep. JPL TR 32-1602*, Calif. Inst. of Technol. Jet Propul. Lab., Pasadena, Calif., 1976.

Birnbaum, G., and S. K. Chatterjee, The dielectric constant of water vapor in the microwave region, *J. Appl. Phys.*, *23*, 220–223, 1952.

Black, H. D., An easily implemented algorithm for the tropospheric range correction, *J. Geophys. Res.*, *83*, 1825–1828, 1978.

Black, H. D., and A. Eisner, Correcting satellite Doppler data for tropospheric effects, *J. Geophys. Res.*, *89*, 2616–2626, 1984.

Born, M., and E. Wolf, *Principles of Optics*, 4th ed., Pergamon, New York, 1970.

Boudouris, G., On the index of refraction of air, the absorption and dispersion of centimeter waves by gasses, *J. Res. Natl. Bur Stand.*, *67D*, 631–684, 1963.

Chao, C. C., A model for tropospheric calibration from daily surface and radiosonde balloon measurements, *Tech. Memo Calif. Inst. Technol. Jet Propul. Lab.*, *391-350*, 17 pp., 1972.

Clark, T. A., et al., Precision geodesy using the MkIII very-long-baseline interferometer system, *IEEE Trans. Geosci. Remote Sens.*, *GE-23*, 438–449, 1985.

CRC Press, *Handbook of Chemistry and Physics*, 55th ed., pp. B-1 and F-223, Boca Raton, Fla., 1974.

Fanselow, J. L., Observation model and parameter partials for the JPL VLBL parameter estimation software "MASTERFIT-V1.0," *JPL Publ.*, *83-39*, 54 pp., 1983.

Fleagle, R. G., and J. A. Businger, *An Introduction to Atmospheric Physics*, pp. 15–16, Academic, Orlando, Fla., 1980.

Gardner, C. S., Correction of laser tracking data for the effects of horizontal refractivity gradients, *Appl. Opt.*, *16*, 2427–2432, 1977.

Glueckauf, E., The composition of the atmosphere, in *Compendium of Meteorology*, edited by T. F. Malone, pp. 3–10, American Meteorological Society, Boston, Mass., 1951.

Hill, R. J., R. S. Lawrence, and J. T. Priestly, Theoretical and calculational aspects of the radio refractive index of water vapor, *Radio Sci.*, *17*, 1251–1257, 1982.

Hopfield, H S., Two-quartic tropospheric refractivity profile for correcting satellite data, *J. Geophys. Res.*, *74*(18), 4487–4499, 1969.

Marini, J W., Correction of satellite tracking data for an arbitrary tropospheric profile, *Radio Sci.*, 7, 223–231, 1972.

Owens, J. C., Optical refractive index of air: Dependence on pressure, temperature, and composition, *Appl. Opt.*, *6*, 51–58, 1967.

Resch, G. M., Inversion algorithm for water vapor radiometers operating at 20.7 and 31.4 GHz, *JPL TDA Progr. Rep. 42-76*, Calif. Inst. of Technol. Jet Propul Lab., Pasadena, Calif., 1983.

Resch, G. M., Water vapor radiometry in geodetic applications, in *Geodetic Refraction*, edited by F. K. Brunner, Springer-Verlag, New York, 1984.

Robertson, D. S., Geodetic and astrometric measurement with very-long-baseline interferometry, Ph.D. thesis, 186 pp., Mass. Inst. of Technol., Cambridge, 1975.

Robertson, D. S., and W E Carter, Earth rotation information derived from MERIT and POLARIS observations, in *High Precision Earth Orientation and Earth-Moon Dynamics*, edited by O. Calame, pp 97–122, D. Reidel, Hingham, Mass., 1982.

Saastamoinen, J., Atmospheric correction for the troposphere and stratosphere in radio ranging of satellites, in *The Use of Artificial Satellites for Geodesy*, *Geophys. Monogr. Ser.*, vol. 15, edited by S. W. Henriksen et al., pp. 247–251, AGU, Washington, D. C., 1972.

Smith, O. E., W. M. McMurray, and H. L. Crutcher, Cross sections of temperature, pressure, and density near the 80th meridian west, *Rep. NASA TND-1641*, 1963.

Thayer, G. D., An improved equation for the radio refractive index of air, *Radio Sci.*, *9*, 803–807, 1974.

Van Vleck, J. H., *The Theory of Electric and Magnetic Susceptibilities*, Oxford University Press, New York, 1965.

J. L. Davis, Department of Earth, Atmospheric, and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139.

G Elgered, Onsala Space Observatory, Chalmers University of Technology, S-43900 Onsala, Sweden

T. A. Herring and I. I. Shapiro, Harvard-Smithsonian Center for Astrophysics, Cambridge, MA 02138.

A. E. E. Rogers, Haystack Observatory, Westford, MA 01886

Appendix B

The small scale horizontal distribution of tropospheric water vapor

Introduction

In this appendix, we develop a model of the small scale (1–10 km) horizontal distribution of water vapor in the troposphere. In the model, we treat the distribution of water as a white-noise process passed through a linear time-invariant filter, whose origin lies in the atmospheric-turbulence field, and whose impulse response is interpreted as a small-scale water-vapor distribution function. We will present data obtained from ground-based water-vapor radiometers, along with an interpretation of those data based on the model, and an analysis based on a statistical description of the tropospheric turbulence field.

The data used in this appendix were originally intended for use in a series of tests designed to study the precision and accuracy of the estimation of the line-of-sight delay due to water vapor by water-vapor radiometers. These instruments were discussed in Chapter 2. The tests consisted of a series of quickly repeated observations in the zenith direction, among other things.

These measurements also provided an opportunity to study the statistical fluctuation of water vapor on a scale from several minutes to several days. The power spectral density of these measurements was already being estimated in conjunction

with the tests mentioned above. The next step was to perform the common academic exercise of treating the measurements as white noise passed through a linear time-invariant (LTI) filter and looking at the filter response.

What was interesting in this case, however, was that the filter response has an obvious, mechanistic, and simple interpretation in terms of the atmospheric turbulence field. To see this, one need only imagine an initially dry atmosphere into which is randomly scattered a series of point-blobs of water vapor. The turbulent motion of the atmosphere would tend to smear out these blobs, acting in fact like a low-pass filter.

This appendix begins by reviewing stochastic processes, LTI systems, and atmospheric turbulence theory. The expression for the filter response is then parameterized in terms of observables, and the results are presented. Also included is the modeling of the input as shot noise, and a comparison of the results to Kolmogorov turbulence and to other results.

B.1 Stochastic processes and linear time-invariant systems†

In preparation for what follows, a very brief review of stochastic processes and linear time-invariant systems is included here. A stochastic, or random, process is defined by an event space Ω and a probability mapping function Pr( ). For every $\omega \in \Omega$, Pr(ω) is the probability that ω occurs. We now associate with every $\omega \in \Omega$ a

† A standard reference for the material contained in this section is Davenport and Root [1958].

waveform $x(t,\omega)$ defined in $-\infty < t < +\infty$, abbreviated simply as $x(t)$. The stochastic process $x(t)$ has the following properties:

- For ω given, $x(t)$ is a deterministic waveform.
- For t fixed, $x(t)$ is a random variable with expectation over Ω of $m_x(t)$.
- For ω not given and t viewed as an independent variable, $x(t)$ is a random waveform whose expectation at any time t is $m_x(t)$.

In addition to $m_x(t)$, $x(t)$ in general possesses a second order moment, the autocorrelation $R_x(t,s)$, given by

$$R_x(t,s) = \mathrm{E}[x(t)x(s)] \qquad (B.1.1)$$

where E[ ] indicates expectation over Ω. There are third and higher order moments as well.

A process is wide-sense stationary if and only if (i) $m_x(t) = m_x(0)$ for all t, and (ii) $R_x(t,s) = R_x(t-s,0)$ for all t and s. Note that the property (ii) allows us to write the autocorrelation as a function of $\tau = t - s$ only. Thus

$$R_x(\tau) = \mathrm{E}[x(t+\tau)x(t)] = R_x(-\tau) \qquad (B.1.2)$$

is used in place of $R_x(t,s)$.

The power spectral density $S_x(f)$ of a wide-sense stationary process $x(t)$ is the Fourier transform of the autocorrelation function:

$$S_x(f) = \int_{-\infty}^{\infty} d\tau \; R_x(\tau) \; e^{-2\pi i f \tau} \qquad (B.1.3)$$

The physical interpretation given $S_x(f)$ is that $S_x(f)df$ is the mean square strength of the process $x(t)$ contained in the Fourier components located between the frequencies f and $f + df$. The stochastic process $x(t)$ is white noise if

$$R_x(\tau) = q\delta(\tau) \tag{B.1.4}$$

where $\delta(\tau)$ is the delta function and q is a constant. Here q is called the spectral height of x, since

$$S_x(f) = q \quad -\infty < f < \infty \tag{B.1.5}$$

It is clear that white noise cannot truly be said to exist, for a process being a white noise process implies that the total expected power is infinite. There are many instances, however, where a process can be modeled as white noise if the spectrum is flat over a desired finite band of frequencies.

We now turn to the subject of linear systems. We will consider a waveform $x(t)$ (not necessarily random) that is passed as input into some system whose resulting output is $y(t)$. The system is linear if $y(t)$ can be written as the convolution of some function $h(t,\tau)$ and the input $x(t)$:

$$y(t) = \int_{-\infty}^{\infty} d\tau\ h(t,\tau)x(\tau) \tag{B.1.6}$$

The system is time-invariant if $h(t,\tau) = h(t-\tau)$. Then

$$y(t) = \int_{-\infty}^{\infty} d\tau\ h(t-\tau)x(\tau) \tag{B.1.7}$$

The function $h(t)$ is known as the impulse response of the system, for if $x(t) = \delta(t)$ we have

$$y(t) = \int_{-\infty}^{\infty} d\tau\ h(t-\tau)\ \delta(\tau) = h(t) \tag{B.1.8}$$

The Fourier transform of the impulse response yields $H(f)$, the transfer function of the system:

$$H(f) = \int_{-\infty}^{\infty} dt\, h(t)\, e^{-2\pi i f t} \qquad (B.1.9)$$

Let us now suppose that $x(t)$ is a wide-sense stationary process with mean m_x and autocorrelation $R_x(\tau)$. Then $y(t)$ will also be a wide-sense stationary process with mean

$$m_y = m_x \int_{-\infty}^{\infty} dt\, h(t) \qquad (B.1.10)$$

and autocorrelation

$$R_y(\tau) = \int_{-\infty}^{\infty} dr \int_{-\infty}^{\infty} ds\, h(r)\, h(s)\, R_x(\tau - r + s) \qquad (B.1.11)$$

The power spectral density $S_y(f)$ of $y(t)$ is by application of the convolution theorem [Bracewell, 1978]

$$S_y(f) = |H(f)|^2 S_x(f) \qquad (B.1.12)$$

Equation (B.1.12) will play a major role in the model developed in Section B.3.

B.2 Statistical description of atmospheric turbulence†

In this section a statistical description of the turbulence field of the atmosphere will be developed. In theory, it should be possible to derive a deterministic expression for the velocity field $\mathbf{u}(\mathbf{x})$ via the Navier-Stokes equation for an incompressible fluid (see Vinnischenko *et al.* [1980]). This solution requires, however, a set of initial conditions which are not readily available. Instead, $\mathbf{u}(\mathbf{x})$ is treated as a random field,

† The bulk of Section B.2 is covered in detail by Lumley and Panofsky [1964].

defined in the same way as the random function of the previous section, except that we now have a three-dimensional function $\mathbf{u}$ as well as a three-dimensional independent variable $\mathbf{x} = (x, y, z)$. Analogous to the one-dimensional autocorrelation function is the correlation tensor $R_{jk}(\mathbf{r})$ given by

$$R_{jk}(\mathbf{x}, \mathbf{r}) = \mathrm{E}[u_j(\mathbf{x} + \mathbf{r}) u_k(\mathbf{x})] \qquad (B.2.1)$$

If $\mathbf{u}(\mathbf{x})$ is wide sense stationary, then $R_{jk}(\mathbf{x}, \mathbf{r}) = R_{jk}(\mathbf{r})$, and we can define a three-dimensional power spectral density tensor $S_{jk}(\mathbf{k})$ given by

$$S_{jk}(\mathbf{k}) = \frac{1}{(2\pi)^3} \int d\mathbf{r}\, R_{jk}(\mathbf{r}) e^{-i\mathbf{k}\cdot\mathbf{r}} \qquad (B.2.2)$$

A random vector field which is stationary is called homogeneous, and appropriately so: the correlation between values of $\mathbf{u}$ at any two points separated by the vector $\mathbf{r}$ depends only on $\mathbf{r}$. (If it depends only on $|\mathbf{r}|$, then the field is also isotropic.) In reality, it is very difficult to imagine that the atmospheric turbulence field is actually homogeneous; it should be clear, for instance, that the mean velocity vector is not a constant over the surface of the earth. Instead, one can introduce a three-dimensional structure function $D_j(\mathbf{r})$, based on the assumption that $\mathbf{u}(\mathbf{x})$ is a process with stationary independent increments [Monin and Yaglom, 1975]:

$$D(\mathbf{r}) = \mathrm{E}[(u_j(\mathbf{x} + \mathbf{r}) - u_j(\mathbf{x}))^2] \qquad (B.2.3)$$

Tatarskii [1961] has stated that the assumption that $\mathbf{u}(\mathbf{x})$ is a stationary independent increments process is correct for $|\mathbf{r}|$ "not too large." On the basis of this stationarity,

the structure function can be related to the diagonal components of the power spectral density tensor via

$$D_j(\mathbf{r}) = 2\int d\mathbf{k}\ \left(1 - e^{i\mathbf{k}\cdot\mathbf{r}}\right)\ S_{jj}(\mathbf{k}) \qquad (B.2.4)$$

Henceforth all the subscripts will be omitted on the assumption that all of the components of $\mathbf{u}$ obey the same statistics. Thus we write simply

$$D(\mathbf{r}) = 2\int d\mathbf{k}\ \left(1 - e^{i\mathbf{k}\cdot\mathbf{r}}\right)\ S(\mathbf{k}) \qquad (B.2.5)$$

It is usual to write the structure function in the plane z = constant in terms of a two-dimensional power-spectral density $S_2(k_x, k_y)$:

$$D(x,y,0) = 2\int_{-\infty}^{\infty} dk_x \int_{-\infty}^{\infty} dk_y\ \left(1 - e^{i[k_x x + k_y y]}\right) S_2(k_x, k_y) \qquad (B.2.6)$$

where

$$S_2(k_x, k_y) = \frac{1}{2\pi}\int dk_z\ S(k_x, k_y, k_z) \qquad (B.2.7)$$

If the atmosphere is isotropic, then (B.2.6) can be written in terms of a Hankel transform [Bracewell, 1978] as

$$D(r) = 4\pi\int_0^{\infty} dk_r\ k_r\ [1 - J_\circ(k_r r)]\, S_2(k_r) \qquad (B.2.8)$$

Finally, one can compute the one-dimensional power spectral density. The transform, analogous to (B.2.7), is

$$S_1(k_x) = \int dk_y S_2(k_x, k_y) \qquad (B.2.9)$$

It is convenient at this point to convert spatial variations to temporal variations, as measurements are often obtained by observing the atmosphere in one direction for a

long time. Thus, if we assume that there exists a unique constant mean velocity $\mathbf{m}_v$ at which the "frozen" features of the atmosphere drift, then we have

$$\mathbf{x}(t) = \mathbf{x}(t_o) + \mathbf{m}_v \times (t - t_o) \qquad (B.2.10)$$

This assumption is known as Taylor's frozen turbulence field hypothesis [Taylor, 1938], and has been experimentally verified for scales of several kilometers, but not greater. The one-dimensional temporal power spectrum corresponding to (B.2.9) is

$$S(f) = \frac{2\pi}{m_{vx}} S_1(k_x)\Big|_{k_x = \frac{2\pi f}{m_{vx}}} \qquad (B.2.11)$$

Any attempt to model the spatial dependence of m_v would result in complications which are beyond the scope of this limited study. Armstrong and Sramek [1982] have investigated the case for which the atmosphere is treated as a series of layers in which each layer possesses a uniform probability density for the mean velocity. Their results indicate that one can expect differences in the observed one-dimensional power spectrum of no more than about 10% compared to the constant mean velocity case.

B.3 The model

The basic assumption of the model presented here is that the total amount of water vapor in the atmosphere is due to the sum of a number of individual small scale ($\sim$ several km) conglomerates of water vapor, distributed horizontally. In each of these conglomerates the horizontal distribution of water vapor is governed by an identical "shape function." We will consider only a one-dimensional sky here. This shape function, henceforth denoted by $h(x)$, is assumed to be symmetric about its

center, and normalized so that $\int dx\ h(x) = 1$ Thus, if the total amount of water vapor in one of these conglomerates is L, and the conglomerate is centered at $x = x_\circ$, then the amount of water vapor located between x and $x + dx$ is $Lh(x - x_\circ)dx$. For convenience, we will measure L in units of centimeters of delay for a light ray traveling from infinity in the zenith direction.

If, as previously discussed, there are a number of these conglomerates distributed at $x_1, x_2, \ldots$, and if the total delay in each of the conglomerates is $L_1, L_2, \ldots$, respectively, then the delay $L(x)$ at the point x due to the contribution of each of the conglomerates, averaged over the length Δx, is is

$$\begin{aligned} L(x) &= \Delta x \sum_j L_j h(x - x_j) \\ &= \Delta x \int_{-\infty}^{\infty} dx'\ p(x')\ h(x - x') \end{aligned} \tag{B.3.1}$$

where

$$p(x) = \sum_j L_j\ \delta(x - x_j) \tag{B.3.2}$$

Henceforth, instead of speaking of horizontal distance in terms of x, it will be in terms of time t, as per Section B.2. The time Δt will become the time between observations (approximately 3 minutes). Thus, (B.3.1) becomes

$$L(t) = \Delta t \int_{-\infty}^{\infty} d\tau\ p(\tau)\ h(t - \tau) \tag{B.3.3}$$

Comparing this equation to (B.1.7), we can see that it is possible to view $L(t)$ as the output from a linear time-invariant filter whose input is $p(t)$ and whose impulse response is $h(t)$. The power spectrum of $L(t)$ is from (B.1.12) given by

$$S_L(f) = |H(f)|^2 S_p(f) \tag{B.3.4}$$

where $S_p(f)$ is the power spectrum of $p(t)$.

If $S_p(f)$ were known, then it would be possible to perform an experiment and obtain the form of $H(f)$, and therefore $h(t)$. Unfortunately, $S_p(f)$ is not known, and it is even doubtful that $p(t)$ is stationary. However, since we are working here on somewhat short time scales, we will make the assumption that $p(t)$ is a white-noise process, so by (B.1.5) we have

$$S_p(f) = q_p \qquad (B.3.5)$$

and an estimate of $S_L(f)$ leads straightforwardly to $H(f)$.

The assumption that $p(t)$ is white is probably not too bad considering that $L(t)$ is nearly an independent increments process (see Section B.2). This assumption implies that $p(t)$ might, for example, be modeled as shot noise [Davenport and Root, 1958] with a constant average arrival rate (see Section B.6).

B.4 The observations

The observations reported here were obtained on 26 and 27 October 1982 from a ground-based water-vapor radiometer (WVR) located at the George R. Agassiz Station (formerly Harvard Radio Astronomy Station) located near Ft. Davis, Texas. The calibration and operation of the WVR was discussed in Chapter 2, and the reduction of data in Chapter 4. In an attempt to achieve a high density of data, measurements

were made at a rate of one observation every three minutes, continuing for nearly forty-six hours before an instrumental malfunction occurred, aborting the experiment.

The data thus obtained have been separated into two twenty-three hour long data sets, shown in Figure B.4.1. The average delay for the first data set (henceforth referred to as Day 1) was 11.0 cm, and the data show relatively little variation for the first twenty-two hours. The data of the second set (Day 2) have a mean delay of 5.9 cm, and display short-term variations with amplitudes of several centimeters. (These means were subtracted from the data for the following analysis.) For both days, the weather was mild with light winds, and the sky was partly overcast. The relative humidity was under 10% and the barometric pressure hovered near 850 mbar.

The spectra of the data are shown in Figure B.4.2. These spectra were estimated first by forming the discrete Fourier transform (DFT) of the data, given by [Bracewell, 1978]:

$$X(f) = \sum_{j=0}^{N-1} x(j\Delta t) e^{-2\pi i f \Delta t} \Delta t \qquad (B.4.1)$$

In practice, the Fast Fourier Transform (FFT) routine FOURG [Gorden, 1976] was used to obtain estimates of $X(f)$ only at the frequencies $f = k\Delta f$, $k = 0, \ldots, N-1$, where $\Delta f = \frac{1}{N\Delta t}$, so that (B.4.1) becomes

$$X(k) = \frac{1}{\sqrt{N}} \sum_{j=0}^{N-1} x(j) e^{-2\pi i j k/N} \qquad (B.4.2)$$

An estimate of the power spectral density $S_x(k)$ at the discrete frequencies indicated above is then given by

$$S_x(k) = \frac{1}{N\Delta f} |X(k)|^2 \qquad (B.4.3)$$

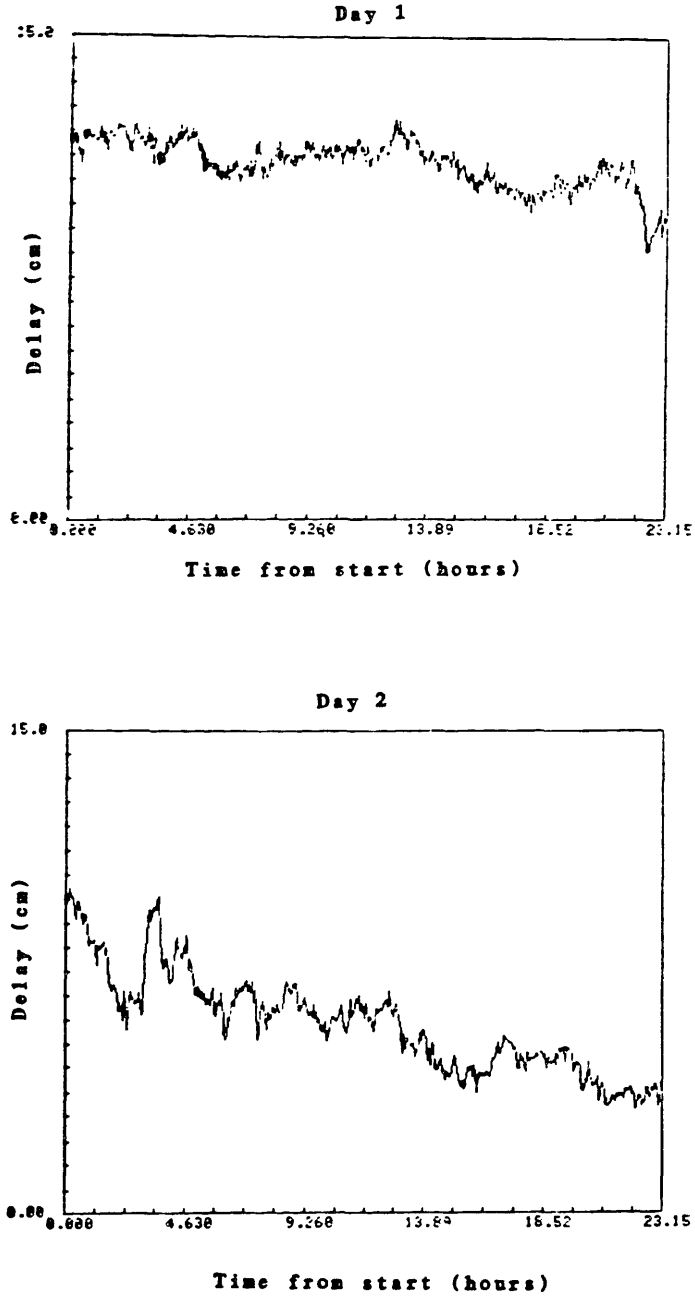


Figure B.4.1. Unsmoothed zenith delays.

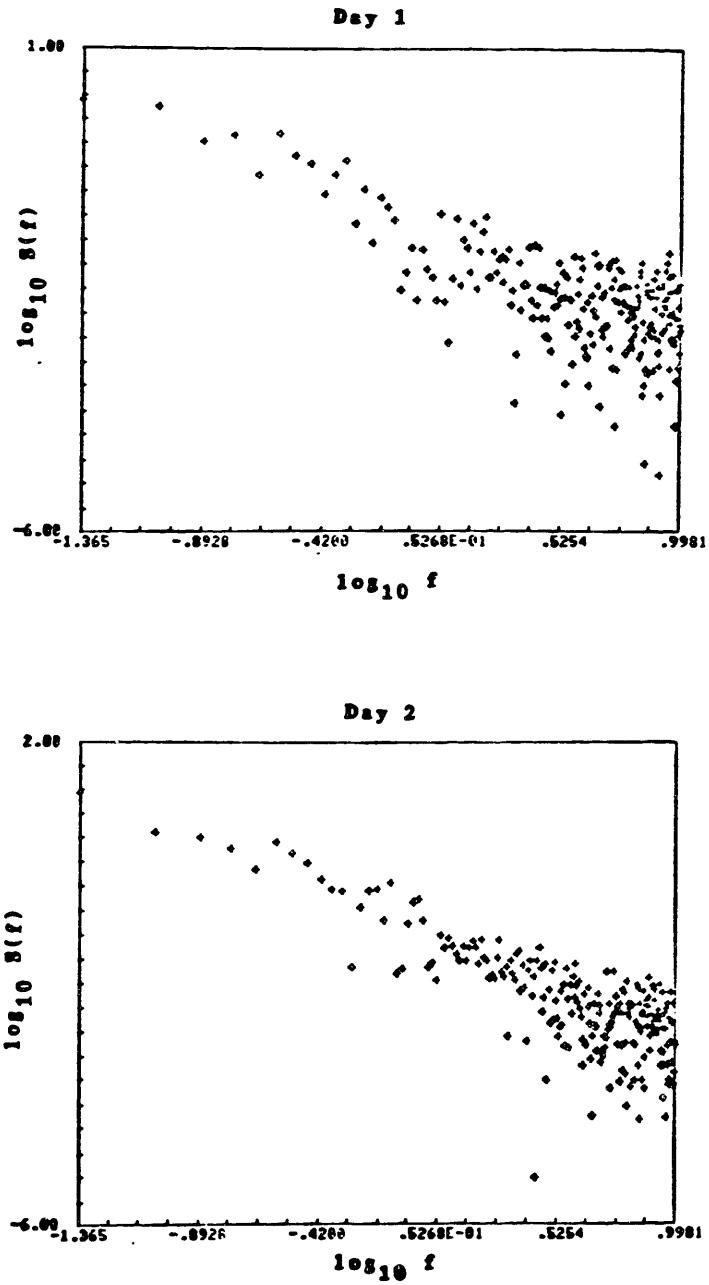


Figure B.4.2. Power spectra of data in Figure B.4.1. The units of spectra are square centimeters per cycle per hour. The units of frequency are cycles per hour. The standard deviation of the estimates is approximately 0.13 cm^2 per cycle per hour on Day 1 and 0.38 cm^2 per cycle per hour on day 2.

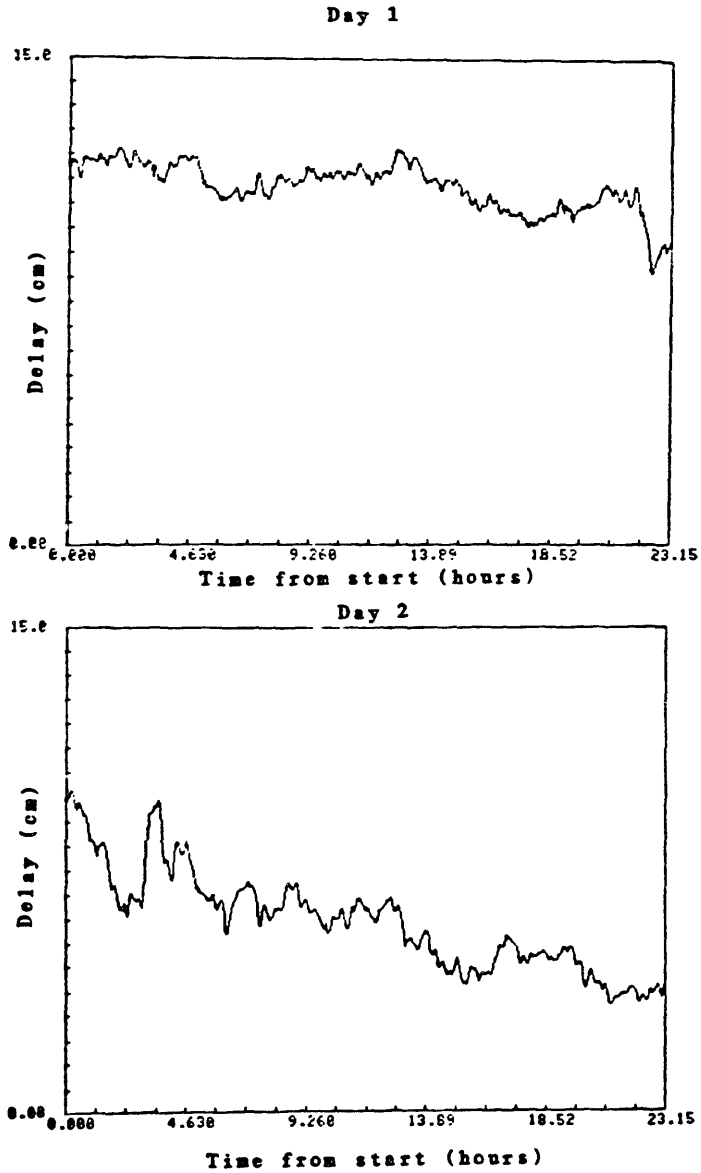


Figure B.4.3. Data of Figure B.4.1 smoothed with a Gaussian filter of FWHM = 9 minutes.

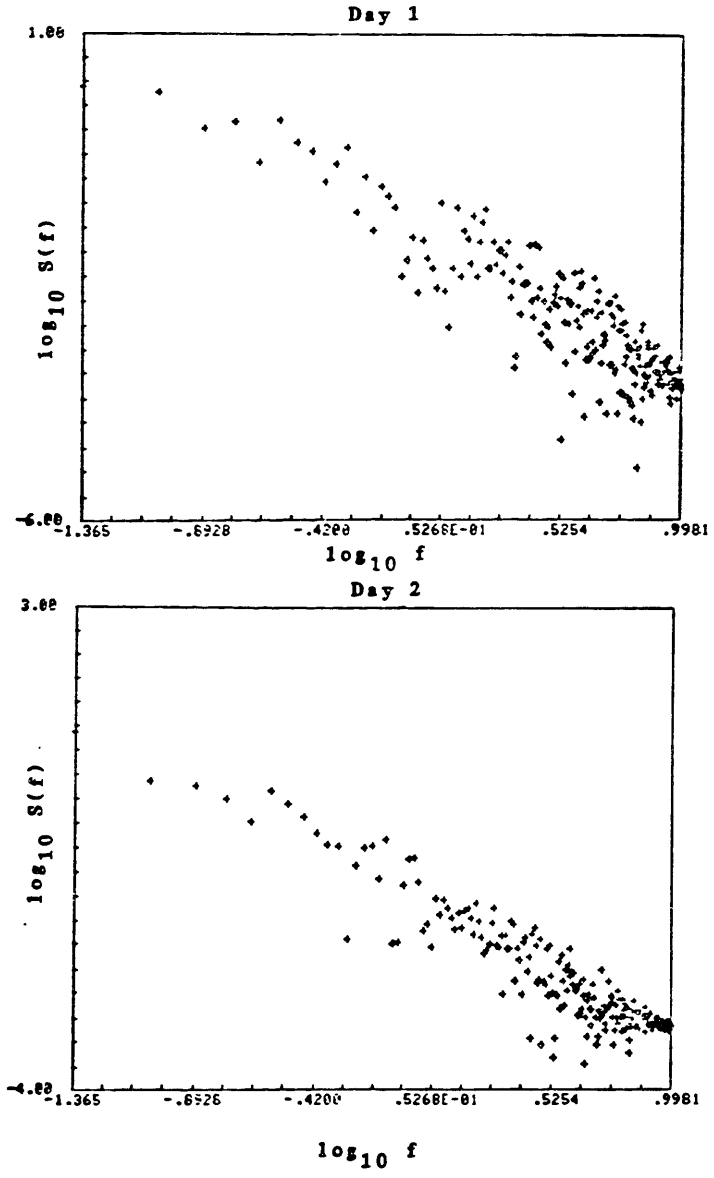


Figure B.4.4. Power spectra of data in Figure B.4.3. All units are the same as those of Figure B.4.2. The standard deviation of the estimates is approximately 0.08 cm^2 per cycle per hour on Day 1 and 0.22 cm^2 per cycle per hour on day 2.

Table B.4.1

Summary of powerlaw fits

Day	$A \times 10^{-10}$	$\beta - 1$	β
1	0.68±0.07	2.205 ± 0.010	3.205 ± 0.010
2	11.4±0.6	2.033 ± 0.005	3.033 ± 0.005

Summary of least squares fit of power spectrum of Figure B.4.4 to $S_L(f) = Af^{1-\beta}$. The model was linearized prior to fit by taking the logarithm. All units are MKS. The uncertainties given are the estimates of the formal standard deviation resulting from the fit, assuming a standard deviation of 1 mm (post-smoothed) for the estimates of the zenith delay. This standard deviation is based on the known magnitude of instrumental fluctuations.

The estimator in (B.4.3) is neither an unbiased nor an efficient estimator.

In handling measurements of this type it is usual to smooth the data either prior to taking the DFT to act as a low pass filter, or to smooth the spectral estimates, thereby alleviating edge effects [Jenkins and Watt, 1968]. Filtering, however, while decreasing the random error in each measurement, introduces correlations and decreases the frequency resolution, thereby having a large effect on the results presented in the next section. Therefore, the data were smoothed using a Gaussian filter, but it was decided to use a FWHM of only 9 minutes. Thus, only three data points are within the main lobe of the weighting function. This is roughly the lower limit of the smoothing window usually employed to smooth these types of data. The smoothed versions of Figures B.4.1 and B.4.2 are shown in Figures B.4.3 and B.4.4, and a summary of the powerlaw fit to $S_L(f) = Af^{1-\beta}$ of the power spectrum of Figure B.4.4 is contained in Table B.4.1.

B.5 Analysis of the observations

A parametrization for the shape function $h(t)$ described in Section B.3 is possible. It has been found [Armstrong and Sramek, 1982] that the two-dimensional phase structure function—analogous to the velocity structure function of (B.2.3)—obeys

$$D(r) = K\, r^{\beta-2} \qquad (B.5.1)$$

where $\beta > 2$ and $K > 0$. Conversion to the structure function of interest to us is obtained simply by multiplication of (B.5.1) by the constant $(c/\omega)^2$, where c is the

speed of light, and ω is the angular frequency to which (B.5.1) is referred. The two-dimensional delay structure function $D_L(r)$ will therefore be written as

$$D_L(r) = K_L\, r^{\beta-2} \tag{B.5.2}$$

with $K_L = (c/\omega)^2 K$. The two-dimensional power spectral density $S_L(k_r)$ is given by (B.2.8). This integral can be evaluated using the integral found on p. 130 of Panchev [1971]:

$$\int_0^\infty dx\, |x|^{-n}\, [1 - J_\circ(x)] = \frac{\pi}{2^{-n}\Gamma^2\left(\frac{1+n}{2}\right)\sin(n-1)\frac{\pi}{2}} \tag{B.5.3}$$

for $1 < n < 3$. Applying this to (B.2.8) yields for the two-dimensional spectral density of L the expression

$$S_{L2} = \frac{-K_L\Gamma^2\left(\frac{\beta}{2}\right)\sin\left(\beta\frac{\pi}{2}\right)}{4\pi^2 2^{1-\beta}} k_r^{-\beta} \tag{B.5.4}$$

where $\Gamma(x)$ is the gamma function. The one-dimensional spatial delay power spectrum S_L can be obtained from (B.2.9) and the following integral identity from the CRC Standard Mathematical Tables (no. 615):

$$\int_0^\infty dx\, \frac{x^a}{(m+x^b)^c} = \frac{m^{\frac{a+1-bc}{b}}}{b}\,\frac{\Gamma\left(\frac{a+1}{b}\right)\Gamma\left(c-\frac{a+1}{b}\right)}{\Gamma(c)} \tag{B.5.5}$$

for $a > -1$, $b > 0$, $m > 0$, and $c > \frac{a+1}{b}$. We therefore find, after transforming to the temporal power spectrum as per (B.2.11), that

$$S_L = -2K_L\Gamma\left(\frac{\beta}{2}\right)\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{\beta-1}{2}\right)\left(\sin\beta\frac{\pi}{2}\right)(2\pi)^{\beta-2}(2M_{vx})^{\beta-2} f^{1-\beta} \tag{B.5.6}$$

Since we are assuming that the shape function $h(t)$ is even, then $|H(f)|^2 = H^2(f)$, and from (B.3.4) we have

$$H(f) = \sqrt{q_p} B(\beta) f^{\frac{1-\beta}{2}} \qquad (B.5.7)$$

where the constants in (B.5.6) have been absorbed by $B(\beta)$. To obtain an expression for $h(t)$, we must use the identity from Tatarskii [1961], p. 269:

$$\int_{-\infty}^{\infty} dx\, |x|^{-n}(1 - \cos ax) = \frac{\pi a^{n-1}}{\Gamma(n)\sin(n-1)\frac{\pi}{2}} \qquad (B.5.8)$$

for $1 < n < 3$. We therefore obtain for $h(t)$:

$$h(t) = h(0) - \frac{2\pi B(\beta)(2\pi)^{\frac{\beta-3}{2}}\sqrt{q_p}}{\Gamma\left(\frac{\beta-1}{2}\right)\sin\frac{\beta-3}{4}\pi}|t|^{\frac{\beta-3}{2}} \qquad (B.5.9)$$

The region over which (B.5.1) is valid is known as the inertial subrange, and extends to several tens of kilometers, although this "outer scale" has never been experimentally determined. Depending on the mean velocity, we would therefore expect (B.5.9) to be valid on time scales of 0.5 minutes to several hours. As previously mentioned, the estimate of $h(t)$ is obtained by back-transforming the complex absolute value of the estimated Fourier transform of the delay. In this sense, $h(t)$ closely resembles an estimate of the autocorrelation function. In general, it is unwise to attempt to estimate the autocorrelation function at values of the lag which are greater than about 10% of the total period during which data were taken, due to the decrease in the number of overlapping data at the greater lags. Applying this reasoning to $h(t)$, then, it was decided to estimate $h(t)$ only for $|t| < 2.3$ hours. Figure B.5.1 contains the resulting shape functions (solid line), along with the least squares fits (dashed line).

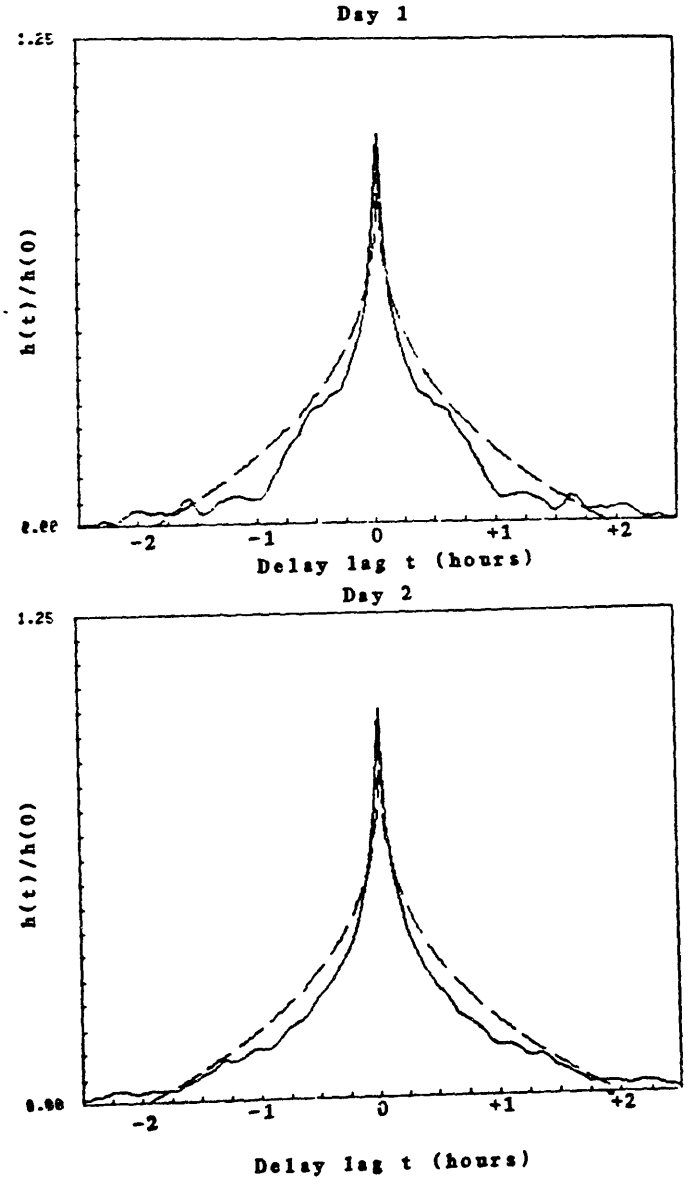


Figure B.5.1. Shape functions. Solid lines result from back-transforming $H(f)$ (see text). Dashed lines are plots of (B.5.7) using the parametrization resulting from the least squares fit as described in Section B.5.

The fits were obtained by normalizing the back-transformed amplitudes to give a value of unity at $t = 0$. The resulting values were then fit in a least squares sense to (B.5.9) in the form

$$\frac{h(t)}{h(0)} = 1 - \frac{A}{h(0)}|t|^{\frac{\beta-3}{2}} \qquad (B.5.10)$$

The estimates of $h(0)$ were obtained by requiring that

$$\int_{-T}^{T} dt\, h(t) = 1 \qquad (B.5.11)$$

for some integration time T. In this case, T was chosen to be 10% of the total time, or 2.3 hours, as previously discussed. This limit seems reasonable, since from Figure B.5.1 it can be seen that $h(t)$ is nearly zero beyond this time.

Although it is possible to fit the data in a least squares sense directly to (B.5.8) with (B.5.11) as a constraint, there are practical considerations which make the process described above more desirable: a linearized least squares fit to (B.5.8) requires derivatives of $h(t)$ with respect to β; thesis derivatives (which depend on $\log|t|$) do not exist for $t = 0$. It was found that the exclusion of the $t = 0$ point allowed a solution for which $\beta < 3$ and $q_p < 0$. Fitting the data to (B.5.10), however, "anchored" the solution at $t = 0$ to a finite value, allowing an unconstrained linearized least squares fit to be done. The "true" value of $h(0)$ was then determined from (B.5.11).

The price paid for this method, however, can be seen in Figure B.5.1: a systematic bias exists between the data and the least squares fit. Although there are alternatives to fitting the data to (B.5.10), none has yet been tried. For instance, it is possible to use a Bayesian approach and fit the data to

$$\frac{h(t)}{h(0)} = \delta - \frac{A}{h(0)}|t|^{\frac{\beta-3}{2}} \qquad (B.5.12)$$

Table B.5.1

Estimated shape function parameters

Day	$h(0) \times 10^4$	$q_p \times 10^{-5}$	$\frac{\beta-3}{2}$	β
1	3.8 ± 0.2	0.69 ± 0.38	0.31 ± 0.04	3.62 ± 0.06
2	3.6 ± 0.2	9.7 ± 3.3	0.30 ± 0.03	3.60 ± 0.04

Summary of least squares fit to (B.5.10) of data shown by solid line in Figure B.5.1. The units of $h(0)$ are sec^{-1}. The units of q_p are $\text{cm}^2\ \text{Hz}^{-1}$. The uncertainties given are the estimates of the formal standard deviations resulting from the fit, with the uncertainty of $h(t)$ estimated from the postfit RMS residual. This residual was 2.8×10^{-5} sec^{-1} on Day 1 and 1.7×10^{-5} sec^{-1} on Day 2.

where the *a priori* variance on the parameter δ is assumed small enough that δ remains in the neighborhood of unity.

The values of q_p and $h(0)$ resulting from the fit to (B.5.10) are included in Table B.5.1. The value of q_p was obtained by first estimating K_L from the least squares fit to the power spectrum and from (B.5.4). All units are MKS, and m_{vx} has been taken to be 1 m s^{-1}, so the results can be easily scaled (see (B.5.5)). For both days, the values obtained for β are higher than those found via the powerlaw fit to $S_L(f)$. This is possibly due to an error in the assumption that $p(t)$ is white noise. Based on the comparison with other studies (see Section B.5.7), it seems that the estimates derived from the powerlaw fit are closer to the "true,' values.

The large differences between the estimates q_p for the two days is expected, since the fluctuations recorded for the second day are so much greater (see Figure B.4.3). The variance of the delay on Day 1 was approximately 0.54 cm^2, whereas for Day 2 the variance was approximately 2.5 cm^2.

B.6 Modeling the input: An example

So far, we have avoided modeling the input $p(t)$ of (B.3.3), except to assume that $p(t)$ is a white noise process. In this section we explore the implications of modeling $p(t)$ as shot noise as suggested in Section B.3. In this model, the L_j in (B.3.2) are assumed to be a constant, L_o. Shot noise is defined by assuming that the number of impulses in non-overlapping time intervals are a set of statistically independent random variables, and that the probability of having k impulses for $r \leq t < s$ is given by

$$P(k|r,s) = \frac{1}{k!}\left[\int_r^s dt\ \lambda(t)\right]^k \exp\left[-\int_r^s dt\ \lambda(t)\right] \qquad (B.6.1)$$

where, it can be shown, $\lambda(t)$ is the average rate of arrival of the impulses. In this case, it can be shown [Davenport and Root, 1958] that the expected value of $p(t)$ is given by

$$m_p(t) = L_\circ \lambda(t) \tag{B.6.2}$$

and that the autocorrelation $R_p(t,s)$ is given by

$$R_p(t,s) = L_\circ^2 \, \lambda(t)\delta(t-s) \tag{B.6.3}$$

We will now assume that $\lambda(t) = \lambda$, a constant, so that by (B.1.10) the expected value of $L(t)$ is given by

$$m_L(t) = \lambda L_\circ \Delta t \int_{-\infty}^{\infty} dt \; h(t) = \lambda L_\circ \Delta t \tag{B.6.4}$$

and that the autocorrelation function of $L(t)$ is by (B.1.11) given by

$$R_L(\tau) = (\Delta t)^2 \lambda L_\circ^2 \int_{-\infty}^{\infty} dr \; h(r) \; h(r-\tau) \tag{B.6.5}$$

Although it seems as though the mean and correlation depend on the time between observations, the term Δt at this point is no more than a scale factor for the normalization for $h(t)$ which we chose. As we will see, the estimate of λ will be independent of Δt—under the assumptions of the model. We can now relate (B.6.4) and (B.6.5) to some observable quantities by noting from (B.1.2) that

$$\sigma_L^2 = R_L(0) - m_L^2 \tag{B.6.6}$$

where σ_L^2 is the variance of $L(t)$. Combining (B.6.4), (B.6.5), and (B.6.6) yields

$$\sigma_L^2 = L_\circ^2 \lambda \left[\int_{-\infty}^{\infty} dt \; h^2(t) - \lambda \right] (\Delta t)^2 \tag{B.6.7}$$

At this point, we have two unknowns, L_o and λ since both σ_L and the integral are easily estimated from the data. However, the Fourier transform of (B.6.3) yields the power spectral density $S_L(f) = q_p$, which gives us the relationship $q_p = \lambda L_o^2$. Substitution of this relationship into (B.6.7) gives

$$\lambda = \int_{-\infty}^{\infty} dt\, h^2(t) - \frac{\sigma_L^2}{q_p(\Delta t)^2} \qquad (B.6.8)$$

Note that under the assumptions of our model, λ is independent of Δt.

For the two days the values for λ have been estimated to be approximately 1.4×10^{-4} Hz for Day 1 and 1.3×10^{-4} Hz for Day 2. This means that the average arrival times between impulses are 2.0 hours and 2.2 hours. It is difficult to tell how plausible these values are simply from looking at Figure B.4.3.

It should be stressed that this is only one possible model for $p(t)$. This particular model has the appealing characteristic of having a simple physical interpretation.

B.7 Comparison with other observations and conclusion

It is possible to compare the results found in this work both to theory and to other observations. For fully developed Kolmogorov turbulence [Tatarskii, 1961], one would expect that $\beta = \frac{11}{3}$ at scales at which the turbulence is isotropic, and $\beta = \frac{8}{3}$ at larger scales. For the atmosphere, one can seemingly place an upper bound on the scale of isotropy in terms of the tropospheric height (~10 km). Both the data presented here, and those of Hogg *et al.* [1981], as well as those of Armstrong and Sramek [1982] seem to offer evidence that the upper bound is on this order, or even greater. It is also possible that the differences between the two values for β found on

each day in this study are due to the fact that the β found by the fit to $h(t)$ depend for the most part on delay lags of less than 2.3 hours, or 8.3 km at 1 m s^{-1}, while the powerlaw fit to $S(f)$ is sensitive to changes with periods of up to 23 hours. This could also explain why the values for β from $h(t)$ are very close to $\frac{11}{3}$, while the values from $S(f)$ are slightly lower, having been corrupted by $f^{\frac{8}{3}}$ variations at low frequencies. It is therefore perhaps correct to model $p(t)$ as white noise for $t < 2.3$ hours.

Figure B.7.1 illustrates the range of values for K and β which have been obtained by this and other works. The constant K of (B.5.1) has been referred to a frequency of 5 GHz. Results obtained from the powerlaw fit to $S_L(f)$ are indicated by asterisks those via fit to $h(t)$ by open circles The results obtained by Armstrong and Sramek [1982] are indicated by crosses and that by Dravskikh and Finklestein [1979] by a pounds sign. It can be seen that the values for β found via the fit to $S_L(f)$ are, for these values of K, more consistent with the values from the other works than those found by the fit to $h(t)$.

The correlation evident in Figure B.7.1 between values of β and K is as yet unexplained, although it could indicate the existence of a "mechanism" for the atmospheric fluctuations such as that described here. That the values for K found here and in Dravskikh and Finklestein seem consistently high could quite possibly be due to differences in altitude. (At lower altitudes the total amount of integrated water vapor, to which all these studies are sensitive, is greater, and the power of the fluctuations tends therefore to increase.) The Armstrong and Sramek data were taken at the VLA near Socorro, NM at an altitude of 2124 m, and those of this report near Ft. Davis,

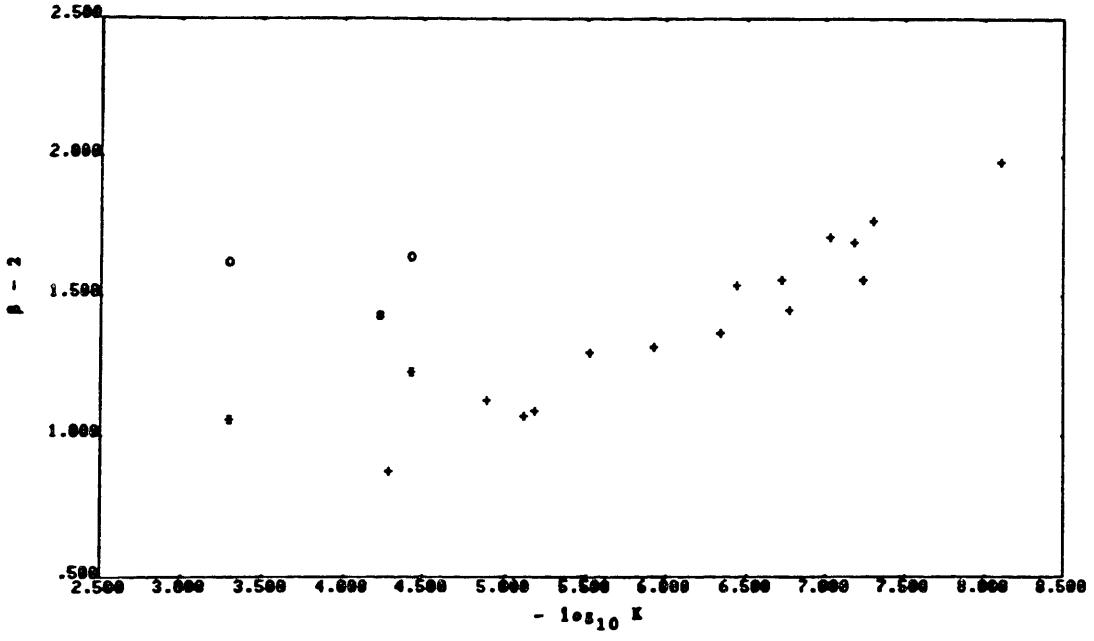


Figure B.7.1. Distribution of β and K values culled from several works. The reference frequency is 5 GHz, the units for K are MKS. The existence of the strong correlation is not well understood. Key: + = Armstrong and Sramek [1982]; # = Dravskikh and Finklestein [1979]; * = this appendix with β from fit to $S(f)$ (Table B.4.1 values); o = this appendix with β from fit to $h(t)$ (Table B.4.2 values).

TX at an altitude of 1603 m. The origin of the Dravskikh and Finklestein data is not known.

In conclusion, two days of estimates of the zenith delay due to the presence of water vapor (the so-called wet delay) have been presented. These data were shown to be consistent with the relationship $S(f) = Kf^{1-\beta}$, with β being within the range of values values found in other works of $2.8 < \beta < 4.0$. A model for the wet delay was developed in which the atmospheric turbulence was treated as a low pass filter with a transfer function of the form $h(t) = h(0) - A|t|^{(\beta-3)/2}$, and whose input is assumed to be white noise of spectral height q_p. This model yields values for β very close to the $\frac{11}{3}$ predicted for Kolmogorov turbulence. Values for q_p obtained were presented although the a physical interpretation of q_p is dependent upon the complete statistical characterization of the input $p(t)$. As an example, we modeled $p(t)$ as shot noise; then q_p is related to the average time between pulses.

An analysis of the type presented in this appendix might be useful both to those who view the turbulence of the wet atmosphere as an error source and to those who are interested in atmospheric and meteorological studies of turbulence as it applies to such subjects as the atmospheric propagation of heat, momentum, and moisture, and atmospheric structure.

References

Armstrong, J.W., R.A. Sramek, Observations of tropospheric phase scintillations at 5 GHz on vertical paths, *Radio Science*, **17**, 1579–1586, 1982.

Barrett, A.H., V.K. Chung, A method for the determination of high-altitude water-vapor abundance from ground-based microwave observations, *J. Geophys. Res.*, **67**, 4259–4266, 1962.

Bean, B.R., E.J. Dutton, *Radio Meteorology*, National Bureau of Standards Monograph 92, 1966.

Becker, G.E., S.H. Autler, Water vapor absorption of electromagnetic radiation in the centimeter wave-length region, *Phys. Rev.*, **70**, 300–307, 1946.

Berman, A.L., The prediction of zenith range refraction from surface measurements, *Tech. Report 32-1602*, Jet Propulsion Laboratory, Pasadena, California, 1976.

Birnbaum, G. S.K. Chatterjee, The dielectric constant of water vapor in the microwave region, *J. Appl. Phys.*, **23**, 220–223, 1952.

Boudouris, G., On the index of refraction of the air, the absorption and dispersion of centimeter waves by gasses, *J. Res. Natl. Bur. Stand.*, **67D**, 631–684, 1963.

Bracewell, R.N., *The Fourier Transform and its Applications*, McGraw-Hill, New York, 1978.

Chandrasekhar, S., *Radiative Transfer*, Dover, New York, 1960.

Chao, C.C., A model for tropospheric calibration from daily surface and radiosonde balloon measurements, *Tech. Mem. 391-350*, Jet Propulsion Laboratory, Pasadena, California, 1972.

Clark, T.A., B.E. Corey, J.L. Davis, T.A. Herring, H.F. Hinteregger, C.A. Knight, J.L. Levine, G. Lundqvist, C. Ma, E.F. Nesman, R.B. Phillips, A.E.E. Rogers, B.O. Rönnäng, J.W. Ryan, B.R. Schupler, D.B. Shaffer, I.I. Shapiro, N.R. Vandenberg, J.C. Webber, A.R. Whitney, Precision geodesy using the Mark-III very-long-baseline interferometer system, *IEEE Trans. GARS*, **GE-23**, 438–449, 1985.

Colegrove, F.D., W.B. Hanson, F.S. Johnson, Eddy diffusion and oxygen transport in the lower thermosphere, *J. Geophys. Res.*, **70**, 4931–4941, 1965.

Corey, B.E., *A model for the JPL WVR's*, unpublished memorandum, Haystack Observatory, Westford, Massachusetts, 1983.

Crane, R.K. Refraction effects in the neutral atmosphere, in *Methods of Experimental Physics, Vol. 12B*, M.L. Meeks, ed., Academic Press, New York, pp. 186–200, 1976.

Davenport, W.B., *An Introduction to the Theory of Random Signals and Noise*, McGraw-Hill, New York, 1958.

Debye, P., *Polar Molecules*, Dover, New York, 1929.

Decker, M.T., E.R. Westwater, F.O. Guiraud, Experimental evaluation of ground-based microwave radiometric sensing of atmospheric temperature and water vapor profiles, *J. Appl. Met.*, **17**, 1788–1795, 1978.

Dravskikh, A.F., A.M. Finklstein, Tropospheric limitations in phase and frequency coordinate measurements in astronomy, *Astrophys. Space Sci.*, **60**, 251–265, 1979.

Elgered, G., *Water vapor radiometry with applications to radio interferometry and meteorology*, Ph.D. Thesis, Chalmers University of Technology, Göteborg, Sweden; and papers therein, 1983.

Elgered, G., G. Lundqvist, *Calibration of the excess propagation path for radio waves in the neutral atmosphere*, Onsala Space Observatory Technical Memorandum 1984.07.06, Sweden, 1984.

Elgered, G., B. Rönnäng, E. Winberg, J. Askne, *Satellite-earth range measurements I: Correction of the excess path length due to atmospheric water vapour by ground based microwave radiometry*, final report, European Space Agency, ESTEC/Contract No. 5910/84/NL/MD, 1985.

Essen, L., K.D. Froome, The refractive indices and dielectric constants of air and its principal constituents at 24,000 Mc/s, *Proc. Royal Soc. B*, **64**, 862–875, 1951.

Evans, G., C.W. McLeish, *RF Radiometer Handbook*, Airtech House, Dedham, Massachusetts, 1977.

Gary, B.L., S.J. Keihm, M.A. Janssen, Optimum strategies and performance for the remote sensing of path-delay using ground-based microwave radiometers, *IEEE Trans. GARS*, **GE-23**, 479–484, 1985.

Gardner, C.S., Correction of laser tracking data for the effects of horizontal refractivity gradients, *Appl. Opt.*, **16**, 2427–2432, 1977.

Gleuckauf, E., The composition of the atmosphere, in *Compendium of Meteorology*, edited by T.F. Malone, pp. 3–10, American Meteorological Society, Boston, Massachusetts, 1951.

Goldstein, H., Attenuation by condensed water, in *Propagation of Short Radio Waves*, edited by D.E. Kerr, McGraw-Hill, New York, 1951.

Gorden, M.A., see computer programs in Chapter 6 of *Methods of Experimental Physics, Vol. 12B*, M.L. Meeks, ed., Academic Press, New York, 1976.

Guiraud, F.O., J. Howard, D.C. Hogg, A dual-channel microwave radiometer for measurement of precipitable water vapor and liquid, *IEEE Trans. Geosci. Electron.*, **GE-17**, 129–136, 1979.

Hecht, E., A. Zajac, *Optics*, Addison-Wesley, Reading, Massachusetts, 1979.

Herring, T.A., *Precision and accuracy of intercontinental distance determinations using radio interferometry*, Ph.D. Thesis, Massachusetts Institute of Technology, Cambridge, Massachusetts, 1983. Also published as Air Force Geophysics Laboratory report AFGL-TR-84-0182.

Herring, T.A., Precision of vertical position estimates from VLBI, *J. Geophys. Res.*, in press, 1986.

Herring, T.A., I.I. Shapiro, T.A. Clark, C. Ma, J.W. Ryan, B.R. Schupler, C.A. Knight, G. Lundqvist, D.B. Shaffer, N.R. Vandenberg, B.E. Corey, H.F. Hinteregger, A.E.E. Rogers, J.C. Webber, A.R. Whitney, G. Elgered, B.O. Rönnäng, J.L. Davis, Geodesy by radio interferometry: Evidence for contemporary plate motion, *J. Geophys. Res.*, in press, 1986.

Hess, S.L., *Introduction to Theoretical Meteorology*, Holt, Reinhart and Winston, New York, 1959.

Hill, R.J., R.S. Lawrence, J.T. Priestly, Theoretical and calculational aspects of the radio refractive index of water vapor, *Radio Science*, **17**, 1251–1257, 1982.

Hogg, D.C., F.O. Guiraud, W.B. Sweezy, The short-term temporal spectrum of precipitable water vapor, *Science*, **213**, 1112–1113, 1981.

Jackson, J.D., *Classical Electrodynamics*, Wiley, New York, 1975.

Janssen, M.A., A new instrument for the determination of radio path delay due to atmospheric water vapor, *IEEE Trans. GARS*, **GE-23**, 485–490, 1985.

Jenkins, G.M., D.G. Watts, *Spectral Analysis and its Applications*, Holden-Day, New York, 1968.

Johansson, J., The site and seasonal dependence of the wet path delay algorithm used in water vapor radiometry, in *Proceedings of the 4th working meeting on European VLBI for geodesy and astrometry*, edited by B. Rönnäng and G. Tang, held at Onsala Space Observatory, Sweden, 3 June 1985.

Kalman, R.E., A new approach to linear filtering and prediction problems, *Journal of Basic Engineering* (ASME Trans.), **82D**, 35–45, 1960.

Kalman, R.E., R.S. Bucy, New results in linear filtering and prediction theory, *Journal of Basic Engineering* (ASME Trans.), **83D**, 1961.

Lanyi, G., Tropospheric delay effects in radio interferometry, *TDA progress report 42–78*, Jet Propulsion Laboratory, Pasadena, California, 1984.

Leslie, D.C., *Developments in the Theory of Turbulence*, Oxford University Press, 1973.

Liebe, H.J., An update model for millimeter wave propagation in moist air, *Radio Science*, **20**, 1069–1089, 1985.

Liebelt, P.B., *An Introduction to Optimal Estimation*, Addison-Wesley, Reading, Massachusetts, 1967.

Lumley, J.L., H.A. Panofsky, *The Structure of Atmospheric Turbulence*, Wiley, New York, 1964.

Lundqvist, G.L., R.I. Potash, *WVR brightness-temperature corrections due to finite beamwidth*, NASA internal memorandum, August 30, 1985.

Ma, C., T.A. Clark, J.W. Ryan, T.A. Herring, I.I. Shapiro, B.E. Corey, H.F. Hinteregger, A.E.E. Rogers, A.R. Whitney, C.A. Knight, G.L. Lundqvist, D.B. Shaffer, N.R. Vandenberg, J.C. Pigg, B.R. Schupler, B.O. Rönnäng, Radio source positions from VLBI, to be submitted to *Astronomical Journal*, 1986.

Marini, J.W., Correction of satellite tracking data for an arbitrary atmospheric profile, *Radio Science*, **7**, 223–231, 1972.

Marini, J.W., C.W. Murray, Correction of radio range tracking data for atmospheric refraction at elevations above 10 degrees, unpublished memorandum, 1974.

Monin, A.S., A.M. Yaglom, *Statistical Fluid Mechanics, Vol. II*, MIT Press, Cambridge, Massachusetts, 1975.

Moran, J.M., B.R. Rosen, Estimation of the propagation delay through the troposphere from microwave radiometer data, *Radio Science*, **16**, 235–244, 1981.

Panchev, S., *Random Fluctuations and Turbulence*, Pergamon Press, New York, 1971.

Reber, E.E., J.R. Swope, On the correlation of the total precipitable water in a vertical column and absolute humidity at the surface, *J. Appl. Met.*, **11**, 1322–1325, 1972.

Resch, G.M., Water vapor radiometry in geodetic applications, in *Geodetic Refraction*, edited by F.K. Brunner, Springer-Verlag, Berlin, 1984.

Resch, G.M., D.E. Hogg, P.J. Napier, Radiometric correction of atmospheric path length fluctuations in interferometric experiments, *Radio Science*, **19**, 411–422, 1984.

Resch, G.M., M.C. Chavez, N.I. Yamane, K.M. Barbier, R.C. Chandlee, Water Vapor Radiometry Research and Development Phase Final Report, *JPL Publication 85-14*, Jet Propulsion Laboratory, Pasadena, California, 1985.

Robertson, D.S., *Geodetic and astrometric measurements with very-long-baseline interferometry*, Ph.D. Thesis, Massachusetts Institute of Technology, Cambridge, Massachusetts, 1975.

Robinson, S.E., Algorithm accuracy for delay estimation from WVR data, JPL unpublished memorandum, 1985.

Rozenkranz, P.W., Shape of the 5 mm oxygen band in the atmosphere, *IEEE Trans. Ant. Propagation*, **AP-23**, 498–506, 1975.

Saastamoinen, J., Atmospheric correction for the troposphere and stratosphere in radio ranging of satellites, in *The Use of Artificial Satellites for Geodesy, Geophys. Monogr. Ser., vol. 15*, edited by S.W. Henriksen *et al.*, pp. 247–251, American Geophysical Union, Washington, D.C., 1972.

Schaper, L.W., D.H. Staelin, J.W. Waters, The estimation of tropospheric electrical path length by microwave radiometry, *Proc. IEEE*, **58**, 272–273, 1970.

Snider, J.B., E.R. Westwater, Atmospheric attenuation at 15, 31, and 53 GHz, *ESSA Tech. Report ERL 156-WPL 11*, U.S. Department of Commerce, Boulder, Colorado, 1969.

Snider, J.B., H.M. Burdick, D.C. Hogg, Cloud liquid measurements with a ground-based microwave instrument, *Radio Science*, **15**, 683–693, 1980.

Staelin, D.H., Measurements and interpretation of the microwave spectrum of the terrestrial atmosphere near 1-centimeter wavelength, *J. Geophys. Res.*, **71**, 2875–2881, 1966.

Tatarskii, V.I., *Wave propagation in a turbulent medium*, R.A. Silverman, trans. McGraw-Hill, New York, 1961.

Taylor, G.I., The spectrum of turbulence, *Proc. Roy. Soc.*, **A164**, 476, 1938.

Thayer, G.D., An improved equation for the radio refractive index of air, *Radio Science*, **9**, 803–807, 1974.

Van Vleck, J.H., V.F. Weisskopf, On the shape of collision-broadened lines, *Rev. Mod. Phys.*, **17**, 227–236, 1945.

Van Vleck, J.H., The absorption of microwaves by oxygen, *Phys. Rev.*, **71**, 413–424, 1947a.

Van Vleck, J.H., The absorption of microwaves by uncondensed water vapor, *Phys. Rev.*, **71**, 425–433, 1947b.

Vinnichenko, N.K., Pinus, N.Z., Shmeter, Shur, *Turbulence in the Free Atmosphere*, F.L. Sinclair, trans., Plenum, New York, 1980.

Whitney, A.R., *Precision geodesy and astrometry via very-long-baseline interferometry*, Ph.D. thesis, Massachusetts Institute of Technology, Cambridge, Massachusetts, 1974.

Waters, J.W., Absorption and emission by atmospheric gasses, in *Methods of Experimental Physics, Vol. 12B*, M.L. Meeks, ed., Academic Press, New York, pp. 142–176, 1976.

Wu, S.C., Optimum frequencies of a passive microwave radiometer for tropospheric path-length correction, *IEEE Trans. Ant. Prop.*, **AP-20**, 181–187, 1979.