ACCURATE AND EFFICIENT LONG-RANGE LIGHTNING
GEO-LOCATION USING A VLF RADIO ATMOSPHERIC
WAVEFORM BANK

A DISSERTATION
SUBMITTED TO THE DEPARTMENT OF ELECTRICAL
ENGINEERING
AND THE COMMITTEE ON GRADUATE STUDIES
OF STANFORD UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Ryan Khalil Said
September 2009

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

_____

(Umran S. Inan)   Principal Adviser

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

_____

(Kenneth L. Cummins
The University of Arizona)

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

_____

(John T. Gill III)

Approved for the University Committee on Graduate Studies.

This dissertation is dedicated to my mother and father
Connie *and* Jim

Abstract

On average there are ∼50 lightning flashes worldwide every second, with activity varying by region and season. Many systems currently exist that detect and locate lightning flashes for a broad range of commercial and scientific applications, including air traffic control, insurance claims, climate modeling, and the investigation of secondary atmospheric and magnetospheric electrical phenomena. These lightning detection systems have varying degrees of coverage area and location accuracy. Commercial ground-based systems that excel at locating return strokes in cloud-to-ground lightning use radio detection in the LF ($30-300$ kHz) band to provide very accurate location data, with a typical accuracy of ∼0.5 km, but they require a dense network of receivers separated by ∼400 km and are therefore primarily limited to monitoring the land areas within the network.

In addition to radiating in the LF band, each lightning strike generates a broadband electromagnetic pulse containing frequencies from a few Hz through to the optical band with a peak component at VLF ($3-30$ kHz). Radio waves at VLF propagate through the waveguide formed by the Earth and the ionosphere with relatively low attenuation (∼3 dB per 1000 km), enabling the detection of these pulses, called radio atmospherics, at great distances from the lightning strike. Several existing networks utilize this efficient guiding to geo-locate lightning strikes often at distances greater than 5000 km from a given receiver. However, the Earth-ionosphere waveguide also presents a complex and time-varying channel that heavily disperses the pulse as it propagates away from the strike location. These networks fail to adequately address the path-dependence of the received impulse and suffer a lower location accuracy as a result (∼20 km).

A new technique of long-range global lightning location is presented that both takes advantage of the efficient propagation at VLF and addresses the path-dependence of the propagation channel. This new technique catalogs the dominant variation in expected received waveforms to form a set of waveform banks, which are then used to estimate the propagation distance and identify features on each waveform that allow for a more accurate determination of the arrival time. Using three stations in a trial network, this new technique is used to demonstrate an accuracy of $1-4$ km, depending on network geometry and the time of day. Furthermore, this technique provides an estimate of the peak current and polarity in the lightning channel, parameters that existing long-range networks do not measure using VLF radio atmospherics.

The propagation distance estimated at each receiver, together with an arrival azimuth measurement, enables accurate geo-location using as few as three sensors. The redundancy offered by this range and azimuth information mitigates the complexity involved with correlating radio atmospherics from multiple sensors and enables a high detection efficiency. An overall stroke detection efficiency between $\sim$40–60% is estimated by correlating individual lightning stroke events to data from a commercial LF reference network. There are a significant number of additional events reported by the trial network that do not time-correlate with data from the commercial network, but the tight spatial clustering of lightning strokes between the two networks suggest that many of the unmatched events are not spurious but rather may correspond to weak cloud-to-ground strokes or cloud flashes.

Acknowledgments

I would like to take this opportunity to express my sincere gratitude to some of the many people whose insight, guidance, and support have helped make this work possible.

First and foremost, I would like to thank my adviser, Professor Umran Inan. Many years ago when I was searching for a Ph.D project, Professor Inan proposed a simple and alluring challenge: to produce a display of the world with real-time lightning location data. Buoyed by his infectious enthusiasm and dedication to his work, and helped a great deal by his guidance and collaboration, I am pleased to submit this thesis as partial fulfillment of his original challenge.

I am also immensely grateful for the input and counsel of my co-adviser, Dr. Ken Cummins. With almost twenty years of research and development experience in lightning geo-location, Dr. Cummins has been an invaluable source of insights, reality checks, and knowledge over the past several years. Without his guidance, this work would not have progressed nearly as far as it has. Finally, I thank Dr. Cummins for his cheerful willingness to fly out to Stanford in order to serve on my defense committee, and for his many detailed and thoughtful comments on this dissertation.

For their time and for their input, I thank Professor John Gill III, who graciously agreed to join my reading committee on short notice; Professor Don Cox, for serving on my defense committee; and Professor Nick Bambos for agreeing to be my defense chair.

This work uses data collected from several geographically remote locations, and I thank the many people who have helped me collect and store this data. Professor Bob Davis at Taylor University has provided constant support with a sensor on his

importance of balancing work with some play). I also thank my sister, Amirah, for her friendship, compassion, and ready ear. I have received an endless amount of encouragement from my grandparents and the rest of my extended family, who have been very understanding as, year after year, I tell them that no, I have not graduated yet. Finally, I would like to thank my beautiful Karen for all the ways she has lovingly supported me over the past few years, from her understanding of my erratic work schedule to her help with reading this thesis.

RYAN K. SAID

Stanford, California
September 18, 2009

Contents

List of Figures

Chapter 1

Introduction

Lightning is an electrical discharge that partially neutralizes accumulated charge in a cloud. Recent satellite-based measurements estimate that there are an average of ~50 lightning flashes per second around the world [*Christian et al.*, 2003]. This work concerns the development of a new method to detect and geo-locate as many of these lightning flashes as possible using a sparse network of ground-based Very Low Frequency (VLF; 3−30 kHz) receivers.

1.1 Lightning

In fair weather conditions, the electric field measured at the Earth's surface is ~-100 V/m (i.e., the field is pointed downward; the negative sign assumes a z-axis outward from Earth), caused by a positively charged ionosphere at a potential of ~300,000 V. Measurements of this electric field on the ground in *Krehbiel* [1986] show an interesting pattern. Before the onset of lightning activity, the polarity of the electric field at the ground reverses. After a lightning strike, the ground electric field suddenly drops off and then it slowly ramps back up to a large positive value.

This remote field observation suggests a large negative charge layer near the base of the thundercloud that is partially neutralized in the course of the lightning discharge and that subsequently builds back up in the minutes following the strike. This picture is supported by many other remote field and in-situ measurements. Balloon
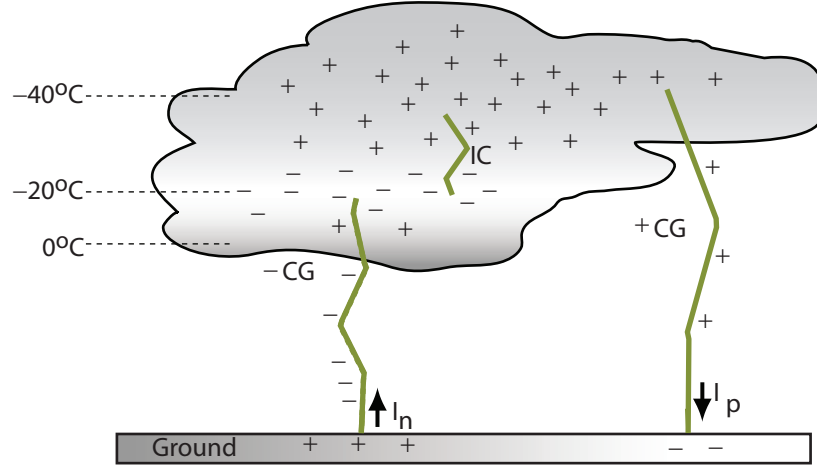
Figure 1.1: Illustration of the charge separation within a cloud and the three categories of lightning strokes.

experiments designed to map out the electric field within a cloud have found a general tripole structure [*Byrne et al.*, 1983], with a large negative charge underneath a large positive charge and above a smaller positive charge. Figure 1.1, partially adapted from *Krehbiel* [1986], illustrates this idealized charge structure.

The mechanisms behind the charge separation are not completely understood. Two primary mechanisms have been proposed and may both contribute to the charge accumulation: the precipitation theory and the convection theory [*Uman*, 2001, p. 65]. The precipitation theory is believed to account for most of the charge build up and posits that heavy precipitation particles falling due to gravity collide with and strip negative charge from small precipitation particles that are traveling upward due to updrafts of wind. The preference of the negative charge to transfer to the larger, falling particles result in a positive layer at the top of the cloud and a large negative layer toward the base. Laboratory experiments [*Jayaratne et al.*, 1983] show that this charge transfer depends on the local temperature in the cloud. For temperatures cooler than $\sim -20°$, negative charge is transferred to the heavier particles. For warmer temperatures, positive charge is transferred, which may account for the small positive layer at the base of the cloud. These temperature separations are supported by remote measurements of the source height of lightning in conjunction with cloud temperature

measurements [*Brook et al.*, 1982] and are indicated in Figure 1.1. This mechanism is frequently referred to as non-inductive charge separation, since the particles do not need to have an initial charge.

The convection theory holds that accumulated charge at the surface of the Earth (or perhaps on a screening layer surrounding the accumulated charge of the thundercloud itself) is moved to the cloud to create the various charge layers. A similar process is thought to contribute to the positive charge layer at the base of the cloud [*Krehbiel*, 1986], whereby positive corona from the ground and positive ions left by cosmic rays are moved in bulk up to the cloud.

Whatever the charging mechanism, convective cumulonimbus clouds contain a charge separation structure which generates strong local electric fields. These fields create an initial breakdown[1] that can eventually form a conductive channel of charge called a leader. For negative cloud-to-ground (CG) flashes, this leader produces a tortuous, step-wise branching pattern; this "stepped leader" propagates toward the ground, depositing negative charge along the way. As it nears the ground, it eventually connects to a small induced upward positive leader. Once contact is made, the negative charge flows toward the earth, creating an upward going current which acts as a large vertical antenna, radiating at all frequencies from a few Hz [*Burke and Jones*, 1992] through to the optical band [*Weidman and Krider*, 1986]. There is a peak in the radiated spectrum in the VLF band, from a few kHz to about 30 kHz; a summary of measured field spectra up to the Very High Frequency (VHF; $30-300$ MHz) band is give in Figure 1.2. This type of a primary discharge is called a negative CG first stroke.

After the primary discharge, the channel remains conductive and small, secondary currents may flow from the cloud to the ground. Subsequent dart leaders that propagate down this same channel precede subsequent strokes, which produce similar current characteristics as the first stroke. The entire process, including the first and subsequent strokes, is called a CG lightning flash [*Cummins and Murphy*, 2009].

---

[1]The measured electric fields are actually too low to account for an electric breakdown of air at the altitude of the cloud. Various mechanisms for the formation of charge channels have been proposed, including the formation of corona streamers around hydro-meteors [e.g., *Richards and Dawson*, 1971].
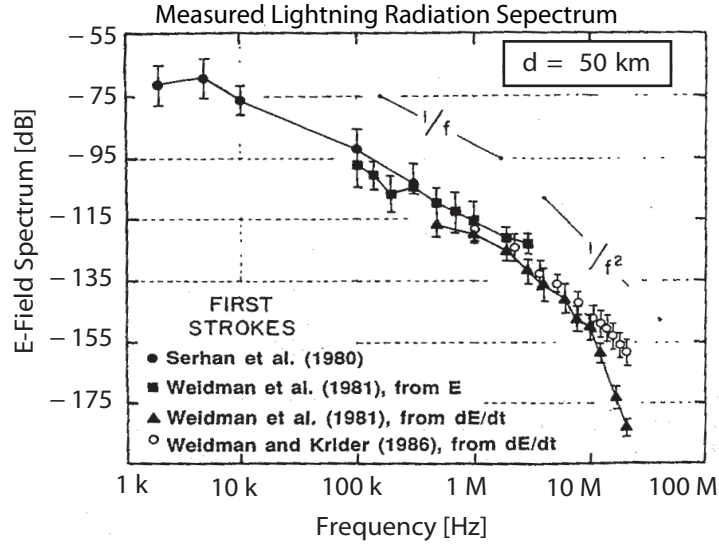
Figure 1.2: Measured radiation spectrum from negative first strokes. Adapted from *Uman* [2001, p. 118].

A similar process may occur which lowers positive charge to ground, resulting in a positive CG flash. The return stroke current and therefore also the radiated field have an opposite polarity with respect to the negative CG flash. Due to the typically larger separation between the upper (positively charged) part of the cloud to the ground, these types of flashes are less common than the negative CG flashes. Positive lightning does occur with relatively more frequency in winter lightning, where, due to the colder atmosphere, the charge separation occurs at a lower altitude with the separation between the main positive layer and the ground being correspondingly reduced [*Brook et al.*, 1982].

The most common type of flash, called a cloud flash (labeled as intra-cloud (IC) in Figure 1.1), does not reach the ground at all and only neutralizes the charge separation within a cloud. These types of discharges tend to have a lower peak current [*Murphy and Cummins*, 1998], radiate a weaker field [*Weidman et al.*, 1981] and, depending on the relative position of the positive and negative charge reservoirs, the vertical component of the current may be quite small and can point either up or down.

The radiated fields used in this work are preferentially associated with current

in the vertical channels in cloud and CG flashes. A conducting layer in the upper atmosphere called the ionosphere guides these radiated fields to great distances in the so-called earth-ionosphere waveguide. This guided electromagnetic pulse is called a radio atmospheric, or sferic, and has a low attenuation in the VLF band [*Wait and Spies*, 1964]. Our ability to measure this impulse at great distances from the strike forms the basis for a long-range lightning geo-location network.

1.2 Geo-Location

The goal of this thesis is to develop a ground-based lightning geo-location network. This section describes the procedure for using remote sensing to locate lightning discharges. The approach described below is used in this work, and variations are used in many of the existing networks currently in use that are reviewed in the next section.

Figure 1.3 shows a cartoon of a CG lightning stroke with four sensors, Rx_1 through Rx_4. Each sensor is equipped with one or more antennae to detect the magnetic and/or electric field from the sferic generated by the lightning discharge. Each sensor measures various properties of the recorded sferic and sends a small amount of information back to a central processor (CP). The CP uses the information from each station to determine the location and time of the strike, plus, in some cases, information about the current in the discharge as well as its polarity.

Figure 1.3 illustrates a two-dimensional example, where the CP only determines the position of the strike on the plane (the earth's surface), specified by two coordinates. This thesis is focused on using the VLF frequency component of the sferic to achieve long-range lightning geo-location, using energy launched primarily in the lower part of the return stroke channel. As such, two coordinates, indicating the location on the earth's surface where the stroke connected to the ground (or, for intra-cloud strikes, the approximate location on the earth directly beneath the cloud discharge), are sufficient. Some sensor networks use VHF or High Frequency (HF; $3-30$ MHz) sensors to detect the radiation from the finer breakdown processes in a lightning discharge [*Thomson et al.*, 1994; *Thomas et al.*, 2000]. These networks
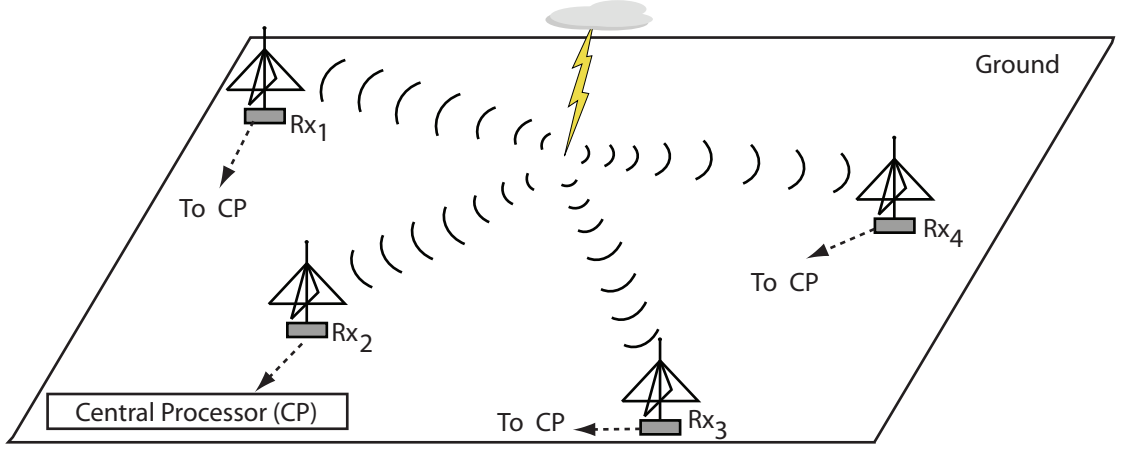
Figure 1.3: Illustration of a cloud-to-ground lightning discharge generating a sferic which is sensed by four receiver stations, each with a communications link to a central processor.

add a third dimension: altitude. The techniques described in this section are directly applicable to geo-location with this third dimension, though the redundancy level is reduced by one dimension to compensate for this additional unknown.

In the general case, each receiver i in the network estimates four parameters from each measured sferic: the arrival angle θ_i^m, usually measured in degrees east from geographic north; the range R_i^m between the receiver and the source location of the lightning strike; the arrival time t_i^m, measuring the GPS-referenced time of arrival of the sferic; and the amplitude (with polarity) A_i^m. These parameters are illustrated in Figure 1.4, with the relevant parameters highlighted in red. The parameters are sent back to the CP, which then correlates them with measurements from other sensors to determine various parameters of the causative discharge. These derived parameters include the discharge position $\mathbf{r}(\lambda, \phi)$, where λ and ϕ are the latitude and longitude of the strike on the earth's surface, respectively; the strike time t_0, referenced to Universal Time (UT), and the discharge's peak current A and polarity.

Using these four measurements, any number of stations can be used to estimate the strike location and strength, though the estimates improve with greater redundancy. Figure 1.5a illustrates a simple approach using up to three receivers. Supposing
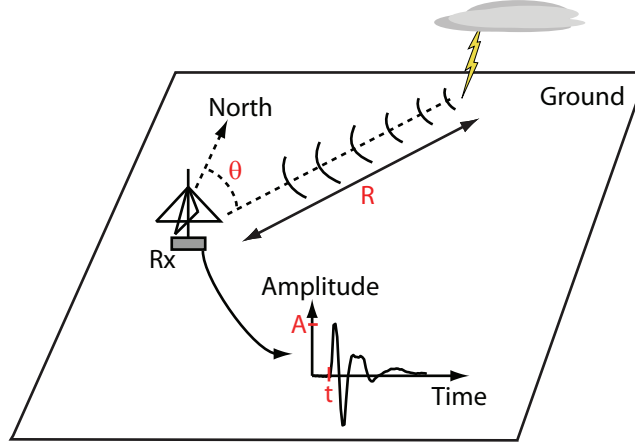
Figure 1.4: Illustration of four parameters measured by each receiver: arrival time t, arrival azimuth θ, amplitude and polarity A, and source-receiver range R.

first that only one receiver, Rx_1 in the figure, detected a sferic, then the solitary receiver can estimate the strike location (i.e., the polar coordinates of the strike, assuming the receiver is at the origin r and ϕ) by estimating the propagation distance R and the arrival azimuth θ, respectively. As detailed in later sections, the estimated propagation distance R typically can only be determined at best with $\sim 20\%$ accuracy. Thus, this single receiver method provides only a crude location estimate, especially at great distances.

Suppose now two receivers, Rx_1 and Rx_2, detected the sferic. In this case the sferic arrival time difference (ATD), labeled as 'ATD 1-2' in the figure, between the two sensors defines a hyperbola (or a hyperboloid if the two sensors lie on an oblate spheroid, such as the earth), on which the strike position must occur. The location of the strike along this curve is defined by the intersection of a projected azimuth line from one of the sensors, indicated by a black line in the figure. The location of the arrival time difference curve depends on an assumed propagation speed v of the sferic from the strike location to the receiver, which, for the present discussion, can be taken as the speed of light c[2].

---

[2]The frequency-dependent phase and group velocity through a waveguide are not, in general, equal to the speed of light, so the chosen propagation speed is determined by the method one uses to determine the arrival time. The technique for determining the arrival time in Chapter 5 is designed so that the central processor may assume a propagation speed of c.
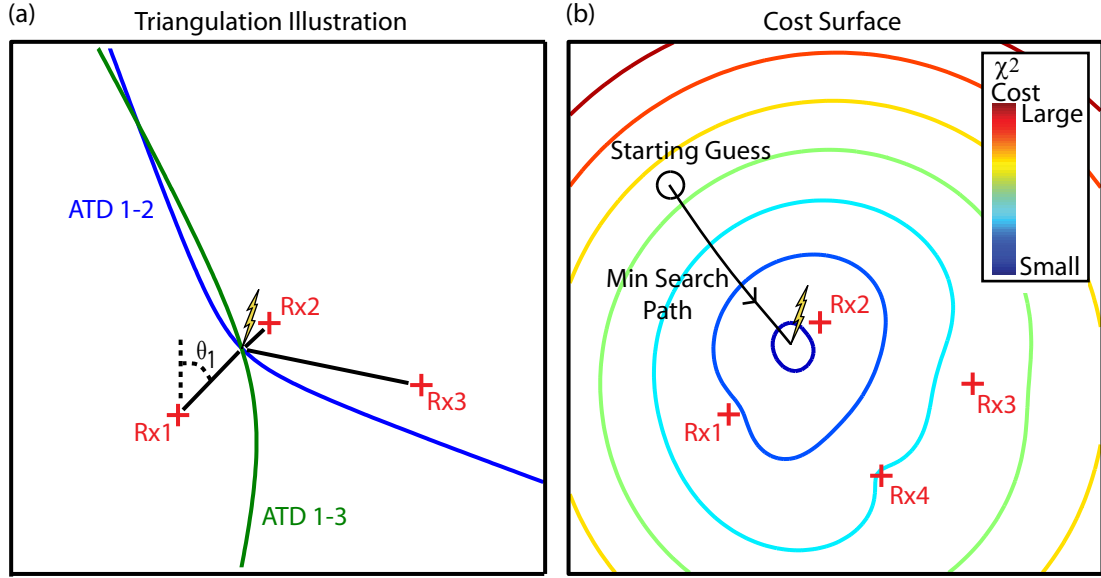
Figure 1.5: Illustration of the triangulation procedure by (a) using intersecting arrival azimuth and/or arrival time difference curves and (b) finding the minimum of a cost surface.

Depending on the network geometry[3], constraining the strike location to an arrival time difference curve can lead to substantial improvements in the location accuracy. Consider a timing accuracy of $\Delta t \simeq 33$ μs. Assuming the strike location is along the baseline between the two sensors, this timing uncertainty contributes to a location uncertainty in the direction of the baseline of $\Delta d = c \Delta t \simeq 10$ km (if the strike location is off of this baseline, a geometric correction factor must be introduced). For configurations where the lightning source is reasonably close to the baseline between two sensors, the location uncertainty is largely a function of the arrival azimuth uncertainty. As an example, if the strike were 1000 km from the sensor and the arrival azimuth was measured with an uncertainty of $\Delta \theta \simeq 1°$, then the related location uncertainty would be $r \Delta \theta \simeq 17$ km. Note that this location error scales linearly with distance; at 5000 km from the receiver, this same angular uncertainty yields a positional uncertainty of 87 km. Depending on the range estimation accuracy, this angular

---

[3]For a closer look at the network geometry's effect on location accuracy, see Appendix C.

uncertainty can still be a substantial improvement over the case where the strike location is determined by one sensor using a range estimate. Suppose the range was accurate to within $\Delta R \simeq 20\%$, then even at 1000 km, this range uncertainty translates to a location uncertainty of 200 km.

Adding a third station defines two ATD curves, which now intersect at two points defining two possible strike locations. Under most conditions, one of these possible locations can be eliminated through analysis of the amplitude, range or azimuth information. The ambiguity is generally best-resolved by adding the azimuth information. The location estimate is typically determined by arrival-time-difference only, thus the location accuracy is solely determined by the arrival time uncertainty. Due to the linear relationship between the strike position and time and the fact that the arrival azimuth-determined position accuracy depends on range, strike coordinates are typically far more accurate using intersecting ATD curves compared to intersecting arrival azimuth lines, especially at large distances.

In the above analysis, more stations were added to the result, "weak" information was steadily discarded to arrive at a solution. With the addition of a second receiver, only the arrival azimuth from one station was used and the range estimates were discarded completely. With the addition of a third station, the arrival azimuths were used only to distinguish between two intersection curves. Adding a fourth station would define a third ATD line, which may not intersect with one of the two existing intersection points. The problem is over-determined, meaning that one can instead minimize a locally convex cost function over a constrained set of parameters. This type of minimization is illustrated in Figure 1.5b. A starting point is guessed, and, through an iterative procedure, the minimum point of the cost function is found and taken to be the strike location.

The cost surface depicted in Figure 1.5b is a function of the two position coordinates λ and ϕ. The cost function may be a function of other parameters as well, such as the strike time t_0. In this case, the surface would be three dimensional.

To calculate an optimal location and time estimate in a minimum mean squared error (MMSE) sense, the cost surface χ^2 is defined as the sum of several squared-error terms, one for each measured variable. Each error term measures the difference

between the inferred measurement given a set of strike coordinates and the measured value, where each difference term is normalized by the standard deviation of that measurement to allow for appropriate weighting with other terms in the summation.

Let λ and ϕ be the strike latitude and longitude, defining a strike position $\mathbf{r}(\lambda, \phi)$, and let t_0 be the strike time. Then the normalized arrival time error at receiver i, given a measured arrival time t_i^m and a timing uncertainty $\sigma_{t,i}$, as a function of the strike position and strike time is

$$(\Delta t)_i = \frac{\frac{1}{v}\widehat{d}\left(\mathbf{r}_i, \mathbf{r}\right) - (t_i^m - t_0)}{\sigma_{t,i}} \tag{1.1}$$

where $\widehat{d}\left(\mathbf{r}_1, \mathbf{r}_2\right)$ is the distance (or, on the earth, the geodesic distance) between coordinates $\mathbf{r}_1$ and $\mathbf{r}_2$, and $\mathbf{r}_i$ is the location of sensor i with (coordinates λ_i and ϕ_i). Similarly, given a measured arrival azimuth θ_i^m at sensor i and an arrival azimuth uncertainty $\sigma_{\theta,i}$, the error term from arrival azimuth as a function of the strike position is

$$(\Delta\theta)_i = \frac{\widehat{\theta}\left(\mathbf{r}_i, \mathbf{r}\right) - \theta_i^m}{\sigma_{\theta,i}} \tag{1.2}$$

where $\widehat{\theta}(\mathbf{r}_1, \mathbf{r}_2)$ is the azimuth at $\mathbf{r}_1$, in degrees East from North, of the line joining $\mathbf{r}_1$ and $\mathbf{r}_2$.

Additional error terms that incorporate the range estimation R_i^m and peak current estimation A_i^m may be added. Let the function $\widehat{A}(I_{\max}, \mathbf{r}_i, \mathbf{r})$ denote the expected received amplitude at $\mathbf{r}_i$ from a strike at location $\mathbf{r}$ with peak current $I_{\max}$. Then the additional error term from the measured amplitude, with a measurement uncertainty $\sigma_{A,i}$, is a function of the strike position and peak current

$$(\Delta A)_i = \frac{\widehat{A}(I_{\max}, \mathbf{r}_i, \mathbf{r}) - A_i^m}{\sigma_{A,i}}. \tag{1.3}$$

The error term from the range estimation, assuming a range uncertainty $\sigma_{R,i}$ which

may be a function of the propagation distance, is also a function of the strike position

$$(\Delta R)_i = \frac{\widehat{d}\left(\mathbf{r}_i, \mathbf{r}\right) - R_i^m}{\sigma_{R,i}}. \tag{1.4}$$

The total cost χ^2 is defined as the sum of the squared error terms, with one term in the summation for each receiver and N total stations:

$$\chi^2(\lambda, \phi, t, I_{\max}) = \frac{1}{4N - 4} \sum_{i=1}^{N} \left[(\Delta t)_i^2 + (\Delta\theta)_i^2 + (\Delta A)_i^2 + (\Delta R)_i^2 \right]. \tag{1.5}$$

The "best estimate" strike latitude $\lambda^\star$, longitude $\phi^\star$, time $t^\star$, and amplitude $I_{\max}^\star$ are found by solving the constrained minimization problem

$$\{\lambda^\star, \phi^\star, t^\star, I_{\max}^\star\} = \arg \min_{\lambda, \phi, t, I_{\max} \in \Psi} \chi^2. \tag{1.6}$$

The vector space Ψ consists of position coordinates on the surface of the earth and time coordinates which lead to propagation times less than half the circumference of the earth divided by c. With a good starting guess, a simple unconstrained optimization procedure may be used, where a flag is set if the optimization falls outside this constrained domain.

The peak current $I_{\max}$ primarily depends on $\{(\Delta A)_i\}$ since the dependence on the other terms is only indirect through the strike location $\mathbf{r}$. Therefore $\{(\Delta A)_i\}$ contributes relatively weakly to the optimization compared to the time and azimuth terms. In practice this term is removed to reduce the number of free variables to three, allowing for a more efficient minimization, and $I_{\max}$ is estimated at the end using the estimated strike position.

A variety of multi-variable minimization algorithms may be used to find the optimal point, though the technique should be chosen wisely based on the shape of the cost surface χ^2. For example, with two stations, the cost surface is very steep on either side of the arrival time difference curve (the condition number of the sublevel sets [*Boyd and Vandenberghe*, 2004, p. 75] of the χ^2 surface may be decreased by increasing σ_t or decreasing σ_θ). A steepest-descent approach would fall onto this line quickly from the

start position, then slowly walk along the line to the optimal point as dictated by the arrival azimuth values [For a discussion on the convergence of the steepest descent method when the optimal search direction is not known, see *Boyd and Vandenberghe*, 2004, pp. 480–484]. A Nelder-Mead simplex method [*Press*, 2002, p. 413] would take a more haphazard walk and may converge to within an acceptable tolerance in fewer time steps. Depending on the starting point and the condition number of the Hessian of the cost function along the search path, the Gauss-Newton approach [*Boyd and Vandenberghe*, 2004, p.484] may converge faster still, but evaluating the Hessian at each time step can eliminate the advantage of this faster convergence if extreme accuracy is not needed. In this work, the Nelder-Mead simplex method is used.

Redundant measurements also allow for an internal consistency check through the final value of the cost function. A low cost indicates that all parameters were simultaneously satisfied. A high cost value indicates that one or more parameters do not agree well with the final strike location, time, and amplitude—indicating a measurement that either has significant error or, if sferics from different lightning strikes are used, is invalid.

For three or more stations, for a sufficiently small σ_t, the location accuracy of the strike position is heavily dependent on the arrival time accuracy and knowledge of the propagation speed v. This thesis thus focuses on determining v and minimizing the timing uncertainty σ_t. To reduce the complexity at the central processor, the velocity is fixed at c, and then σ_t is minimized to the extent possible. The following two sections review the existing lightning geo-location technologies and itemize the scientific contributions in this thesis.

1.3 VLF lightning location systems

All ground-based long-range lightning detection systems use the Extremely Low Frequency (ELF; 300 Hz$-$3 kHz) and VLF portion of the electromagnetic spectrum, utilizing the low attenuation through the earth-ionosphere waveguide at these frequencies. VLF lightning geo-location systems can be divided into two classes: systems that use arrival time information and systems that use magnetic direction finding

(MDF) either solely or in conjunction with arrival time information.

As shown above, at long ranges, even small angle errors, due to polarization, site-error, or signal-to-noise (SNR) limitations, can lead to large location errors. For this reason, existing long-range systems rely on using timing information to geo-locate the lightning strike. With only timing information, four stations are needed[4] to geo-locate an event.

There are two existing long-range technologies that geo-locate lightning strikes using exclusively timing information of VLF measurements of individual sferics. The first approach, which is referred to as simply the Arrival Time Difference (ATD) technique [*Lee*, 1986], determines the arrival time difference between sensors by cross-correlating recorded sferic waveforms. In this approach, equation (1.1) needs to be modified so that each term calculates an arrival time difference between pairs of sensors. The second approach calculates a Time of Arrival (TOA) at each receiver by calculating an averaged group delay arrival time, a value referred to as the Time of Group Arrival (TOGA) [*Dowden et al.*, 2002].

1.3.1 ATD networks: UK Met Office, STARNET, and ZEUS

In 1986 Lee proposed a method for lightning geo-location based on cross-correlation of individual waveforms [*Lee*, 1986]. In this method, the time difference between a sferic recorded at two receivers, defined as the ATD, is calculated as the time at which the cross-correlation between the waveforms reaches a maximum. For a pair of receivers, Lee envisions one waveform as the reference signal, $w_0(t)$, and the other waveform as a scaled and shifted version of the reference signal, plus noise: $w_1(t) = \alpha w_0(t - \delta) + n_1(t)$. In this formulation, the coefficient α can be determined by minimizing the mean square noise $\overline{n_1^2}$ with respect to α, giving

$$\alpha = \frac{\mathrm{CORR}(\delta)}{\overline{w_0^2}}, \tag{1.7}$$

---

[4]Recall that with three stations, any two arrival time difference pairs give a location. At least one redundant measurement is needed to verify that the sferics used in calculating the arrival times originated from the same strike.

where $\text{CORR}(\delta)$ is the time-lag correlogram between $w_1(t)$ and $w_0(t)$ at offset δ (i.e., $\text{CORR}(\delta) = \int w_1(t)w_0(t - \delta)dt$) and $\overline{w_0^2}$ is the mean square of the reference signal. Lee then derives the mean power signal-to-noise ratio

$$R(\delta) = \frac{\alpha^2 \overline{w_0^2}}{n_1^2} = \frac{\text{CORR}^2(\delta)}{w_0^2 w_1^2 - \text{CORR}^2(\delta)} \tag{1.8}$$

as the quantity to be maximized, which yields an improved discrimination between the peak correlation offset and other values of δ when compared to $\text{CORR}(\delta)$ [*Lee*, 1986]. Lee further suggests that the second peak in $R(\delta)$ be used as a metric for the reliability of the correlation offset.

Finally, he achieves sub-sample accuracy in the offset δ through an iterative process. First, an initial value δ_0 is found via quadratic interpolation of the maximum of $R(\delta)$. The waveform $w_1(t)$ is then shifted by an amount necessary to shift this peak to the nearest sample. The process then repeats, starting with the peak interpolation.

If n receivers register a particular sferic, there will be $m = \binom{n}{2}$ pairs of ATD measurements. The cost surface is defined in terms of the ATD error:

$$\chi^2(\phi, \lambda) = \frac{1}{m-2} \sum_{r=1}^{m} \left[\frac{\text{ATD}_{\text{TH}}(r, \phi, \lambda) - \text{ATD}_{\text{M}}(r)}{\sigma(r)} \right]^2 \tag{1.9}$$

where $ATD_{\text{TH}}(r, \phi, \lambda)$ is the theoretical arrival time difference assuming a propagation phase velocity v for a strike at coordinates (ϕ, λ), $ATD_{\text{M}}(r)$ is the measured value, $\sigma^2(r)$ is the variance for the ATD pair r determined by the signal strength at each station and the main and subsequent peaks of $R(\delta)$, and the $m-2$ normalizing factor accounts for the two degrees of freedom lost in calculating the two spatial coordinates.

To mitigate against offsets due to propagation effects, *Lee* [1989] formed a composite χ^2 using multiple received sferics (much in the same way site error is corrected for, see Section 3.2.1). Lee used the extra degrees of freedom afforded by minimizing a group of location errors to calculate a series of offsets to each station. By removing these systematic offsets, Lee achieved a location accuracy for clusters of lightning strikes on the order of $\sim$1$-$2 km.

Lee deployed seven stations during his pilot study between October 1978 and

October 1979. While he could only compare against a substantially less accurate cathode ray direction finding (CRDF) system, he achieved root-mean-square (RMS) χ values between 3.3 and 10.9 μs. This method does not prove an absolute accuracy, but it does speak to a remarkable degree of consistency. Taking the RMS spread to be related to a positional error by approximately $\Delta x = c\Delta t$, this method yields a consistency on the order of $\sim 1-3$ km. The improved accuracy was achieved in the United Kingdom (UK), where there were five stations within 1200 km of each other.

UK Met Office

Lee's ATD technique was first put to operational use by the UK Met Office in June 1988, replacing the labor-intensive CRDF [*Lee*, 1990]. This system used seven stations, with five clustered in the UK and Shetlands, and one receiver each in Gibraltar and Cyprus. Each receiver measured the vertical electric field via a short whip antenna and used a 4.88 kHz half-bandwidth centered around 9.76 kHz. Each detected sferic above a certain threshold was recorded for a 13.1 ms time window and sent back to a central processor.

In *Lee* [1990], four storm clusters were analyzed using this new network. After removing systematic arrival time difference offsets around each storm using redundant ATD measurements, location errors on the order of 1 km were achieved for regions in the interior of the network for short time intervals when the propagation effects were not variable.

STARNET

Between July 1997 and February 1998 an experimental network of five receivers, named as Sferics Timing and Ranging NETwork (STARNET), was established along the United States (US) eastern seaboard and Puerto Rico [*Morales et al.*, 2002b]. This network used the same ATD approach outlined in *Lee* [1986], but with receivers designed by Resolution Displays Inc. and outfitted with GPS timing for improved reliability. Similar to the UK Met system, each receiver consisted of a vertical electric field antenna, with a bandwidth of 6 kHz centered around 9.8 kHz. With lightning

geo-location in the US, results could now be compared to the established National Lightning Detection Network (NLDN) [*Cummins et al.*, 1998a]. STARNET applied constant daytime and nighttime phase velocities of $1.0025c$ and 1.0055, respectively[5]. Location errors, with respect to NLDN strokes, over much of the eastern US and towards the center of the network were between $0-10$ km. The averaged error increased to ~ 50 km toward the western US, due to the accumulation of phase velocity errors over a longer distance [*Morales et al.*, 2002b]. This system also used cloud temperature profiles to weight the strike positions, a statistical correction that is only practical with sufficient supplementary monitoring equipment.

Morales et al. [2002a] estimates the detection efficiency (DE) of the network by comparing individual sferics to NLDN stroke data over a 100 km region in Florida. This region is also monitored by the Lightning Detection and Ranging (LDAR) system [*Lennon and Maier*, 1991]. By assuming only CG detection by NLDN and full CG and cloud flash (IC) detection by the LDAR system, CG (IC) detection efficiencies of $\sim 72\%$ ($\sim 6\%$) and $\sim 54\%$ ($\sim 15\%$) were computed for daytime and nighttime, respectively.

Zeus

A third ATD network that is currently in operation, named Zeus, consists of 10 receivers manufactured by Resolution Displays, Inc. over Europe and Africa, operated by the University of Connecticut and National Observatory of Athens [*Chronis and Anagnostou*, 2003, 2006]. This network employs a much larger receiver separation than the previous ATD networks. Due to the long baselines between each receiver, many sferics may fall into a valid window to be cross-correlated between each station pair. Many different combinations of sferic pairings need to be tested before a global χ^2 minimum is found; to avoid this potential combinatorial explosion, the operation of the network is divided between receivers in Europe and Africa.

Intra-network accuracy was calculated with a median of ~ 25 km within Europe,

---

[5]The cross-correlation will line up the zero-crossings of each waveform, providing a phase difference offset; hence, in this context, it is appropriate to use the phase velocity.

using Spanish National Lightning Network data as reference [*Chronis and Anagnostou*, 2003]. *Chronis and Anagnostou* [2006] evaluated the location accuracy using the Lightning Imaging Sensor (LIS) [*Christian et al.*, 1999] data, to find a median error of ~20 km.

A recent paper [*Morales et al.*, 2007] introduces a technique for estimating the polarity of strokes reported by the Zeus network. In addition to recording the VLF waveform from individual sferics, each sensor also records an ELF waveform (from 100–1000 Hz). Both records are transmitted back to the central processor. The VLF waveform contributes to the geo-location algorithm as usual. Once a strike location is determined, a simple ELF propagation model is used to generate a predicted ELF response at each sensor for a reference polarity. These modeled results are then correlated with the measured ELF waveforms. The correlation results are combined using a weighted average and the polarity is then estimated by the sign of this combination. A quality factor, estimated from the quality of the cross-correlations, is also associated with the result. For distances out to 2000 km from the ZEUS network center in Africa, ~70% of the strokes had a medium to high quality factor, indicating a high confidence in the polarity estimate. This figure dropped below 10% at 10,000 km from the network center. No direct comparison is made in this thesis between the performance of the polarity estimation technique introduced in Chapters 4 and 5 and the polarity estimation technique introduced to augment the ZEUS network. One advantage of the method in this thesis is that it uses the same VLF waveforms used to geo-locate the lightning stroke, whereas the ZEUS method relies on a sufficiently large ELF response.

The ATD method of geo-locating lightning strikes has two key disadvantages. First, a constant differential propagation speed must be assumed between each station. As the characteristics of the waveguide change or are different between the strike position and two receivers, this is not a safe assumption. The resulting bias errors may be eliminated using empirical offsets derived from redundant measurements, as was done for several test cases in *Lee* [1990], but this may potentially lead to a large number of necessary correction factors to account for all possible storm locations and receiver configurations. Second, the lightning waveforms from the same

strike must have a similar shape. If the channel between the source and two different receivers is very different—by, for example, a vastly different length or an ionospheric reflection height—the shapes may not be similar, compromising the ATD method of geo-location. This would explain the poor location accuracy for the widely-spaced STARNET and Zeus systems. To compensate for different propagation paths, *Lee* [1990] proposes that a channel response compensation be applied to each sferic. This method has not yet been fully implemented, and it may face problems when correcting for sferics with amplitude nulls due to modal interference or long propagation paths.

1.3.2 WWLLN

The ATD networks rely on cross-correlation of individual waveforms, which requires each receiver to send back the whole waveform. *Dowden et al.* [2002] introduces another approach which captures the arrival time on each sferic independently so that only one number need be sent back to the central processor for each sferic. This method measures an averaged group delay to estimate the arrival time.

For a single propagating mode (see Chapter 2), the phase of any field component is $\phi(t, \omega) = \omega t - \beta(\omega) d + \phi_0(\omega)$, where β is the waveguide mode number and $\phi_0(\omega)$ captures constant terms and has a frequency dependence on the source waveform. Dowden assumes that ϕ_0 is a constant, which is well-justified at VLF since most of the return stroke power originates in the lower 2 km of the return stroke [*Lee*, 1989]. Under this assumption, the derivative of the measured phase gives

$$\frac{d\phi}{d\omega} = t - r\frac{d\beta}{d\omega} = t - \frac{d}{v_g(\omega)} \tag{1.10}$$

where $v_g(\omega)$ is the frequency-dependent group velocity. The time at which $d\phi/d\omega = 0$ therefore gives the group arrival time assuming a group velocity v_g at a particular frequency ω. Dowden uses an averaged time of group arrival (TOGA) by estimating a single derivative using a linear fit to $\phi(\omega)$ in the range $6-22$ kHz. The arrival time difference in (1.9) is then calculated using the averaged time of group delay with a velocity $v = \widehat{v}_g$, where $\widehat{v}_g$ corresponds to a representative group velocity in the middle

of the detection frequency range.

This technique for determining the arrival time forms the basis of the World Wide Lightning Location Network (WWLLN), which uses vertical electric field antennae to preferentially measure the transverse magnetic polarization. WWLLN has been periodically upgraded since it's first inception in 2002 [*Rodger et al.*, 2004, 2005; *Rodger et al.*, 2009] and included 32 receivers as of July 2008. By comparing with events reported by the Los Alamos Sferic Array (LASA), *Jacobson et al.* [2006] report a spatial accuracy of ∼15 km with a detection efficiency of ∼4% for strikes with a peak current over 30 kA. *Rodger et al.* [2009] report an algorithm upgrade which boosts the detection efficiency to ∼35% for strikes stronger than 100 kA, dropping to ∼20% for 50 kA strikes, and then dropping below a 10% detection efficiency for strikes weaker than 25 kA.

1.3.3   Other Lightning Location Networks

The previous two sections covered long-range lightning geo-location networks that measured VLF fields to achieve long-range detection. Many other networks exist that measure other portions of the electromagnetic spectrum, each with an associated set of benefits and trade-offs.

The NLDN consists of over 100 sensors which measure the arrival time, arrival azimuth, or both using the VLF/LF portion of the ground wave from individual lightning strokes *Cummins et al.* [1998b,a]. The arrival azimuth is measured using a broadband magnetic direction finding technique developed in *Krider et al.* [1976] where the angle is extracted from the early part of the ground wave. This portion is excited by the early, more vertical portion of the ground stroke and so minimizes polarization errors. The arrival time is also extracted from the rising edge of the ground wave. Using these techniques, the arrival time and azimuth are measured with accuracies of ∼1.5 μs and ∼1°, respectively, achieving a 50th percentile location range of ≲400 m [*Cummins et al.*, 1998a], with a ground flash detection efficiency of ∼90%.

In an effort to extend the range of the NLDN, four sensors have been deployed on

northern Pacific islands to form the Pacific Lightning Detection Network (PacNet) [*Pessi et al.*, 2009]. These sensors have a higher gain compared to their NLDN sensor counterparts to increase their sensitivity and therefore detection radius. The sensor data from the PacNet is integrated with NLDN and the Canadian Lightning Detection Network (CLDN) to form a larger Long-range Lightning Detection Network (LLDN), covering North America and the northern Pacific Ocean. By comparing geo-location results with lightning data from NASA's lightning imaging sensor (LIS) on board the Tropical Rainfall Measuring Mission (TRMM, see below) and performance models which account for sferic attenuation rates, an assumed distribution of peak current, and an amplitude-dependent sensor detection efficiency, the daytime and nighttime flash detection efficiencies were estimated at 17%–23% and 40%–61%, respectively. As with the NLDN, arrival time estimates were determined using an amplitude threshold mechanism. The authors provide a detailed exploration of the amplitude and time-of-arrival behavior of the ground-wave and first two ionospheric "hops". However, their location algorithm did not distinguish between the various trigger features, and distance-dependent timing errors were not considered. The overall median location accuracy (LA) was between 13–40 km.

Moving to a higher operating frequency range, an arrival time network operating in the VHF range can resolve the 3-D structure of the electrical breakdown paths in the lightning flash. This technology was developed into an operational Lightning Mapping Array (LMA) [*Thomas et al.*, 2000], which can resolve simultaneous leader branches from the main charge region to the upper charge region (cloud flashes) or to the ground (CG flashes). These networks provide extremely high resolution, but have a limited coverage area limited by line-of-sight propagation.

Moving now into the optical band, the Optical Transient Detector (OTD) was launched in 1995 on a low earth orbit satellite on a 70° inclination. Through the year 2000, the optical sensor provided a field of view of 1300×1300 km and provided ∼10 km spatial resolution and ∼50% detection efficiency [*Christian et al.*, 2003]. While this space-based lightning geo-location approach has the advantage of covering the entire planet within the orbit path, there are some significant drawbacks at this time. The CCD records transient optical events, and so can not distiguish between a

cloud and and a ground stroke. Also, the satellite can only monitor a given spot on earth for 15 minutes, returning to the same spot at the same time of day every 55 days. Nevertheless, over the initial 5-year duration of OTD measurements, global diurnal and seasonal variations were constructed. A global flash rate of 45±5 flashes/second was estimated, as well as a land/ocean flash ratio of ∼10:1 [*Christian et al.*, 2003].

Another very similar optical sensor was placed on board the TRMM, which launched in 1997 with a lower altitude (350 km) and an inclination of 35°. While only the tropical regions may be monitored due to the low inclination, the lower altitude allows 5 km spatial resolution [*Christian et al.*, 1999] and the instrument is still in operation.

1.4   Contributions of this Work

This thesis aims to develop a long-range lightning geo-location network with similar goals as the existing long-range VLF systems. In particular, this work seeks to improve on the location accuracy and detection efficiency of existing long-range systems, and to introduce peak current and polarity measurements, which have not yet been incorporated into any other long-range network using VLF measurements. Like the NLDN network, the proposed network uses a combination of arrival time, arrival azimuth, and amplitude to measure the strike time, peak current, and location.

The contributions realized in this dissertation are as follows:

- The introduction of a nonlinear method for removal of VLF transmitters that pervade the important frequency range of radio atmospherics.

- The establishment and evaluation of the properties of radio atmospheric waveforms over varying distances and propagation path profiles.

- The introduction of a new method for lightning detection and parameter estimation at global distances using a radio atmospheric waveform bank.

- The introduction of a systematic method for measuring and correcting for path-dependent distortions to the radio atmospheric waveforms.

- The application of the new method of lightning detection to develop and demonstrate a long-range lightning geo-location network capable of measuring peak current and polarity, with a simultaneous location accuracy and detection efficiency comparable to existing medium-range commercial networks, and much better than all of the long range networks described above.

Chapter 2

Sferic Propagation

The algorithms outlined in Chapter 5 that achieve long-range and accurate geolocation results hinge on an understanding of how the received waveform shape depends on both the distance and the properties of the path between the lightning flash location and the receiver. The sensor algorithm also makes use of a collection of sferic waveforms that are representative of the dominant variation in received sferic waveforms that the receiver is expected to see. While the results given in this thesis are obtained using a measured set of received waveforms, which are derived in Chapter 4 by using reference stroke locations reported by the NLDN, an understanding of the factors that influence the path response between the source and receiver is needed to produce an informed choice of sample profiles needed to capture this variation. Furthermore, the methods of Chapter 5 may also be used with a modeled set of waveforms, an approach that may prove useful for predicting the propagation effects of paths that lie outside the range of a reference network. In order to lay a foundation for the waveform analysis in Chapter 4 and to serve as a starting point for numerical models that reproduce the range of expected received waveforms, this chapter describes the factors influencing the propagation of sferics in the earth-ionosphere waveguide and derives the functional form of the received sferic waveforms.

In general, at VLF the received waveform at a receiver $r(t)$ is related to the current-moment of a lighting strike $s(t) = I(t)l(t)$ and the impulse response of the

channel between the source and receiver $h(t)$:

$$r(t) = s(t) * h(t) \tag{2.1}$$

In the absence of an atmosphere, the electromagnetic energy from a lightning strike would rely on diffraction to propagate around the globe. Ionizing radiation from the sun and background cosmic radiation, however, maintain free electrons and ions in the relatively diffuse upper atmosphere, forming a plasma [*Sechrist Jr.*, 1977, p. 103]. At VLF/ELF frequencies, the lower portion (called the D region) of this weakly ionized plasma reflects a significant portion of the incident wave, forming the so-called earth-ionosphere waveguide. This waveguide allows more efficient guiding of the radiated energy in the form of a sferic. This chapter considers the properties of the source and transfer function to outline a basic propagation model that assists with the interpretation of measured sferics at various distances from the source location.

2.1 Wave Propagation in a Plasma

The weakly ionized portion of the upper atmosphere is called the ionosphere. The reflection coefficient and the wave propagation properties are strongly dependent on the local properties of the ionosphere. As a result, one must first derive the wave propagation properties in this medium.

Consider a species α (electrons or a particular ion) with mass m_α, charge q_α, and collision rate ν_α, which is the rate of inelastic collisions between a charged particle and the much more abundant neutral species of the gas. Suppose as well that there is a driving electric field $\overline{\mathcal{E}}$ and a static earth magnetic field $\mathbf{B}_\mathrm{E}$ permeating the plasma. Then, the force balance equation describing the motion of the position $\overline{r}_\alpha$ is

$$m_\alpha \frac{\partial^2 \overline{r}_\alpha}{\partial t^2} = q_\alpha \overline{\mathcal{E}} + q_\alpha \frac{\partial \overline{r}_\alpha}{\partial t} \times \mathbf{B}_\mathrm{E} - m_\alpha \nu_\alpha \frac{\partial \overline{r}_\alpha}{\partial t}. \tag{2.2}$$

Multiplying (2.2) by $q_\alpha N_\alpha$, where N_α is the number density per unit volume of species

α, yields an equation relating the current density $\overline{\mathcal{J}}_\alpha = q_\alpha \overline{v}_\alpha N_\alpha$ to the electric field $\overline{\mathcal{E}}$

$$\frac{\partial \overline{\mathcal{J}}_\alpha}{\partial t} = \epsilon_0 \omega_{p\alpha}^2 \overline{\mathcal{E}} + \frac{q_\alpha}{|q_\alpha|} \omega_{H\alpha} \overline{\mathcal{J}}_\alpha \times \mathbf{b}_0 - \nu_\alpha \overline{\mathcal{J}}_\alpha \tag{2.3}$$

where $\overline{v}_\alpha = \partial \overline{r}_\alpha / \partial t$ is the particle velocity. The plasma frequency

$$\omega_{p\alpha} = \sqrt{\frac{N_\alpha q_\alpha^2}{m_\alpha \epsilon_0}} \tag{2.4}$$

describes the natural response rate of the charged particles to a perturbation, the gyro-frequency

$$\omega_{H\alpha} = \frac{B_0 |q_\alpha|}{m_\alpha} \tag{2.5}$$

is the rate of rotation around an ambient magnetic field of strength $B_0 = |\mathbf{B}_\mathrm{E}|$, and $\mathbf{b}_0 = \mathbf{B}_\mathrm{E}/|\mathbf{B}_\mathrm{E}|$ is a unit vector pointing in the direction of the static magnetic field $\mathbf{B}_\mathrm{E}$. In the ionosphere, $\mathbf{B}_E$ is the static magnetic field of the earth (hence, the subscript E), and $\overline{\mathcal{E}}$ is the electric field vector from an electromagnetic wave. The influence on the particle of the magnetic field of the wave is ignored, as its effect is typically much less than the influence of the electric field for particles with velocities much less than the speed of light. An idealized cold plasma, where the charged species are motionless until acted on by the incident wave, has also been assumed.

The propagation of an electromagnetic field in the plasma is governed by the curl terms in Maxwell's equations:

$$\nabla \times \overline{\mathcal{E}} = -\frac{\partial \overline{\mathcal{B}}}{\partial t}; \qquad\qquad \nabla \times \overline{\mathcal{B}} = \mu_0 \epsilon_0 \frac{\partial \overline{\mathcal{E}}}{\partial t} + \mu_0 \overline{\mathcal{J}}_\mathrm{tot} \tag{2.6}$$

where $\overline{\mathcal{J}}_\mathrm{tot} = \sum_\alpha \overline{\mathcal{J}}_\alpha$, $\overline{\mathcal{B}} = \mu_0 \overline{\mathcal{H}}$, and μ_0, ϵ_0 are the permeability and permittivity of free space, respectively.

Suppose now a time-harmonic solution where fields vary as $\exp[j(\omega t - \mathbf{k} \cdot \mathbf{r})]$, where $\mathbf{k}$ is the wave vector: the phase is constant over planes represented by the equation $\mathfrak{Re}\{\mathbf{k} \cdot \mathbf{r}\} = \text{constant}$ and the amplitude is constant over planes represented by the equation $\mathfrak{Im}\{\mathbf{k} \cdot \mathbf{r}\} = \text{constant}$. The time-dependent vectors $\overline{\mathcal{E}}$, $\overline{\mathcal{B}}$, and $\overline{\mathcal{J}}$ can then be

replaced with the phasors **E**, **B**, and **J**. One can now define a 3×3 conductivity tensor $\widehat{\sigma}_{\text{tot}}$ such that $\mathbf{J}_{\text{tot}} = \widehat{\sigma}_{\text{tot}}\mathbf{E}$ using (2.3), where $\widehat{\sigma}_{\text{tot}} = \sum_{\alpha} \widehat{\sigma}_{\alpha}$ includes contributions from each species α:

$$\widehat{\sigma}_{\alpha} = (j\omega\epsilon_0)\widetilde{X}_{\alpha}\begin{bmatrix} \dfrac{1}{\widetilde{Y}_{\alpha}^2 - 1} & \dfrac{-jq_{\alpha}/|q_{\alpha}|\widetilde{Y}_{\alpha}}{\widetilde{Y}_{\alpha}^2 - 1} & 0 \\ \dfrac{jq_{\alpha}/|q_{\alpha}|\widetilde{Y}_{\alpha}}{\widetilde{Y}_{\alpha}^2 - 1} & \dfrac{1}{\widetilde{Y}_{\alpha}^2 - 1} & 0 \\ 0 & 0 & -1 \end{bmatrix} \tag{2.7}$$

where for a species α (the subscript α is dropped for brevity):

$$\widetilde{X} = \frac{X}{1 - jZ}; \qquad\qquad \widetilde{Y} = \frac{Y}{1 - jZ}. \tag{2.8}$$

Here X, Y, and Z describe the normalized plasma frequency, gyro frequency, and collision rate, respectively; i.e.,

$$X = \frac{\omega_p^2}{\omega^2}; \qquad\qquad Y = \frac{\omega_H}{\omega}; \qquad\qquad Z = \frac{\nu}{\omega}. \tag{2.9}$$

In (2.7), it was assumed that $\mathbf{B}_{\text{E}} = B_0\widehat{\mathbf{z}}$. The presence of off-diagonal terms results from the anisotropy introduced by this magnetic field. Since X is inversely proportional to the mass, the contributions to the total current from all species other than the electrons can usually be neglected at VLF. For the remainder of this discussion, $\widehat{\sigma}_{\text{tot}} = \widehat{\sigma}$ is therefore assumed to include only the plasma current due to the electrons.

If one now defines a relative permittivity tensor[1] $\widehat{\epsilon}_p = 1 + \widehat{\sigma}/(j\omega\epsilon_0)$, one can write (2.6) as, respectively,

$$\mathbf{k} \times \mathbf{E} = \omega\mathbf{B}; \qquad\qquad \mathbf{k} \times \mathbf{B} = -\omega\mu_0\epsilon_0\widehat{\epsilon}_p \cdot \mathbf{E}. \tag{2.10}$$

---

[1]Some authors define $\widehat{\epsilon}_p = \begin{bmatrix} S & jD & 0 \\ -jD & S & 0 \\ 0 & 0 & P \end{bmatrix}$, where S, D, and P are readily derived from (2.7).

These two curl equations combine to give, with $k^2 \equiv \mathbf{k} \cdot \mathbf{k}$,

$$\mathbf{k}(\mathbf{k} \cdot \mathbf{E}) - k^2 \mathbf{E} + \frac{\omega^2}{c^2} \widehat{\epsilon}_p \cdot \mathbf{E} = 0. \tag{2.11}$$

In a source-free and isotropic medium, where $\nabla \cdot \epsilon_0 \mathbf{E} = 0$ and $\widehat{\epsilon}_p$ is a scalar, (2.11) reduces to the scalar wave equation. Writing (2.11) as

$$\left(\mathbf{k}\mathbf{k}^T - k^2 \mathbf{I} + \frac{\omega^2}{c^2} \widehat{\epsilon}_p \right) \cdot \mathbf{E} = 0 \tag{2.12}$$

the wave solutions may then be found by setting the determinant of the 3×3 matrix $\mathbf{k}\mathbf{k}^T - k^2 \mathbf{I} + (\omega^2/c^2)\widehat{\epsilon}_p$ to 0, yielding the square of the index of refraction $n^2 = k^2 c^2/\omega^2$, given by the Appleton Hartree equation [*Ratcliffe*, 1959, p. 19]

$$n^2 = 1 - \frac{X}{1 - jZ - \dfrac{Y^2 \sin^2 \theta}{2(1 - X - jZ)} \pm \left\{ \left[\dfrac{Y^2 \sin^2 \theta}{2(1 - X - jZ)} \right]^2 + Y^2 \cos^2 \theta \right\}^{1/2}} \tag{2.13}$$

where θ is the angle between the wave vector $\mathbf{k}$ and the magnetic field $\mathbf{B}_{\mathrm{E}}$. For a given wave direction, there will be two solutions for n.

If the effect of collisions dominate over the effect of the magnetic field, i.e., if $\nu \gg \omega_H$ or, equivalently, if $Z \gg Y$, then the conductivity tensor may be approximated as a scalar

$$\sigma^{\mathrm{isotropic}} = \frac{\epsilon_0 \omega_p^2}{\nu + j\omega}. \tag{2.14}$$

If, in addition, $\nu \gg \omega$, so that $Z \gg 1$, then $\sigma = \epsilon_0 \omega_p^2/\nu$ and $n^2 = 1 - j\omega_r/\omega$, where $\omega_r = \omega_p^2/\nu$, i.e., the effective conductivity is $\sigma_{\mathrm{effective}} = \epsilon_0 \omega_r$.

Figure 2.1a shows representative curves for the electron density for a daytime and nighttime profile, calculated using the model in *Lehtinen and Inan* [2007] which incorporates an atmospheric chemistry model with the International Reference Iono-sphere (IRI) [*Bilitza*, 2001]. Also shown is the electron-neutral collision rate, which is proportional to the atmospheric density [*Schunk and Nagy*, 1980]. Figure 2.1b shows the corresponding X, Y, and Z curves at 15 kHz. The condition $Z \gg 1$ is satisfied
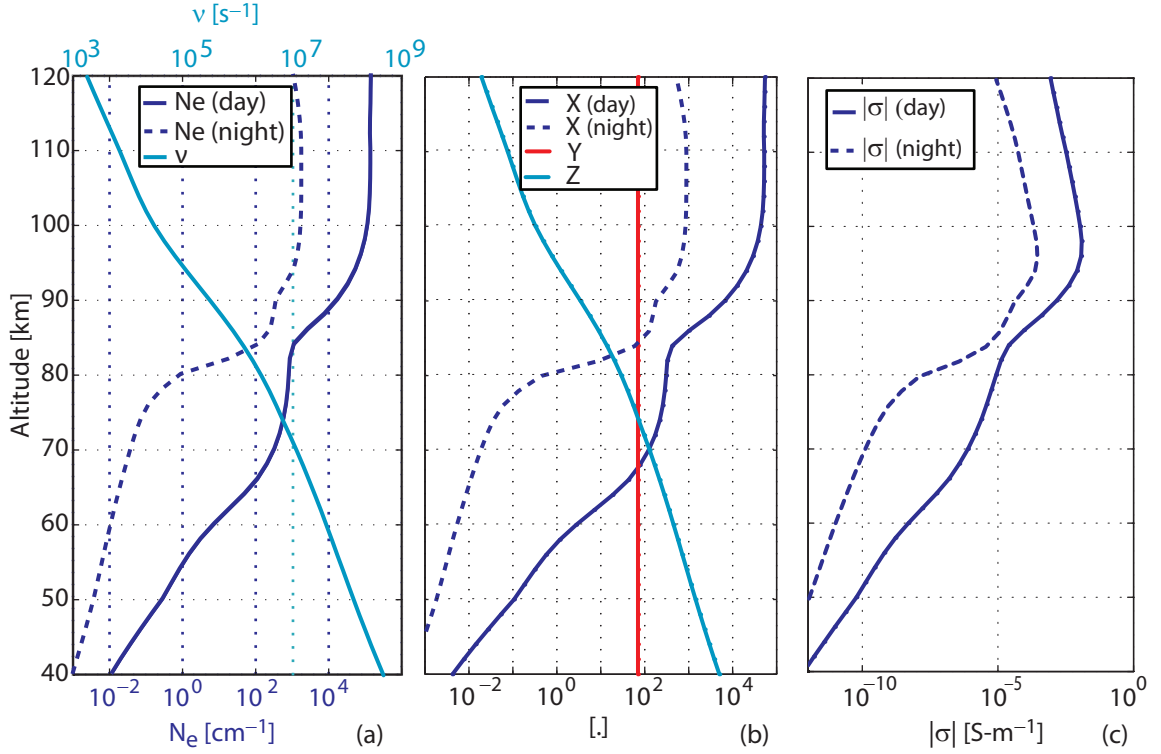
Figure 2.1: Day and night ionospheric profiles. (a) Electron density N_e and collision rate ν. (b) X, Y, and Z parameters at $f_0=15$ kHz. (c) Effective ionospheric conductivity at $f_0=15$ kHz.

for altitudes less than ~ 90 km, but the assumption $Z \gg Y$ is invalid except at very low altitudes. Nevertheless, useful insight may be gained by assuming an isotropic ionosphere. Under isotropic conditions, at VLF, *Ratcliffe* [1959, p. 92] shows that reflection occurs approximately at a height where $X=Z$. Below this point, the collisions keep $|n|$ from deviating too much from unity and the wave passes through (though attenuate due to the imaginary component of n). For an in-depth discussion of the reflection process in a lossy ionosphere, see *Budden* [1961b, Ch. 16]. During the nighttime ionosphere, $X=Z$ occurs at ~ 84 km; during daytime, at ~ 74 km. Due to the higher collision rate at the lower reflection altitude for the latter case, a lower reflection coefficient for the daytime ionosphere is expected.

2.2 Reflection from a Boundary

The boundaries of the earth-ionosphere waveguide affect the wave propagation through the reflection coefficient relating the field of the reflected wave to the field of the incident wave. This reflection coefficient depends on the polarization of the incident wave. For a boundary between free space, with index of refraction 1, and an isotropic sharp boundary with index of refraction n, the reflection coefficient of an incident wave whose electric field is respectively parallel and perpendicular to the plane of incidence is [*Budden*, 1961a, p. 91]

$$R_\parallel = \frac{n\cos\theta - \cos\psi}{n\cos\theta + \cos\psi}; \qquad\qquad R_\perp = \frac{\cos\theta - n\cos\psi}{\cos\theta + n\cos\psi} \qquad (2.15)$$

where θ is the angle of incidence measured from the perpendicular line to the plane, and the reflected wave angle ψ is related to θ by Snell's law [*Budden*, 1961a, p. 88]

$$\sin\theta = n\sin\psi, \qquad (2.16)$$

which allows us to eliminate ψ in (2.15). $R_\parallel$ describes the reflection coefficient for the perpendicular (to the plane of incidence) magnetic field H_y or the vertical electric Hertz vector[2] U_z (see below); $R_\perp$ specifies the reflection coefficient for the perpendicular electric field E_y or the vertical magnetic Hertz vector V_z.

The ionosphere is not a sharp boundary, however, and so one must consider a varying wave vector with altitude. Define $q = n\cos\psi$ so that

$$n^2 = q^2 + \sin^2\theta. \qquad (2.17)$$

If there is no reflection as the wave passes through the medium, which would be the case if the medium is slowly varying with respect to the wavelength (the so-called Wentzel-Kramers-Brillouin (WKB) approximation [*Budden*, 1961b, p. 129]), the field

---

[2]Some authors define $R_\parallel$ with the perpendicular component of H switching signs after reflection from the boundary. See, for example, *Inan and Inan* [2000, p. 177].

varies in a fashion which exhibits phase memory:

$$F = F_0 \exp\left[-ik\left(x\sin\theta + \int_0^z q dz'\right)\right].$$ (2.18)

This variation suggests a reflection coefficient of the form

$$R = \exp\left[-2jk\int_0^{z_0} q dz\right]$$ (2.19)

where z_0 is the reflection height. But the WKB approximation breaks down where $q \simeq 0$. In *Budden* [1961b, p. 319], q^2 is treated as a linear function of z over a small range to express the required solution as an Airy Integral function; under this assumption *Budden* [1961b, p. 325] shows that

$$R = j \exp\left[-2jk\int_0^{z_0} q dz\right]$$ (2.20)

where q is complex and the contour integral path over z terminates at a point where $q = 0$.

The WKB approximation is only valid when the transition length of the refractive index is large compared to a wavelength. In our application at VLF, the wavelengths are long compared to the distance over which ionosphere changes from a dielectric to an effective conductor. In this case, for a stratified reflection boundary, a full-wave numerical solution must be used. A numerical approach for an isotropic ionosphere is discussed in *Wait and Walters* [1963], and a more general full-wave approach for an isotropic ionosphere is discussed in *Budden* [1985, Ch. 18], and, more recently, in *Lehtinen and Inan* [2008]. In all cases, the reflection coefficient is calculated through an iterative procedure. At the uppermost boundary, it is assumed that $R=0$. At each slab below this level, continuity of the transverse field components (or, equivalently, impedance matching) is used to calculate the reflection coefficient in terms of the reflection coefficient of the previous slab. A simple case for a homogeneous ionosphere that is used later in this thesis is described in *Wait* [1970, p. 405]. For the benefit of the reader, an outline of the more general procedure, taken from *Lehtinen and Inan*

[2008], is provided here for an anisotropic ionosphere.

The reflecting medium is divided into anisotropic but homogeneous slabs where each slab has a thickness chosen to capture the changes in the permittivity tensor. Each slab i is therefore characterized by $A\widehat{\epsilon}_p^i A^{-1}$ where A is a rotation matrix to account for the deviation of the magnetic field from the $\widehat{\mathbf{z}}$ direction. Defining $\mathbf{n} = \mathbf{k}c/\omega$, the solution to (2.12) may be written

$$\left|n_i^2\mathbf{I} - \mathbf{n}_i\mathbf{n}_i^T - \widehat{\epsilon}_p^i\right| = 0 \tag{2.21}$$

where, assuming without loss of generality that the wave is propagating in the $\widehat{\mathbf{x}}$ direction, $\mathbf{n} = [n_x, 0, q]^T$, where q is the $\widehat{\mathbf{z}}$ component of $\mathbf{n}$ as before and n_x remains constant due to Snell's Law. Solving (2.21) for q leads to four solutions (the so-called "Booker Quartic"), which may be ordered into two upward and two downward propagating waves according to the sign of the imaginary component [*Lehtinen and Inan*, 2008]. By requiring the horizontal $\mathbf{E}$ and $\mathbf{H}$ fields to remain continuous across each boundary and noting that there is no downgoing field above the top boundary, the reflection matrix (2.40) may be calculated through an iterative procedure (below the bottommost slab of the ionosphere, the coordinate system needs to be set so that the matrix elements align with the TM and TE field elements, defined below).

2.3 Ray-hop Analysis

At close distances the time separation between successive ionospheric reflections is great enough that the waveform may be modeled as a summation of a series of ray hops. The received waveform consists of the source waveform convolved with an impulse response of the channel, consisting of a ground wave component and several ionospheric reflections.

Figure 2.2 illustrates the various components in the ray-hop model. The ground wave diffracts over the earth surface and has the shortest propagation distance. This wave is followed by a succession of sky waves. The total received field consists of the source term convolved with the impulse response associated with the ground wave
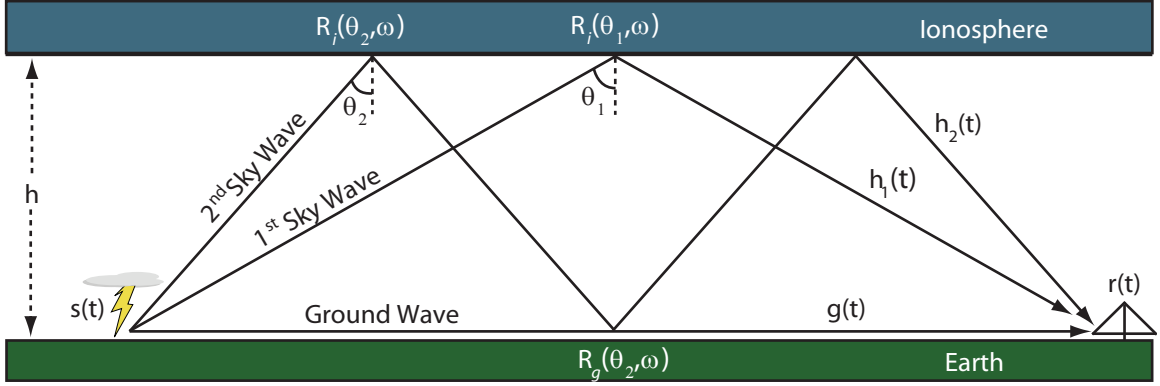
Figure 2.2: Illustration of the ground and multiple sky waves for a flat earth.

and each successive sky wave:

$$r(t) = s(t) * \left[g(t) + \sum_{m=1}^{\infty} h_m(t) \right], \tag{2.22}$$

where $g(t)$ represents the ground wave propagation and $h_m(t)$ represents the propagation for sky hop m.

A sample waveform is plotted in Figure 2.3, showing the various components of the received waveform, including the ground wave and the successive sky waves. The ground wave, which has the most direct path, is seen arriving first, followed by the first sky wave, the second sky wave, and so on. The inversion of the first sky wave is discussed in Section 4.2.2.

Due to its simple functional dependence on an electrical source, it is useful to consider a Hertz vector $\overline{u}$, defined so that

$$\overline{\mathcal{E}} = -\mu \frac{\partial^2 \overline{u}}{\partial t^2} + \frac{1}{\epsilon} \nabla (\nabla \cdot \overline{u}), \tag{2.23}$$

$$\overline{\mathcal{B}} = \mu \frac{\partial}{\partial t} \nabla \times \overline{u}. \tag{2.24}$$

Consider a vertically oriented dipole at the origin with dipole moment

$$\mathcal{M} = \int I l \, dt \tag{2.25}$$
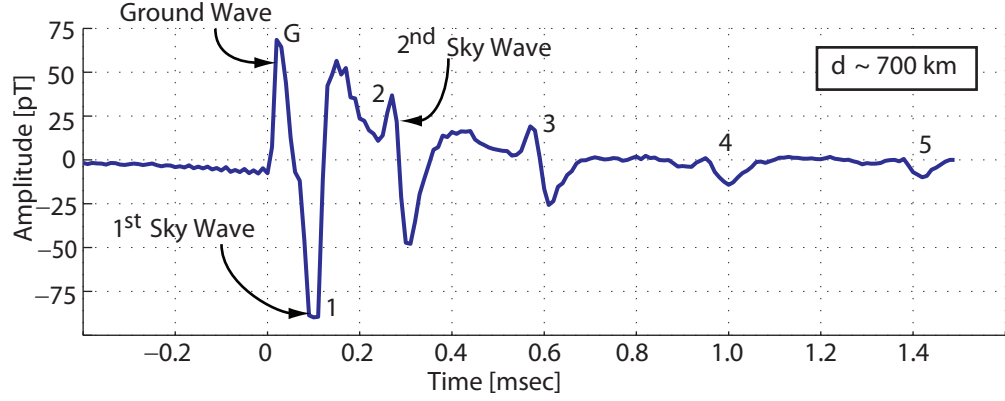
Figure 2.3: Sample sferic that has propagated 700 km under a nighttime ionosphere. The ground wave and subsequent ionospheric reflections are labeled on the plot.

where l is the length and I is the current in the dipole (note that many authors begin with the current moment $\mathcal{M}_c = i(t)l(t)$ instead of the dipole moment $\mathcal{M}$). If the length does not change, then also $\mathcal{M} = Ql$, where Q is the dipole charge. If $\mathcal{M} = M_0 \exp[j\omega t]$, then one may replace $\mathcal{M}$ with the phasor $M = (Il)/(j\omega)$. In a spherical coordinate system with the source at the origin and the distance, elevation, and azimuthal coordinates denoted by r, θ, and ϕ, respectively, the resulting Hertz vector for a wave from this source is [*Budden*, 1961a, p. 43]

$$\mathbf{U} = U_z\widehat{\mathbf{z}} = \frac{M \exp[-jkr]}{4\pi r}\widehat{\mathbf{z}}. \qquad (2.26)$$

In free space (2.23) and (2.24) may be used to find the electric and magnetic field values. The horizontal magnetic field, for example, is

$$\mathcal{B}_\phi = \frac{\mu_0}{4\pi}\cos\theta\left(\frac{\dfrac{\partial[\mathcal{M}]}{\partial t}}{r^2} + \frac{\dfrac{\partial^2[\mathcal{M}]}{\partial t^2}}{cr}\right) \qquad (2.27)$$

where the first term is an induction term and the second is a radiation term (the expression for E_θ is simply scaled by c and includes an electrostatic term proportional to $[\mathcal{M}]/r^3$). Here $[\mathcal{M}]$ contains the dipole moment expressed as a function of the time retarded variable $t' = t - r/c$. For $kr \gg 1$ only the radiation term needs be considered.

In the *Bruce and Golde* [1941] model, the current is constant from the ground up to the highest altitude point of the return stroke. Under this assumption and keeping only the radiation term, the azimuthal magnetic field is [*Uman*, 2001, p. 332]

$$\mathcal{B}_\phi = \frac{\mu_0 \cos\theta}{4\pi cr} \frac{d\left[l(t')i(t')\right]}{dt'}, \tag{2.28}$$

where $l(t') = \int_0^{t'} v(\tau)d\tau$ and v is the velocity of the return stroke up the channel. If one assumes a time-harmonic dependence then $[\mathcal{M}] = M_0 \exp\left[j\omega(t - r/c)\right]$ contains a phase retardation term to account for the propagation distance r. Over a perfectly conducting Earth the horizontal magnetic field component measured on the ground ($\theta = 0$, with $d = r$ as the propagation distance) is multiplied by 2 and becomes (in the time-harmonic analysis)

$$B_\phi = \frac{j\mu_0 Ilk}{2\pi d}\mathrm{e}^{j\omega(t-d/c)} = \frac{j\mu_0 Ilf}{dc}\mathrm{e}^{j\omega(t-d/c)} \equiv B_0, \tag{2.29}$$

where here k is the free-space wavenumber $k = \omega/c$. In this equation, Il is the Fourier Transform of the current-channel length product: $Il = \mathcal{F}\{i(t)l(t)\}$; the source- and distance-dependent B_0 in (2.29) is defined to simplify the expression of the total field equation below.

In reality, the ground has a finite conductivity and is curved according to the spherical shape of the earth, and so a correction factor is needed. Beginning with a process similar to the one outlined in Section D.2 to calculate the resulting field from a time harmonic source on a spherical surface[3], *Wait* [1956] derives the ratio W between the ground wave component and a field from an identical source over a flat perfectly conducting earth, shown here for the special case where the transmitter and

---

[3]Section D.2 contains a transformation between a series and a contour integral to arrive at a more rapidly converging mode series. *Wait* [1956] employs an additional transformation to capture the behavior of the ground wave in a minimum number of terms, leading to (2.30).

receiver are both on the surface:

$$W = F_0(p) - \frac{\delta^3}{2} \left[1 - j(\pi p)^{1/2} - (1 + 2p)F_0(p) \right]$$

$$+ \delta^6 \left[1 - j(\pi p)^{1/2}(1 - p) - 2p + \frac{5}{6}p^2 + \left(\frac{p^2}{2} - 1 \right) F_0(p) \right] \qquad (2.30)$$

$$+ [\text{terms in } \delta^9, \delta^{13}, \text{ etc.}]$$

where

$$p = -jkd\frac{\Delta^2}{2}; \qquad \delta^3 = \frac{j}{kR_E\Delta^3}; \qquad F_0(p) = 1 - j(\pi p)^{1/2}\text{erfc}(jp^{1/2})e^{-p} \qquad (2.31)$$

and, for a boundary between two homogeneous mediums with propagation constants γ_0 and γ_1 [*Saxton*, 1964, p. 179],

$$\Delta = \frac{Z}{\eta_0} = \frac{\gamma_0}{\gamma_1} \left(1 - \frac{\gamma_0^2}{\gamma_1^2} \right)^{1/2}. \qquad (2.32)$$

The variable Z denotes the tangential surface impedance, taken as the ratio of the tangential electric and magnetic fields for a vertically polarized plane wave at grazing incidence, and $\eta_0 = \sqrt{\mu_0/\epsilon_0}$ is the impedance of free space. If, for example, region 0 is free space ($\gamma_0 = jk$) and region 1 has conductivity σ and permittivity ϵ so that $\gamma_1 = \sqrt{j\omega\mu(\sigma + j\omega\epsilon)}$ [e.g., *Inan and Inan*, 2000], then (2.32) may be readily calculated using $\gamma_0/\gamma_1 = [j\omega\epsilon_0/(\sigma + j\omega\epsilon)]^{1/2}$. In the context of (2.22), assuming the quantities in (2.30) are all frequency-dependent, $s(t) = \mathcal{F}^{-1}\{(I(f)l(f))\}$ and $g(t) = \mathcal{F}^{-1}\{B_0W/(Il)\}$, where $\mathcal{F}^{-1}$ denotes the Inverse Fourier Transform (IFT).

A common model for VLF measurements is the *Bruce and Golde* [1941] model, where the current and velocity are modeled as a double and single exponential, respectively

$$i(t) = i_0(e^{-\alpha t} - e^{-\beta t}); \qquad\qquad v(t) = V_0 e^{-at}. \qquad (2.33)$$

For a transmitter and receiver both on the ground ($z = 0$) and assuming a vertical dipole source, the mth sky hop may be generally expressed as [adapted from *Watt*,

1967, p. 204]

$$B_{\phi,m} = B_0 \cos\psi F_t F_r c_m R_m(\phi,\psi) \frac{\mathrm{e}^{-jk\Delta d}}{1 + \Delta d/d} \tag{2.34}$$

where ψ is the launching angle at the ground and ϕ is the incident angle at the ionosphere. The $\cos\psi$ term accounts for the antenna factor associated with the orientation of the transmitting dipole. F_t and F_r are transmitter and receiver correction factors, respectively, which account for the effect the finite conductivity at the source and receiver has on the effective antenna gain and may be approximated by $F = 1 + R_\parallel^g$ [*Watt*, 1967, p. 205]. The factor c_m is a focusing term which takes account of the geometry of a spherical earth (see Appendix D). It contains a factor which accounts for a focusing effect from reflecting off an upper curved boundary and a horizontal spreading term which transforms the d^{-1} dependence in B_0 to the spherical spreading term $[a \sin(d/a)]^{-1}$, where a is the radius of the earth. The R_m term accounts for reflections off of the ionosphere and the ground. If an anisotropic ionosphere is considered with a vertical source and if it is assumed that only the vertical term is measured, then, with m ionospheric reflections and $m-1$ ground reflections, R_m is taken as the parallel polarization term (associated with a vertically-polarized wave) of the product of successive ionospheric and ground reflection matrices (equations (2.40) and (2.41) explicitly write out the ionospheric and ground reflection matrices, respectively)

$$R_m(\phi,\psi) = \begin{bmatrix} 1 & 0 \end{bmatrix} \mathbf{R}_i^m(\phi) \mathbf{R_g}^{m-1}(\psi) \begin{bmatrix} 1 \\ 0 \end{bmatrix} \tag{2.35}$$

Note that there may also be a contribution to the measured horizontal magnetic field from the orthogonal polarization. With an isotropic ionosphere, perfectly conducting earth, and a vertically oriented source, this expression contains only the reflection term for a vertically polarized incident and reflected wave: $R_m = {}_\parallel R_\parallel^m$. The final term $\mathrm{e}^{-jk\Delta d}/(1 + \Delta d/d)$ accounts for the extra phase and spread due to the extra distance Δd traveled along the ray hop compared to the ground wave distance d:

$\Delta d = d_{s,m} - d$, where $d_{s,m}$ is the total distance traveled along the mth ray hop

$$d_{s,m} = 2n \left[R_{\text{E}}^2 + (R_{\text{E}} + h)^2 - 2R_{\text{E}}(R_{\text{E}} + h) \cos \left(\frac{d}{2nR_{\text{E}}} \right) \right]^{1/2} \tag{2.36}$$

$$\simeq \left[(2nh)^2 + d^2 \left(1 + \frac{h}{R_{\text{E}}} \right) \right]^{1/2}. \tag{2.37}$$

The impulse response $h_m(t)$ in (2.22) is simply $\mathcal{F}^{-1} \{B_{\phi,m}/(Il)\}$. Equation (2.34) is used to construct a basic wave-hop model in Section 4.2.2.

2.4 Modal Analysis

At larger distances from the source, many terms in the summation in (2.22) are necessary as the differential distance between each ray hop decreases and the reflection coefficient for subsequent hops increases. It is therefore useful to transform the problem to a summation of a sequence of modes, which may converge to a solution more rapidly.

2.4.1 Mode Condition

Consider a time harmonic source with time dependence $\exp [j\omega t]$. Figure 2.4 depicts a steady-state solution as a superposition of up- and down-going plane waves incident at an angle θ_m to the normal of the boundary. The functional form for any of the fields is [*Budden*, 1961a, p. 115]

$$\begin{aligned}
F_1 &= F_0 \exp[-jk(x \sin \theta + z \cos \theta)] \\
F_2 &= F_0 R_i \exp[-jk(x \sin \theta - z \cos \theta)] \exp[-2jkh \cos \theta] \\
F_3 &= F_0 R_g R_i \exp[-jk(x \sin \theta + z \cos \theta)] \exp[-2jkh \cos \theta]
\end{aligned} \tag{2.38}$$

where $\exp [-2jkh \cos \theta]$ is included so $F_2/F_1 = R_i$ at $z = h$. Since the second reflection should recover the up-going plane wave, one can set $F_1 = F_3$ to arrive at the modal
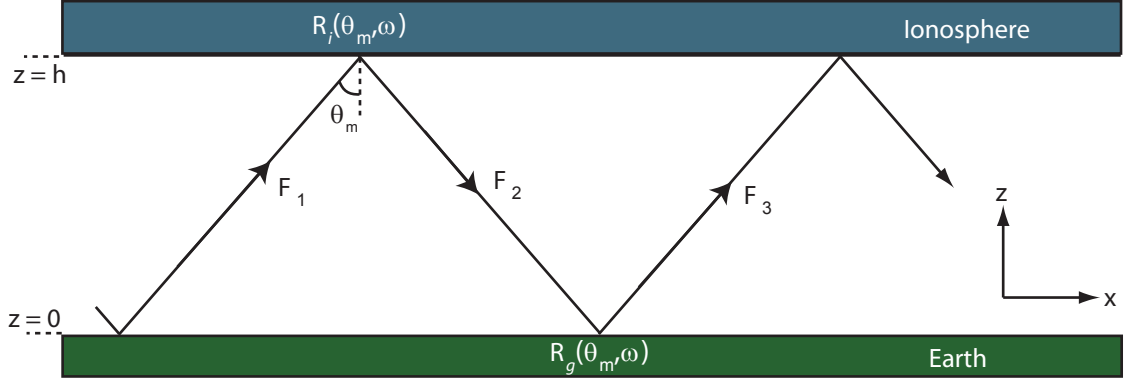
Figure 2.4: Depiction of a waveguide mode as the superposition of an upgoing plane wave F_1 and a downgoing plane wave F_2.

equation for the eigenangle θ_m

$$R_i(\theta_m)R_g(\theta_m)e^{-2jkh\cos\theta_m} = 1 = e^{j2\pi m}. \tag{2.39}$$

In this formulation, R_g is referenced to the ground and R_i is referenced to a height $z = h$. The reflection coefficients are given by (2.15) and depend on the incident polarization. For a smoothly varying ionospheric height this same modal equation may be used by choosing a suitably low reflection height. The waveguide wavenumber describing the phase advance down the waveguide is $\beta = k\mathfrak{Re}\{\sin\theta_m\}$ and the attenuation rate in nepers per unit length is $\alpha = -k\mathfrak{Im}\{\sin\theta_m\}$. The phase and group velocity are, respectively, $v_p = \omega/\beta$ and $v_g = d\omega/d\beta$.

Useful intuition may be gained by assuming idealized boundaries. For example, with perfectly conducting boundaries so that $R_i = R_g = 1$, which provides a crude approximation of the earth-ionosphere waveguide with a steep (nearly vertical propagation) eigenangle, $\cos\theta_m = m\pi/(kh)$ and so $\beta = k\sqrt{1 - [m\pi/(kh)]^2}$, where, assuming free-space between the two waveguide boundaries, $k = \omega/c$. The cutoff frequency, defined as the frequency where $\beta = 0$ as β transitions from a real (propagating wave) to an imaginary (evanescent wave) quantity, may be used to approximate the ionospheric height; in Figure 2.6, for example, the cutoff frequency of 1750 Hz gives a

nighttime ionospheric reflection height of $c/(1750\cdot2)$=85.7 km. With $R_g = R_i = 1$, F_1 and F_2 in (2.38) may be combined to recover the functional form of the z-dependence in the waveguide at $\theta = \theta_m$; for example, $H_y \sim \cos(kz\cos\theta_m) = \cos(zm\pi/h)$. This altitude dependence is the so-called height-gain function. For the first mode, at $z = h/2$, $H_y = 0$, so that a receiver (transmitter) at this altitude would not be able to receive (transmit) a H_y field component; the transmitter or receiver would be most effective at the boundaries.

If the upper boundary has a reflection coefficient of -1, corresponding to a perfect magnetic conductor (which is a better approximation for the earth-ionosphere waveguide for shallow angles, e.g., for lower order modes away from cutoff), then the modal condition becomes $\cos\theta_m = (m - 1/2)\pi/(kh)$.

In a parallel plate waveguide with isotropic boundaries, the field solution may be partitioned into two independent solutions. The geometry of Figure 2.5 assumes a waveguide with infinite extent in the $\widehat{\mathbf{y}}$ direction, so $\frac{\partial}{\partial y}(\ldots) = 0$. Under this assumption and in a source-free region, in Cartesian coordinates (2.6) expands to (noting $B = \mu H$):

$$\nabla \times \mathbf{E} = -j\omega\mu\mathbf{H} \qquad\qquad \nabla \times \mathbf{H} = j\omega\varepsilon\mathbf{E}$$

$$-\frac{\partial E_y}{\partial z} = -j\omega\mu H_x \qquad\qquad -\frac{\partial H_y}{\partial z} = j\omega\varepsilon E_x$$

$$\frac{\partial E_x}{\partial z} - \frac{\partial E_z}{\partial x} = -j\omega\mu H_y \qquad\qquad \frac{\partial H_x}{\partial z} - \frac{\partial H_z}{\partial x} = j\omega\varepsilon E_y$$

$$\frac{\partial E_y}{\partial x} = -j\omega\mu H_z \qquad\qquad \frac{\partial H_y}{\partial x} = j\omega\varepsilon E_z$$

which can be expanded into two sets of independent differential equations, one involving E_z, H_y, E_x, the other H_z, E_y, H_x. Solutions involving the first set are called Transverse Magnetic (TM) waves, so named because the magnetic vector is strictly transverse to the direction of propagation. TM waves have an electric field in the plane of incidence to the boundary, so the reflection coefficient is given by $R_\parallel$. Solutions to the second set are called Transverse Electric (TE) waves, which have an electric field strictly perpendicular to the plane of incidence and are thus governed by $R_\perp$. For a vertical electric source such as the base of a CG lightning strike, the TM
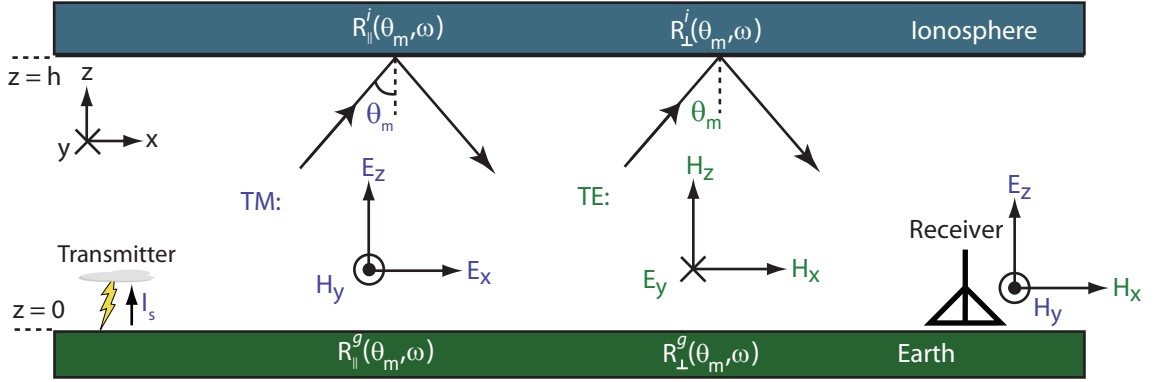
Figure 2.5: Field components for the TM and TE mode polarizations.

mode is preferentially excited.

All measurements used in this thesis were obtained using magnetic loop antennae which measure H_y and H_x. If only a TM mode were present, then the recorded magnetic field would be perpendicular to the direction of propagation. If a TE mode were additionally present, then the recorded field on each channel would not necessarily have the same phase and a parametric plot may show an ellipse instead of a line at each frequency. A vertical electric field antenna would only measure the TM component of the incident wave.

The presence of earth's magnetic field violates our assumption of an isotropic boundary. An anisotropic boundary couples the two polarizations, resulting in a reflection matrix referenced to an altitude h,

$$\mathbf{R_i}^h(\theta) = \begin{bmatrix} {}_\| R_\| & {}_\| R_\perp \\ {}_\perp R_\| & {}_\perp R_\perp \end{bmatrix}, \qquad (2.40)$$

where the first (second) subscript denotes the incident (reflected) polarization. The ground is isotropic, so the ground reflection matrix contains only diagonal terms:

$$\mathbf{R_g}(\theta) = \begin{bmatrix} {}_\| R_\|^g & 0 \\ 0 & {}_\perp R_\perp^g \end{bmatrix}. \qquad (2.41)$$

The modal solutions for the eigenangles $\{\theta_m\}$ are now found by expanding (2.39)

to matrix form

$$\left|\mathbf{R_g}\mathbf{R_i}^h - \mathrm{e}^{2jkh\cos\theta}\mathbf{I}\right| = 0. \tag{2.42}$$

Since these solutions no longer separate into independent polarizations, they are referred to as Quasi-TM (QTM) and Quasi-TE (QTE) modes.

The above formulation assumed a parallel plate waveguide. In spherical coordinates, (2.6) no longer separates into two independent sets of equations, so that once again one has something akin to QTM and QTE modes. In reality, of course, the earth is curved and has an anisotropic ionosphere.

Figure 2.6 illustrates the ray-hop and modal points of view of the received waveform. An example sferic is shown in (a) which persists for over 20 milliseconds. Two spectrograms are shown in (b) and (c), each produced with a different frequency bin width. The larger frequency bin allows for a higher timing resolution and shows the multiply reflected ionospheric hops. The smaller frequency bin highlights the modal nature of these hops, showing the modal cutoff frequencies for at least the four lowest modes. Sferics that posses this late-time waveform which resolves into multiple modal cutoff frequencies are called tweeks [*Yamashita*, 1978].

2.4.2   Point Source in a Parallel Plate Waveguide

The propagation characteristics, including the attenuation rate, phase velocity, and height-gain functions, were derived in the previous section for a waveguide mode given the boundary reflection coefficients. This work measures sferics from lightning strikes, which, at VLF, may be approximated by a point source. This section therefore shows the derivation of the fields due to a point source, which involves the determination of the excitation efficiency of each mode by such a source. Propagation under an isotropic ionosphere is considered first, followed by the anisotropic case.

Two formulations of this problem commonly found in the literature are from *Budden* [1961a] and in a series of paper by Wait, which are summarized in *Wait* [1970]. Both authors[4] start with the Hertz vector $\mathbf{U}$. Budden's formulation is first

---

[4]Wait uses a similar definition for the vector $\mathbf{\Pi}$, which is related to $\mathbf{U}$ through $\mathbf{\Pi} = \mathbf{U}/\epsilon_0$. To keep the notation consistent, $\mathbf{U}$, which carries the units Coulomb, shall be used exclusively and the
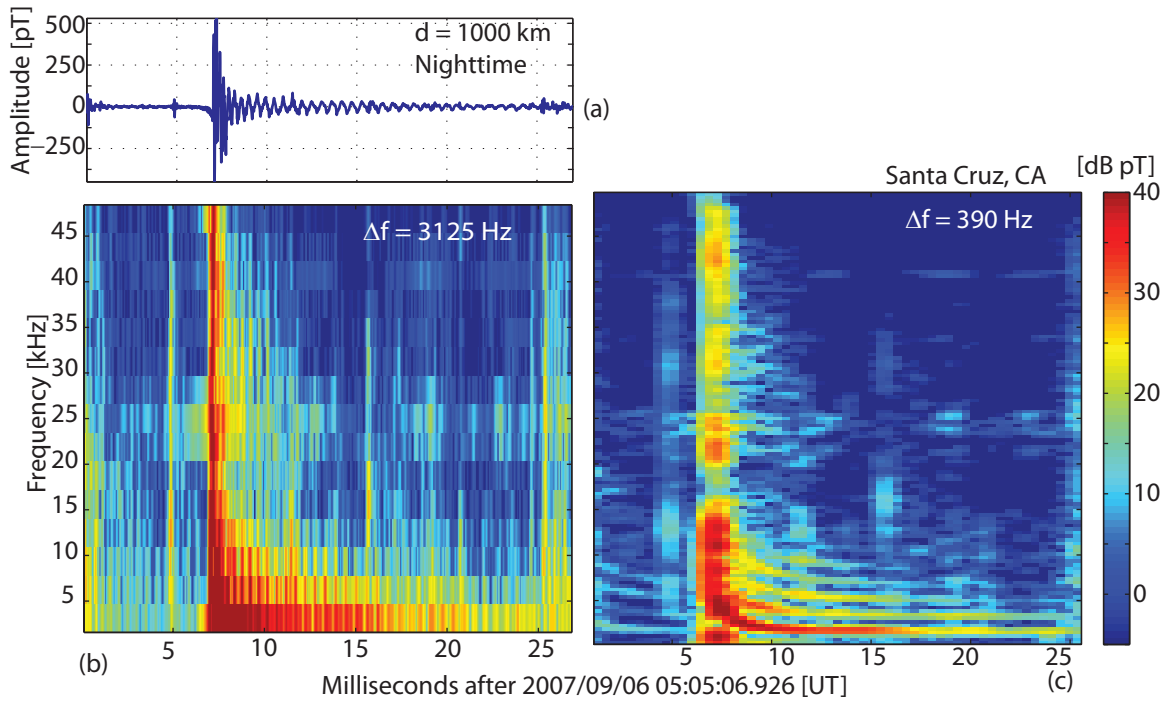
Figure 2.6: (a) Time-domain waveform of a nighttime tweek. (b) Spectrogram with a 3125 Hz frequency bin resolution, showing the late-time waveform as a series of impulses. (c) Spectrogram with a 390 Hz frequency bin resolution, which resolves the late-time waveform as a series of modes near cut-off.

introduced, followed by Wait's approach.

Budden first considers an infinite line dipole along the y-axis made up of an infinite number of vertically-oriented dipoles. Using (2.26), the resulting Hertz vector is

$$U_z = \frac{M}{4\pi} \int_{-\infty}^{\infty} \frac{\exp[-jk\sqrt{r^2 + y^2}]}{\sqrt{r^2 + y^2}} dy = \frac{-jM}{4} H_0^{(2)}(kr) \tag{2.43}$$

where $H_0^{(2)}(x)$ is a Hankel function of the second kind. Budden then considers a line quadrapole source, which consists of two line dipoles parallel to the y-axis of strengths $\pm M$ centered $\pm \frac{1}{2}\delta x$ from the origin in the $z = 0$ plane. Taking the derivative with respect to the x-coordinate of (2.43) gives [*Budden*, 1961a, p. 51]

$$U_z = \frac{jQ}{4} \cos\theta \frac{\partial}{\partial r} H_0^{(2)}(kr) = \frac{-jkQ}{4} \cos\theta H_1^{(2)}(kr) \tag{2.44}$$

where $Q = \lim_{\delta x \to 0} M\delta x$ and where θ is the angle measured from the x-y plane[5]. A line quadrapole is considered for the following reason: the point source radiation term in (2.26) may be expanded as [*Budden*, 1961a, p. 55]

$$\frac{e^{-jkr}}{4\pi r} = \frac{-k}{8\pi} \int_{-\frac{\pi}{2}-j\infty}^{\frac{\pi}{2}+j\infty} \cos(\alpha) H_1^{(2)}(ks) dw \tag{2.45}$$

where w is the (complex) angle from the x-axis in the x-y plane, and $s(w)$ and $\alpha(w)$ correspond to the radial distance and elevation from the source, respectively, where each may be complex and is calculated in the usual manner assuming a Cartesian grid but using the complex angle w. Thus, a point source of strength M may be regarded as a spectrum of line quadrapoles of strength $M/(2\pi j)$. For simplicity, Budden considers first only the line quadrapole along the y-axis ($w=0$). This quadrapole may in turn be expanded as an angular spectrum of plane waves by expanding $H_1^{(2)}(kr) \cos\theta$

equations used by Wait are altered accordingly.

[5]Note that this definition differs from the previous definition where θ was measured from the vertical. This latter definition is returned to occasionally below, but Budden's definition is kept here for simplicity and continuity in the equations.

[*Budden*, 1961a, pp. 299,53]; (2.44) assuming this source expands to

$$U_z = \frac{kQ}{4\pi} \int_{-j\infty}^{\pi+j\infty} \exp[-jk(x\cos\theta - z\sin\theta)]\cos\theta d\theta. \tag{2.46}$$

Thus, a point source has been decomposed into a spectrum of plane waves suitable for further analysis. This quadrapole decomposition into plane waves is used to solve for modes with eigenangles $\{\theta_m\}$, each of which have an x-dependence of $\exp[-jkx\cos\theta_m]$, as described in the next paragraph. To recover the field generated by a point source, one needs to integrate over all quadrapole angles w as in (2.45). Remembering to scale by $1/(2\pi j)$, Budden shows [*Budden*, 1961a, p. 58] that one needs to make the substitution (here ρ is the distance measured along the ground from the source)

$$Qe^{-jkx\cos\theta} \rightarrow -\frac{M}{2}jH_0^{(2)}(k\rho\cos\theta). \tag{2.47}$$

This source is now placed at a height z_0 in the waveguide. Adjusting for the source height, Budden expands this source per (2.46) and treats the lower and upper boundaries as sharp reflectors at heights $z = 0$ and $z = h$ with reflection coefficients R_g and R_i, respectively. With this geometry, there are an infinite number of effective radiators above and below the waveguide at $2Nh \pm z_0$ where N ranges over all integers (positive, negative, and zero). These effective radiators form a diffraction grating which produces two plane waves at $\pm\theta$. Due to the lossy reflections from each boundary, the path of the N$^{\text{th}}$ radiator suffers a total of N reflections from a combination of the lower and upper boundaries. These reflections can be represented with an infinite geometric series. By swapping the order of the contour integral and the series summation, a closed form expression for U is found to be [*Budden*, 1961a, p. 257; *Budden*, 1962, p. 548]:

$$U_z = \frac{Q}{2\pi j} \int_{-j\infty}^{j\infty+\pi} \frac{jk}{2} F(\theta) e^{-jkx\cos\theta} \cos\theta d\theta \tag{2.48}$$

where $F(\theta)$ is given by [*Budden*, 1962, pg. 548, with scalar quantities and R_g referenced to $z = 0$]

$$F(\theta) = \frac{\left(e^{jkS(z+z_0)} + R_g e^{jkS(z-z_0)}\right)\left(e^{-2jkSz} + R_i e^{-2jSkh}\right)}{1 - R_g R_i e^{-2jShk}}. \tag{2.49}$$

To evaluate (2.48) Budden distorts the contour so that it is symmetrical about $\theta = 0$ in the complex θ-plane, running from $\pi/2 - j\infty$ to $\pi/2 + j\infty$. The contour may need to be distorted to loop around a critical line[6], but contributions from the integral surrounding this branch cut are minimal [*Budden*, 1962]. Since $F(\theta)$ is Hermitian, the integral along this new symmetric path is zero (this simplification is partly why Budden chose to measure θ from the horizontal). The total integral from distorting the path is therefore given by $2\pi j$ times the sum of the residues of the poles crossed. These poles are at locations θ in the complex plane which satisfy $1 = R_g R_i e^{-2jkSh}$. With the proper definition of θ, this condition is identical to (2.39). Thus the modes are determined by the location of the poles of the integrand in (2.48).

The residue of a simple pole at z_0 of a function $f(z) = g(z)/h(z)$ with $h(z_0) = 0$, $h'(z_0) \neq 0$, and $g(z_0) \neq 0$ with $g(z_0)$ finite is $g(z_0)/h'(z_0)$ [*Boas*, 1983, p. 599]. If $F(\theta) = 1/(1/F(\theta))$ has poles at θ_n given by (2.39), then the residues are given by

$$\left[\frac{\partial}{\partial\theta}\frac{1}{F(\theta)}\right]^{-1}_{\theta=\theta_n}.$$

Hence the Hertz vector for the quadrapole source with $w = 0$ is

$$U_z = Q\sum_n \frac{jk}{2}\left[\frac{\partial}{\partial\theta}\frac{1}{F(\theta)}\right]^{-1}_{\theta=\theta_n} e^{-jkx\cos\theta}\cos\theta \tag{2.50}$$

$$= Q\sum_n \frac{jk}{2}\left[\frac{\partial}{\partial\sin\theta}\frac{1}{F(\sin\theta)}\right]^{-1}_{\theta=\theta_n} e^{-jkx\cos\theta} \tag{2.51}$$

where the last equality follows from (2.48) using $\cos\theta\, d\theta = d\sin\theta$.

---

[6]This branch cut is linked to the line cut in the complex θ-plane associated with the reflection coefficients. This line cut separates solutions where the transmitted wave normal moves away from the boundary from solutions where the transmitted wave normal moves toward the boundary.

As an example, consider the simplified case where $R_g = 1$. Then (2.49) becomes

$$F(\theta) = \frac{\left(e^{jkS(z+z_0)} + e^{jkS(z-z_0)}\right)\left(e^{-2jkSz} + R_i e^{-2jkSh}\right)}{1 - R_i e^{-2jkSh}} \tag{2.52}$$

$$= 2\left[\frac{1 + R_i e^{-2jkSh}}{1 - R_i e^{-2jkSh}}\right]\cos(kSz_1)\cos(kSz) - 2j\cos(kSz_0)\sin(kSz). \tag{2.53}$$

The residues are readily found by taking the derivative of the denominator and noting that $R(\theta)\exp[-2jhk\sin\theta]|_{\theta=\theta_n} = 1$. Note that the second part of (2.53) contains no poles and so this part of the integral vanishes. The expression for U_z becomes

$$U_z = Q\sum_{n=0}^{\infty}\frac{jk}{2}r_n f_n(z_1)f_n(z)\exp[-jkx\cos\theta_n], \tag{2.54}$$

where θ_n is a solution to (2.39) (with $\cos\theta_n \to \sin\theta_n$), $f_n(z) = \cos(kz\sin(\theta_n))$, and r_n is a dimensionless quantity

$$r_n = \frac{2}{jkh}\left[1 + \frac{j}{2kh\cos\theta_n R(\theta_n)}\left(\frac{\partial R(\theta)}{\partial\theta}\right)_n\right]^{-1} \tag{2.55}$$

$$= \frac{2}{jkh}\left[1 + \frac{j}{2khR(\sin\theta_n)}\left(\frac{\partial R(\sin\theta)}{\partial\sin\theta}\right)_n\right]^{-1}. \tag{2.56}$$

The form (2.56) was obtained from (2.55) via the chain rule and is the form for r_n if the integration variable is $S = \sin\theta$.

If, in addition, $(\partial R/\partial\theta)_n \ll 2kh\cos\theta_n R(\theta_n)$, or at least the reference height h for $R(\theta)$ is chosen so that this condition is met, then (2.54) reduces to

$$U_z = \frac{Q}{h}\sum_n f_n(z_1)f_n(z)e^{-jkx\cos\theta_n}.$$

Wait [1970][7] takes a different approach. He starts with an infinite line source in the vertical direction centered on the point source. This line source radiates fields with radial dependence $H_0^{(2)}(k\rho\cos\theta)$, a functional form derived analogously as (2.26).

---

[7]Recall that in his papers, Wait references θ to the plane normal. For continuity, the following discussion rewrites Wait's formulation using Budden's definition of θ measured from the horizontal. To recover Wait's original formulas, simply swap C ($\cos\theta$) for S ($\sin\theta$) in the equations.

This presupposition suggests that one starts with a different expansion of a point source,

$$U_z = \frac{M \exp[-jkr]}{4\pi r} = \frac{Mjk}{8\pi} \int_\Gamma H_0^{(2)}(kC\rho)e^{-jkSz}dS. \tag{2.57}$$

Here, $S = \sin\theta$ is the sine of the (complex) angle from the normal of the boundary, $C = \sqrt{1 - S^2}$, ρ is the radial distance from the source, and z is the height above the source (assuming the source is at $z = 0$). This expansion suggests a solution of upgoing and downgoing exponential terms [*Wait*, 1970, p. 138]. Assuming this functional dependence and homogeneous boundaries, Wait uses the boundary conditions to solve for the field below, inside, and above the waveguide (an inhomogeneous waveguide may be considered by choosing a reflection height suitably low such that the medium in all altitudes below this height may be regarded as free space). The boundary conditions for U are found by using the continuity of E_ρ and H_ϕ across plane surfaces to relate the waveguide U_z to the ionosphere term $U_z^{(i)}$ at $z = h$ and the ground term $U_z^{(g)}$ at $z = 0$:

$$\left.\begin{array}{rcl} k^2 U_z & = & k_l^2 U_z^{(l)} \\[2mm] \dfrac{\partial U_z}{\partial z} & = & \dfrac{\partial U_z^{(l)}}{\partial z} \end{array}\right\}, \ l = g, i \text{ at } z = 0, h. \tag{2.58}$$

Wait arrives at a contour integral analogous to (2.48), but directly using $H_0^{(2)}(kC\rho)$ for the radial dependence:

$$U_z = \frac{1}{2\pi j}\frac{jM}{2}\int_\Gamma \frac{jk}{2}F(S)H_0^{(2)}(kC\rho)dS \tag{2.59}$$

where $F(s)$ is given by

$$F(S) = \frac{\left(e^{jkSz} + R_g e^{-jkSz}\right)\left(e^{-jkS(h-z_0)} + R_i e^{-jkS(h-z_0)}\right)}{e^{ikSh}\left(1 - R_g R_i e^{-2jkSh}\right)} \tag{2.60}$$

which differs by (2.49) by $-2j\sin[kS(z - z_0)]$. This extra term does not contribute anything to the integral since it contains no poles. Thus, (2.59) recovers Budden's

result (2.48) after using the substitution (2.47) to within a minus sign[8].

Using (2.60), equation (2.54) is reproduced except with $f_n(z)$ modified to incorporate a general ground reflection coefficient:

$$f_n(z) = \frac{e^{jkS_n z} + R_g(S_n)e^{-jkS_n z}}{2\sqrt{R_g(S_n)}} \tag{2.61}$$

This work uses measurements of the horizontal magnetic field. In cylindrical coordinates with propagation in the ρ-direction and $\mathbf{U} = U_z \hat{\mathbf{z}}$, (2.24) the horizontal magnetic field is $B_\phi = -j\mu\omega\partial U_z/\partial\rho$. After the substitution (2.47), one can simplify the expression further for $k\rho\cos\theta \gg 1$ by using the expansion $H_\nu^{(2)}(x) \simeq \sqrt{2/(\pi x)}\exp\left[-jx + j(2\nu + 1)\pi/4\right]$ for $x \gg 1$ [*Gradshteyn and Ryzhik*, 1980, p. 962]. Under this condition (2.54) gives (where here θ is still referenced to the horizontal plane)

$$B_\phi(\rho, z) \approx \frac{j\mu k^{3/2}Ile^{j\pi/4}}{\sqrt{8\pi\rho}}\sum_{n=0}^{\infty}\sqrt{\cos\theta_n}\,r_n f_n(z_1)f_n(z)e^{-jk\rho\cos\theta_n}. \tag{2.62}$$

The terms in this expression have corresponding physical meanings. The leading constant is a frequency-dependent gain factor which depends on the current moment of the source. The $\rho^{-1/2}$ distance dependence reflects amplitude roll-off in a cylindrically symmetric waveguide where power falls off as $1/\rho$. The field consists of many modes n, though in practice only a few have appreciable amplitude. The term $\sqrt{\cos\theta_n}\,r_n f_n(z_1)$ is an excitation factor that reflects the relative effectiveness with which the source excites the given mode in the waveguide. As an example, if $f_n(z_1)$ falls off with height, a higher source is less effective at exciting the TM mode. The term $f_n(z)$ is a height-gain function and reflects the altitude dependence of the measured field; in this work all fields are measured at $z = 0$. Finally, the exponential describes propagation of the wave, which defines the waveguide number β_n and attenuation rate α_n as defined in the previous section (but with $C \to S$).

---

[8]*Budden* [1962] shows that this change in minus sign gives an equivalent result as simply changing the sign of θ (when it is measured from the horizontal).

2.4.3 Anisotropic Ionosphere

The extension to an anisotropic ionosphere follows the exact formulation in the previous section. In Budden's formulation, a vertically oriented electrical line quadrapole source with strength Q which gives rise to a vertical Hertz vector $U_z\hat{\mathbf{z}}$ is once again presumed. Under an isotropic ionosphere, this polarization propagates as a TM mode with the electric field strictly perpendicular to the direction of propagation, where the reflection coefficient is given by $R_\parallel$. With the presence of Earth's magnetic field, the incident polarization is coupled to a TE polarization where the electric field is strictly perpendicular to the direction of propagation. This polarization is captured through the use of an analogous Hertz vector V, which is excited by a magnetic line quadrupole of strength Q_{M} (with an isotropic ionosphere, this polarization would be governed by $R_\perp$). The off-diagonal components of (2.40) couple energy between the TM U component and the TE V component, so that the mode condition (2.42) leads to solutions $\mathbf{h} = [U, V/\eta_0]^T$ which contain components of each polarization. The polarizations are now called QTM and QTE based on the dominant polarization of the solution.

Given source electric and magnetic quadrapoles, after forming a diffraction grating by extending the sources to $z_0 \pm 2Nh$ and summing the contributions, *Budden* [1962, p. 548] derives the analogous result to (2.46)

$$\mathbf{h} = \begin{bmatrix} U \\ -V/Z_0 \end{bmatrix} = \frac{1}{2\pi j} \int_{-j\infty}^{j\infty+\pi} \frac{jk}{2} \mathbf{F}(\theta) \mathrm{e}^{-jkCx} \cos\theta \begin{bmatrix} Q \\ -Q_M/Z_0 \end{bmatrix} d\theta \qquad (2.63)$$

where, with $z_0 = z = h = 0$,

$$\mathbf{F}(\theta) = (\mathbf{I} + \mathbf{R_i})\mathbf{W}(\mathbf{I} + \mathbf{R_g}) \qquad (2.64)$$

where $\mathbf{W} = [\mathbf{I} - \mathbf{R_i R_g}]^{-1}$ and $\mathbf{R_i}$, $\mathbf{R_g}$ are both referenced to $z = 0$ and are given by (2.40) and (2.41), respectively. The mode condition is $|\mathbf{W}^{-1}| = 0$ which, with $h = 0$, recovers (2.42). Adjusting the integration path as before to go from $-\pi/2 - j\infty$ to $\pi/2 + j\infty$ and adding up the contribution from the residues of the crossed poles,

the vertically polarized component (which is what is ideally measured in this work, though a horizontal field measurement additionally measures a component of V) is given by

$$U_z = \sum_n \begin{bmatrix} 1 & 0 \end{bmatrix} \frac{jk}{2} \mathbf{\Lambda_n} \begin{bmatrix} Q \\ 0 \end{bmatrix} e^{-jkC_n x} \qquad (2.65)$$

where $\mathbf{\Lambda_n}$ is the excitation factor matrix

$$\mathbf{\Lambda_n} = \frac{(\mathbf{I} + \mathbf{R_i})\mathbf{X}(\mathbf{I} + \mathbf{R_g})}{\left(\dfrac{\partial \Delta}{\partial S}\right)_n} \qquad (2.66)$$

with $\Delta = |\mathbf{W}^{-1}|$ and $\mathbf{X} = \lim_{\theta \to \theta_n} \mathbf{W}\Delta$. The row vector in (2.65) picks out the vertical polarization component; the 0 in the column matrix reflects the ideal case where one has a strictly vertical source term. This expression should be compared to (2.54), where Λ_n corresponds to $r_n f_n(z_1) f_n(z)$; height-gain functions could be recovered in the anisotropic case by raising the reference height of the reflection matrix. Assuming a small TE component, the procedure for recovering the measured field is the same as before. Therefore the equivalent expression to (2.62) is trivially obtained by substituting $[1, 0]\mathbf{\Lambda_n}[1, 0]^T$ for $r_n f_n(z_1) f_n(z)$.

Wait's formulation may also be extended to an anisotropic ionosphere by substituting $\mathbf{R}_g$ and $\mathbf{R}_i$ in (2.60) and following the rules for matrix inversion; the details may be found in *Wait* [1970, p. 248].

2.4.4 Curved Earth

There are two approaches to accommodate a curved earth. The first is to introduce a modified refractive index between the ground and the base of the ionosphere. A straight ray between the lower and upper boundary of a curved surface appears to an observer on the ground as an arc, i.e., the transversal will not produce congruent angles on the lower and upper boundaries. This effect may be captured by assuming a parallel plate waveguide with a real refractive index that varies as $\mu = \exp\left[(z - z_1)/a\right]$, where a is the radius of the earth [*Budden*, 1961a, p. 140]. The reference height z_1 is often set to the effective reflection height h so that the refractive index is unity at the

upper boundary. The exponential may additionally be expanded in a Taylor series to simplify the mathematics.

If this modified refractive index changes slowly over a wavelength, a WKB approximation may be used to modify the modal condition as follows:

$$R(\theta) \exp\left[-2jk \int_0^h q\,dz\right] = 1 \qquad (2.67)$$

where $q^2 = \mu^2 - \cos^2 \theta$. This WKB approximation works for the frequencies above ~ 8 kHz; below this frequency, a numerical full-wave solution [e.g., *Lehtinen and Inan*, 2008] should be used where the variable μ is layered.

The second approach, taken by *Wait* [1970] and *Galejs* [1972], is to directly consider a spherical waveguide. An approximate solution, valid at lower frequencies, may be arrived at by considering a WKB or second-order approximation to the spheric wave functions. This approximation leads to a modal solution which is equivalent to a WKB approximation of a parallel plate waveguide with a linear refractive index (compare *Wait* [1970, p. 157] to *Budden* [1961a, pp. 141,284]).

A more precise approach starts with an integer series expansion of spherical harmonics which is transformed to a quickly convergent series with modal terms analogous to the parallel plate waveguide. Both of these approaches are outlined in Appendix D.

Finally, the cylindrical spreading factor needs to be transformed to reflect the spherical geometry. As shown in Appendix D, the necessary substitution is

$$\frac{1}{\sqrt{\rho}} \rightarrow \frac{1}{\sqrt{R_{\mathrm{E}} \sin\left(\dfrac{\rho}{R_{\mathrm{E}}}\right)}}. \qquad (2.68)$$

2.5 General Trends

Numerical calculations for the reflection coefficient and for waveguide modal parameters, such as phase velocity, attenuation, and excitation factors, have been conducted extensively in the literature. These results give us insight when analyzing measured

waveforms and looking for the dominant factors affecting the propagation channel.

A summary of numerical results derived using the model outlined in Appendix D is found for $f > 8$ kHz in *Wait and Spies* [1964]. In the simulation results, an exponential ionosphere is assumed with the effective conductivity parameter ω_r given by

$$\omega_r = (2.5 \times 10^5 s^{-1}) \exp\left[\beta(z - h')\right] \tag{2.69}$$

where β determines the sharpness of the boundary and h' gives the reference height at which $\omega_r = 2.5 \times 10^5 s^{-1}$. The daytime (nighttime) condition is often represented with $\beta = 0.3$ km^{-1} ($\beta = 0.5$ km^{-1}). A static magnetic field is introduced in *Wait and Spies* [1964] which is strictly horizontal and transverse to the direction of propagation and plots are provided using the parameter

$$\Omega = \frac{Y}{Z|_{z=h}}. \tag{2.70}$$

The sign of Ω indicates the relative direction of propagation. A negative (positive) value indicates west-to-east (east-to-west) propagation. As seen in Figure 2.1, for daytime propagation $Z \simeq Y$ at the reflection height (when $X \simeq Z$), and so $|\Omega| \simeq 1$; under nighttime conditions, $Y > Z$ when $X \simeq Z$, so $|\Omega| > 1$. Note that $\Omega = \omega_H/\nu|_{z=h}$ depends on the magnetic field strength and the collision rate; a common expression for ν which is used by Wait is

$$\nu = (5 \times 10^6 s^{-1}) \exp\left[-0.15(z - 70)\right]. \tag{2.71}$$

Using these parameters, *Wait and Spies* [1964] show that the reflection coefficient is higher for higher values of β and for more negative values of Ω, indicating a minimum in propagation loss for eastward propagating waves under a nighttime ionosphere and a maximum in propagation loss for westward propagating waves under a daytime ionosphere. The phase of the reflection coefficient is also dependent on β and ω, where, with $\cos\theta = 0.1$ (grazing incidence) and $f = 10$ kHz, the phase between $\Omega = \pm 3$ is less than $5°$. This dependence affects the phase measured on the ground much less than the reflection height h'. For example, at 10 kHz ($\lambda \simeq 30$ km), a reference height

increase of 10 km changes the measured phase by $\sim(360)(10)/30 = 120°$.

The effect of the ionosphere on the overall propagation characteristics of waveguide modes is qualitatively similar to the reflection coefficient results. Attenuation is lower for all ionospheric profiles for west-to-east propagation compared to east-to-west propagation. Attenuation is slightly lower for higher β values and increases with a lowering reference height. The lowest attenuation for the least attenuating mode in daytime ($h \simeq 70$ km) is at ~ 18 kHz; for nighttime ($h \simeq 90$ km), ~ 12 kHz. Finally, the attenuation increases as the ground conductivity decreases.

Attenuation profiles for $f < 20$ kHz in *Wait* [1957] show a maximum in attenuation for the lowest attenuated mode[9] near ~ 5 kHz. This finding can be seen clearly in Figure 2.7, which shows a spectrogram of a sferic which has propagated 4000 km under a daytime ionosphere. The amplitude is strongest near 15 kHz with rapidly attenuating energy below 5 kHz and above 20 kHz. There is also a weak ELF response following the VLF impulse, with a peak amplitude between 9 and 15 milliseconds in the figure. This response is the so-called "Slow-tail" which propagates in the quasi-transverse electromagnetic (QTEM, or, equivalently, QTM_0) mode below 1 kHz [*Reising et al.*, 1999].

Broadly speaking, the earth-ionosphere waveguide therefore affords long-range detection at VLF, but the path profile defining the channel between the lightning and the receiver depends on several time-varying and position-dependent factors. Figure 2.8 illustrates the many path-dependent factors that may affect the received waveform, including an anisotropic and inhomogeneous ionosphere and a varying ground conductivity along the path profile. The lightning geo-location network design developed in this thesis will strive to correct for these varying factors in a systematic way to improve the overall location accuracy.

---

[9]For a waveguide with a lossy boundary, the mode number assignment of the eigenangles may not be obvious [*Budden*, 1961a, p. 144]. The lowest-attenuating mode at a given frequency refers to that mode which suffers the least attenuation.
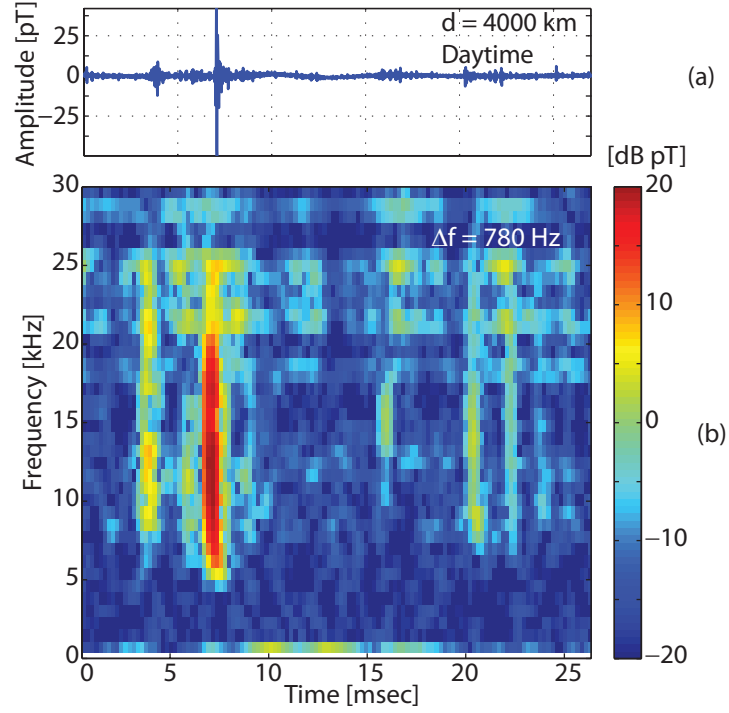
Figure 2.7: (a) Time-domain and (b) spectrogram plot of a sferic after a 4000 km propagation under a daytime ionosphere.
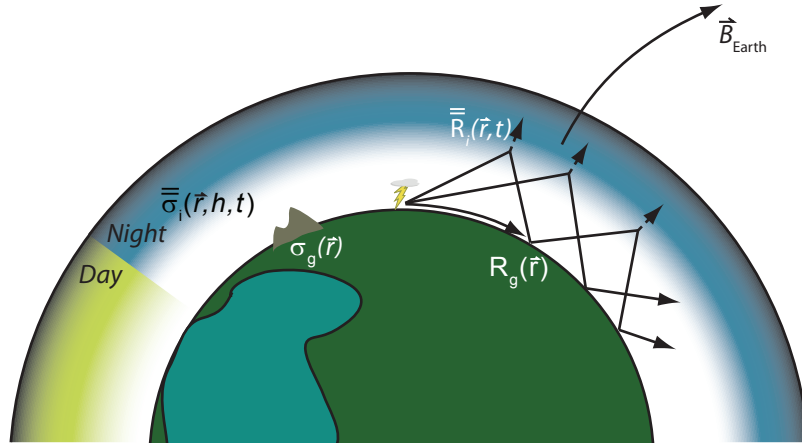
Figure 2.8: Illustration of the many complicating factors in the real Earth-ionosphere waveguide, including a spatially-varying ground conductivity and a spatially- and temporally-varying ionospheric profile.

Chapter 3

Broadband VLF Data Analysis

The lightning geo-location network developed in this thesis processes individual sferics recorded at several ground-based and geographically separated receivers. This chapter discuses how these sferics are recorded and prepared for further processing.

Chapter 5 develops a receiver algorithm that first cross-correlates waveforms from individual lightning strikes with a bank of matched filters and, subsequently, derives timing features from the time-domain waveform. Only noise sources that lie inside the energy band of each matched filter affect the first step; however, the second operation is wideband and is affected by the entire bandwidth. We are therefore interested in maximizing the signal to noise ratio of the received sferic waveforms across the entire recorded bandwidth. A modest 6 dB increase in SNR can nearly double the effective range of a long-range sferics sensor. This chapter demonstrates a way to improve the SNR by mitigating interference from man-made noise sources such as powerline hum and VLF transmitters.

The first section discusses the receiver architecture used in this work. The second section is devoted to minimizing interference sources to maximize the sferic SNR. Finally, the use of two-channel data to measure the arrival azimuth at each receiver to extract the arrival angle θ for use in (1.2) is discussed.
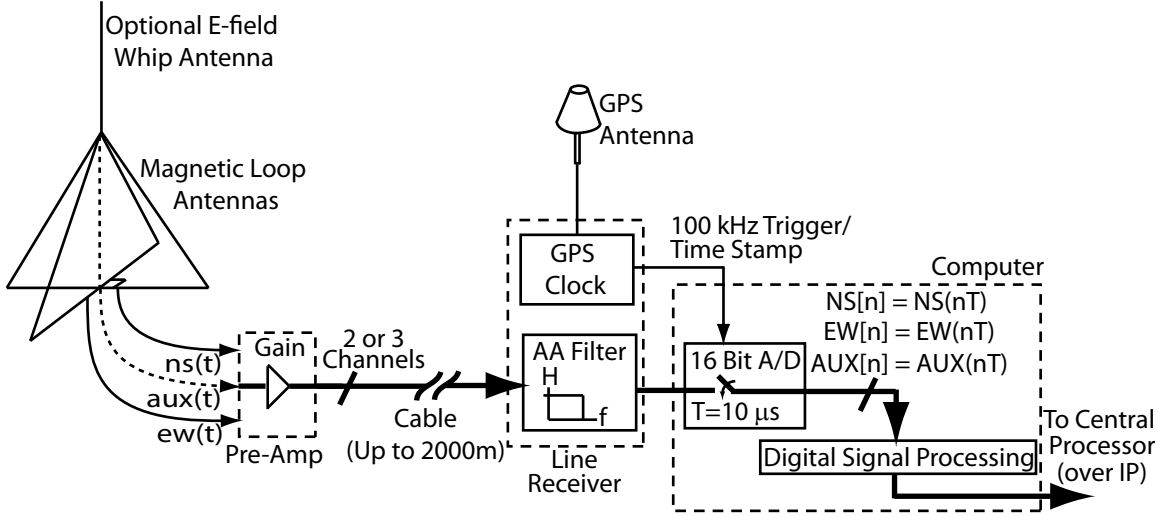
Figure 3.1: Block diagram of the receiver hardware, showing the standard two-channel system with the orthogonal magnetic loop antennas and an optional third channel using a vertical electric field whip antenna.

3.1   Data acquisition and preprocessing

Figure 3.1 contains a block diagram of the receiver. Using two orthogonally aligned air-core magnetic loop antennas, each receiver measures the horizontal magnetic field between ~DC and 50 kHz. An optional component is a vertical whip antenna, which measures the vertical component of the electric field. The current induced on each antenna is amplified in a preamp and then sent over a long cable to the receiver electronics. This separation minimizes electromagnetic interference coupling into the antenna from both the electronics and the power source.

The signal at the end of this cable is passed through an anti-aliasing filter and then sampled, with 16-bit resolution, at 100 kSamples/second using GPS-synchronized timing. The sensor contains a digital signal processing unit and, if real-time operation is desired, has an Internet connection. Data for this thesis were collected using desktop PCs, equipped with a commercial National Instruments DAQ card. The design details of the receiver may be found in *Cohen et al.* [2009]. The receiver sensitivity and calibration are discussed in Appendix B.

Figure 3.2 contains a spectrogram of raw broadband data from one of the magnetic loop antennas of a sensor at Palmer Station on the Antarctic Peninsula. This figure shows that the incoming sferics, seen as the impulsive (wideband) vertical lines in the spectrogram, do not arrive in a noise-free environment. The SNR is limited by the system noise (Appendix B) and naturally and artificially generated signals. Artificially generated signals include minimum-shift-keyed (MSK) modulated US Navy transmitters, which operate above 20 kHz and act as communication links with submarines [*Watt*, 1967, Ch. 2]. Also seen in the spectra are pulsed Russian alpha beacon signals, which follow a regular frequency pattern from three different sources [*Inan et al.*, 1984]. While not visible in this data sample, harmonics from nearby power-lines are often visible up to 5 kHz or higher; this particular receiver is located more than a kilometer away from the nearest building and enjoys an unusually quiet background.

In addition to the artificially-generated signals, several sources of natural emissions reduce the effective SNR. One such source, common at more extreme magnetic latitudes, are whistlers, which result from lightning-generated radiation that has propagated through the magnetosphere along field-aligned ducts of enhanced conductivity and re-entered the earth-ionosphere waveguide [*Helliwell*, 1965, p. 4]. Another natural noise source originating from the magnetosphere is chorus, which are emissions consisting of quasi-monochromatic signals from hundreds of hertz to 5 kHz [*Sazhin and Hayakawa*, 1992].

The naturally occurring noise sources are incoherent and broadband, and, therefore, cannot be removed without also filtering out the broadband signals of interest. However, most artificial sources occupy a relatively narrow bandwidth, and, as is the case with powerline hum and communication narrowband transmitter signals, are coherent. As a result, artificial sources can be partially subtracted from the data.

The next two sections discuss the techniques used in this work to mitigate interference from powerline hum and narrowband interference sources. In both cases, the procedure for coherent noise mitigation follows the design of an open-loop adaptive filter. An attempt is made to first isolate the noise source and then coherently subtract the result from the original data. Techniques from the literature are adapted to suit our needs, and sub-optimal techniques are introduced that may be used if limited
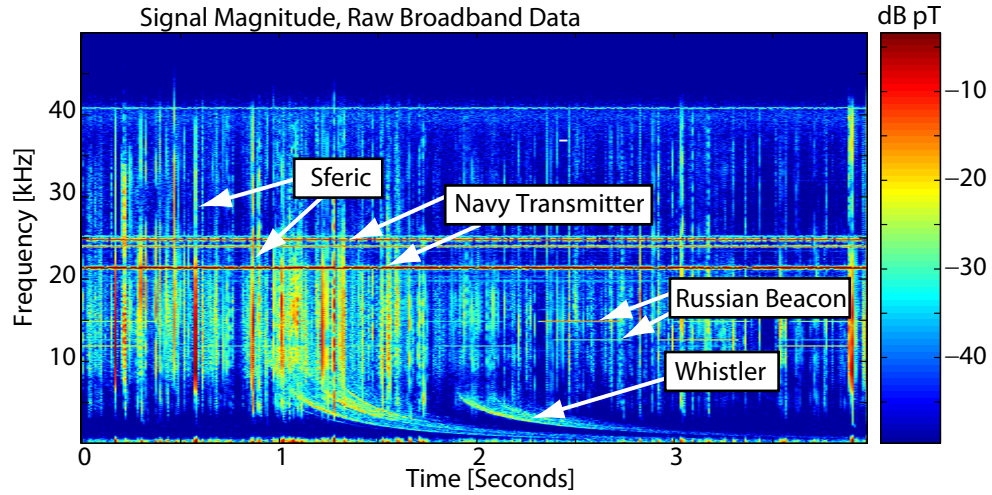
Figure 3.2: Spectrogram of unfiltered data, showing the broadband sferics together with various artificial and man-made noise sources.

processing resources are available, which may become a factor in applications that require real-time processing of such data.

3.1.1 Powerline Hum

Powerline hum is generated by nearby power lines, which operate at a base frequency of 50 or 60 Hz, depending on the location. The fundamental frequency f_0 changes with time due to varying loads on the power line grid. Figure 3.3 shows an estimate of the fundamental frequency $f_0(t)$ versus time at a receiver at Kodiak Island, Alaska.

Interfering powerline harmonics are seen well into the VLF band, representing contributions from up to ∼100 harmonics. With a fundamental deviation of only 0.1 Hz, as in Figure 3.3, these upper harmonics may deviate by several Hz. Therefore, both the estimation of the fundamental frequency f_0 and the amplitudes and phases of each harmonic need to be adjusted over time.

The problem of removing powerline noise has been considered extensively in geophysics to clean seismic data [e.g., *Nyman and Gaiser*, 1983; *Butler and Russell*, 2003; *Saucier et al.*, 2006]. The canonical approach, which is outlined and optimized in *Saucier et al.* [2006], uses a maximum likelihood estimation (MLE) and treats the desired signal $x(t)$ as a white Gaussian noise process. Assume first that f_0 is constant
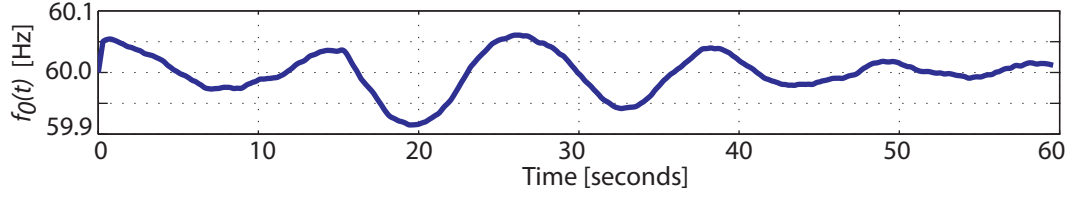
Figure 3.3: Fundamental powerline harmonic f_0 at Kodiak Island in Alaska for a 60 second period.

over a certain time interval. Assuming M harmonics, the powerline signal may be written as

$$s(t) = \sum_{k=1}^{M} [A_k \cos(2\pi k f_0 t + \phi_k)] + x(t) \tag{3.1}$$

where A_k and ϕ_k are the unknown amplitudes and phases of the powerline harmonics, respectively, and $x(t)$ is an added noise source, which includes the sferics and all other noise sources. Sampling at an interval of Δt and converting to a linear equation in terms of the unknowns yields

$$s(n) = \sum_{k=1}^{M} [a_k \cos(2\pi k f_0 n \Delta t) + b_k \sin(2\pi k f_0 n \Delta t)] + x(n), \quad n = 0, 1, \ldots, N - 1. \tag{3.2}$$

This equation can be written as a system of linear equations: $\mathbf{s} = H\boldsymbol{\theta} + \mathbf{x}$, where $\mathbf{s} = (s(0), s(1), ..., s(N - 1))^T$, $\boldsymbol{\theta} = (a_1, b_1, ..., a_M, b_M)^T$, $\mathbf{x} = (x(0), ..., x(N - 1))^T$, and H is an $N \times 2M$ matrix with $H_{n,i} = \cos(i2\pi f_0 n \Delta t)$, $H_{n,i+1} = \sin(i2\pi f_0 n \Delta t)$ for $n = 0, 1, \ldots, N - 1$, $i = 1, 3, 5, \ldots, 2M - 1$.

If $x(t)$ is assumed to have a Gaussian distribution, then the maximum likely estimator of $\mathbf{s}$ is $H\widehat{\boldsymbol{\theta}}$ where $\widehat{\boldsymbol{\theta}}$ is the standard least-squares solution

$$\widehat{\boldsymbol{\theta}} = \min_{\boldsymbol{\theta}} ||\mathbf{s} - H\boldsymbol{\theta}||^2 = (H^T H)^{-1} H^T \mathbf{s}. \tag{3.3}$$

The orthogonality of sinusoids may be exploited to evaluate $H^T H$ analytically; the relevant expressions are given in *Saucier et al.* [2006] and are readily derived from a well-known result of the Discrete Fourier Transform (DFT) [*Oppenheim and Schafer*, 1989, Ch. 8]. Calculating $H^T H$ in this way reduces the computation overhead, especially for long time records. Computation speed and accuracy may also be improved

by lowpass filtering and downsampling, which reduces the dimensions of each matrix and minimizes the noise power in $x(t)$. Even with an analytic solution for $H^T H$, however, building H^T and taking the inverse of $H^T H$ takes considerable computational power. Computing H^T involves $(N)(M)$ multiplications and a matrix inverse scales as $O(M^3)$ [*Press*, 2002, p. 51].

In this least-squares formulation, the fundamental frequency f_0 is found by minimizing the residual of the least-squares estimation, which itself is a function of f_0, $r(f_0) = ||\mathbf{s} - H(f_0)\widehat{\boldsymbol{\theta}}||^2$. Note that $r(f_0)$ is not a convex function, but if the domain is constrained to be within a few Hz of the nominal powerline frequency, then the problem may be bounded and a single minimum found. If the powerline harmonics decay sufficiently fast as a function of frequency, one need remove only the lower order harmonics during this iteration.

This minimization to find the fundamental frequency can be quite costly depending on the number of iterations required. While several techniques have been presented to reduce the computation time at a slight cost in performance [*Saucier et al.*, 2006], we found that a one-step Gauss-Newton optimization [*Boyd and Vandenberghe*, 2004] is sufficient given a good starting guess. To obtain this initial estimate of $f_0(t)$, we first segment the data and take a Fourier Transform of each section to obtain a short-time Fourier Transform (STFT, [*Smith*, Accessed 2009]), as is done in generating the spectrogram. Using a Gaussian window, we use quadratic interpolation to estimate the center frequency of several harmonics, which are then averaged using a weighting proportional to the linear amplitude and harmonic number to estimate the fundamental frequency:

$$f_0(l) = \frac{\sum_k f_k^l A_k^l k}{\sum_k A_k^l k} \tag{3.4}$$

where l is the segment index in the STFT. Figure 3.4a compares the mean squared error $(\widehat{f} - f_0)^2$ of the initial estimate $\widehat{f}_0$ based on the STFT, and after one Gauss-Newton iteration step. A 30 dB SNR[1] with respect to the powerline hum was assumed,

---

[1]Even if the peak amplitude of a sferic exceeds the amplitude of the powerline hum, 30 dB is not an unreasonable SNR. During the powerline isolation procedure, the length of time over which the powerline parameters are assumed constant is on the order of several hum cycles: at 60 Hz and a six-cycle duration, for example, the signal length is 100 ms. The short signal duration of a sferic (~1 ms), when averaged over the full estimation window, significantly reduces the contribution of
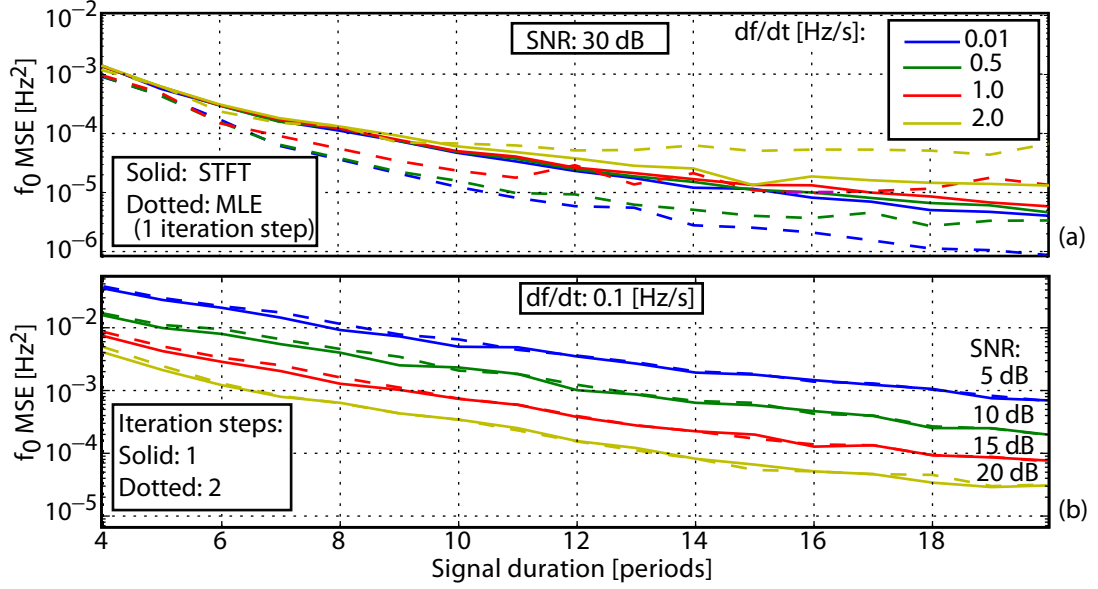
Figure 3.4: (a) Evaluation of the STFT and MLE estimation of f_0 under varying harmonic slopes. (b) Justification for using only one Gauss-Newton step for finding f_0 with the MLE.

and performance was evaluated for increasing slopes in the fundamental frequency f_0 as a function of the signal duration used for the estimate.

The least-squares procedure assumes a constant f_0 over the signal duration used in the matrix H. Note in Figure 3.4a that there is a more accurate estimation of f_0 for longer signal durations for cases where $df/dt < 1.0$. This same signal duration was used as the segment length in the initial STFT estimate of f_0.

Figure 3.4b justifies our use of one iteration step to arrive at the estimate for f_0. A constant 0.1 Hz/s slope was assumed for the fundamental frequency with four different SNR values. The solid lines show the mean-squared-error (MSE) after one iteration; the dotted line after two iterations. The results are virtually identical. Thus, the minimization to find f_0 requires only three evaluations of $r(f_0)$ (to calculate the Gauss-Newton step).

While the MLE approach provides the best estimate of the powerline hum in the presence of Gaussian noise, in our experience, the computational complexity is

the sferic to the overall SNR.

prohibitive when $\gtrsim 50$ harmonics must be removed[2]. The results in this thesis were obtained by using a faster, less optimal approach whereby a dynamically changing comb filter is used to isolate the powerline hum. This procedure necessarily removes some of the energy in the signal of interest. In order to minimize the impact on sferics, the data is first lowpass filtered and downsampled to only encompass the powerline harmonic frequency range. The time-dependent filter is

$$h(t, \tau) = \left[\sum_{k=-\infty}^{\infty} \delta(t - k\frac{1}{f_0(\tau)}) \right] w(t) \qquad (3.5)$$

which, at time τ, isolates harmonics every $kf_0(\tau)$

$$H(f, \tau) = \left[\sum_{k=-\infty}^{\infty} \delta(f - kf_0(\tau)) \right] * W(f). \qquad (3.6)$$

The windowing function $w(t)$ should be short enough such that f_0 does not change appreciably over the time window and parallels the choice of window length in the MLE approach. If the fundamental frequency changes more rapidly, then $W(f)$ widens to capture the range of the changing harmonic frequencies. In discrete time, $h(t)$ is readily implemented by windowing the data around a particular point and summing a down-sampled set composed of every $\text{nint}(f_s/f_0)$ sample, where f_s is the sampling frequency and 'nint' denotes the nearest integer function. In its simplest form, $w(t)$ is a rectangular window, and filtering by $h(t)$ reduces to averaging the nearest L samples in the down-sampled set. This last simplification reduces to a very intuitive result: averaging L samples spaced out every $\sim f_0^{-1}$ seconds overlaps the periodic signal onto itself ("superpose-epoch" averaging), while all other periodicities incoherently average to a small number. A similar suboptimal approach to estimating f_0 and isolating the powerline hum is given in *Cohen* [2009, Appendix D].

To evaluate these two methods in the presence of a varying fundamental frequency, the same metric used in *Saucier et al.* [2006] to evaluate the signal improvement is

---

[2]While, for each estimation window, only a few harmonics are needed to estimate the fundamental frequency f_0, the entire $\widehat{\boldsymbol{\theta}}$ vector needs to be calculated to estimate the amplitude and phase of each harmonic.
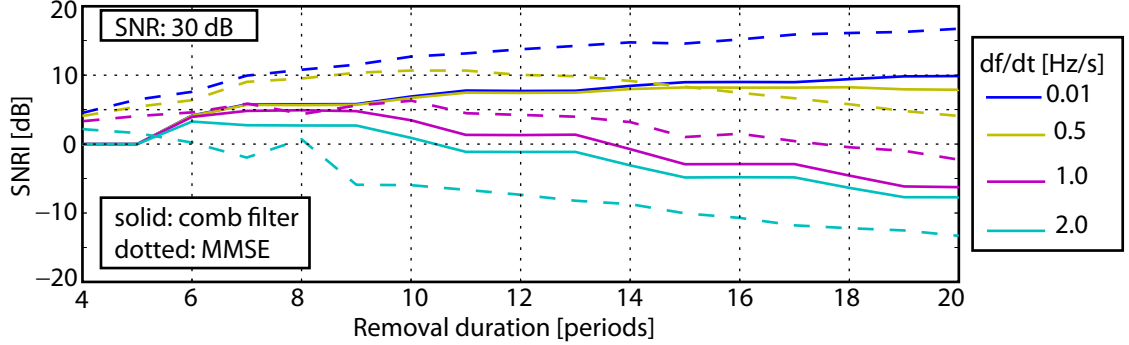
Figure 3.5: SNRI after removing powerline hum for various harmonic slopes using the comb filter method (solid lines) and the MMSE method (dotted lines).

applied, except that here an f_0 which has a constant slope to simulate real-world powerline signals is introduced. The initial SNR of the powerline hum, which is what one needs to isolate and then coherently subtract, is $||\mathbf{h}||^2/||\mathbf{x}||^2$, where $\mathbf{h}$ is the pure powerline hum $h(t) = \sum_{k=1}^{m} A_k \cos(2\pi k f_0 t + \phi_k)$, and $\mathbf{x}$ is everything else. The SNR of the isolated signal is $||\mathbf{h}||^2/||\mathbf{h} - \widehat{\mathbf{h}}||^2$, where $\widehat{\mathbf{h}} = H\widehat{\boldsymbol{\theta}}$ is the estimation of the hum. A measure of our ability to isolate the hum is therefore given by

$$\text{SNRI} = \frac{\text{SNR}_{\text{after}}}{\text{SNR}_{\text{before}}} = \frac{||\mathbf{x}||^2}{||\mathbf{h} - \widehat{\mathbf{h}}||^2}. \tag{3.7}$$

Figure 3.5 shows the SNRI metric using both methods, again plotted against the window length, assuming a powerline SNR of 30 dB and indexed by the fundamental powerline slope. As the slope increases, the SNRI performance degrades and the optimal removal duration decreases. This result is expected: for a higher slope, the region over which the single f_0 estimate is valid shortens. For very high slopes ($df/dt \gtrsim 2$ Hz/s), the MLE method does not remove a sufficient amount of the hum and is out-performed by the delta-train method. Nevertheless, for these slopes, the optimal window length is very short, so the corresponding window $W(f)$ removes much of the desired signal as well. Consequently, under these conditions, it is advisable to find a quieter receiver site.

In summary, the specific method used to remove powerline hum therefore depends on computation time and the rate of change of the powerline frequency. The comb

filter method removes all harmonics using a single time-varying filter, and the STFT to compute $f(t)$ may be implemented very efficiently by downsampling so as to only process the first several harmonics. Shortening the window provides an easy way to remove rapidly-varying hum, but the method intrinsically removes the desired signal as well. Since this method removes all harmonics, the broadband data signal should be lowpass-filtered before applying the filter so that frequencies above the highest powerline harmonic are not corrupted. The increase in computation complexity in the MLE, however, provides a better estimate of the powerline hum for well-behaved signals. Since a single value of A_k and ϕ_k is computed for each harmonic, this method removes less of the other signals of interest.

3.1.2 Narrowband Transmitters

While the energy is confined to one frequency interval, anthropogenic narrowband interference signals from VLF transmitters are intermittent and occupy a larger bandwidth around the center frequency. As an example, the Navy transmitters above 20 kHz often transmit MSK signals at 50 Baud and occupy a bandwidth of $\sim$200 Hz.

The powerline removal process was able to take advantage of the fact that the frequency of the fundamental harmonic does not deviate too much from some nominal value, either 50 or 60 Hz. A priori, one may make no such assumption about which frequencies will be corrupted by narrowband transmitters: there are thousands of these transmitters (especially if the possibility of a strong higher-frequency transmitter aliasing into the measured VLF band is considered) and the strength of a given transmitter will be dependent on the distance between the transmitter and the receiver, the propagation path, and the operating schedule of the transmitter itself. To calculate the center frequency of the strongest transmitters, we follow a similar procedure as the STFT-derived approximation of the powerline fundamental. Since the transmitters are man-made and are utilized for long-range communications, it is assumed that the center frequency of each detectable interference source does not change. We therefore average the magnitudes of the STFT, which is equivalent to within a scale factor to Welch's method for calculating the power spectral density

(see Appendix B), thus creating a smooth amplitude spectrum with peaks centered around frequencies with a consistent energy contribution. The frequencies $\{f_i\}$ of the transmitters are simply identified by the center of these resulting peaks (the methods described below do not depend on an exact lock on the transmitter's center frequency, so this non-coherent frequency estimation method is sufficient) and are ordered according to their strength. Usually removing the three or four strongest transmitters is sufficient.

Once the transmitter frequency is known, the problem of isolating the narrowband signal has been considered in other contexts. As part of an experimental technique to monitor the frequency shift and group delay of ducted whistler mode waves at the conjugate region of a VLF transmitter, *Thomson* [1981] presented a technique to coherently demodulate an MSK-modulated subionospheric narrowband channel. However, in full demodulation, each MSK bit maps to a single amplitude value. In our application, the goal is to remove the broadband MSK signal to maximize the SNR of the sferics. Figure 3.6 plots 0.2 seconds of data at a receiver in Hawaii along with the amplitude (in green) of a nearby transmitter (with a call sign of NLK) broadcasting at 21.4 kHz. The amplitude is seen to switch between two different values depending on the bit-dependent transmit frequency, with a complicated bit transition amplitude waveform. The phase trellis for a strong signal may be decoded; see, for example *Shafer* [1994, Ch. 3]. For a weaker signal, however, a phase decode error becomes increasingly likely. Due to the bit-variable amplitude and potential hazards due to decreasing SNR with decoding the phase trellis, we seek an alternative method for mitigating narrowband transmitter interference which has less dependence on an ability to decode the digital signal.

A simple approach is illustrated in Figure 3.6 whereby the demodulated amplitude[3] is filtered to mitigate the influence of the impulsive sferics. In this example, the amplitude range of the transmitter oscillates between 500 and 540, depending on the bit pattern. During a strong sferic, the amplitude is perturbed beyond this range (dashed green lines). The curve shown in blue is the result of saturating the

---

[3]The demodulated amplitude, or the baseband amplitude, is obtained by multiplying the broadband data by $\exp\left[-2\pi f_i t\right]$, applying a complex lowpass filter with a cutoff frequency greater than half of the signal bandwidth, and then taking the absolute value of the result.
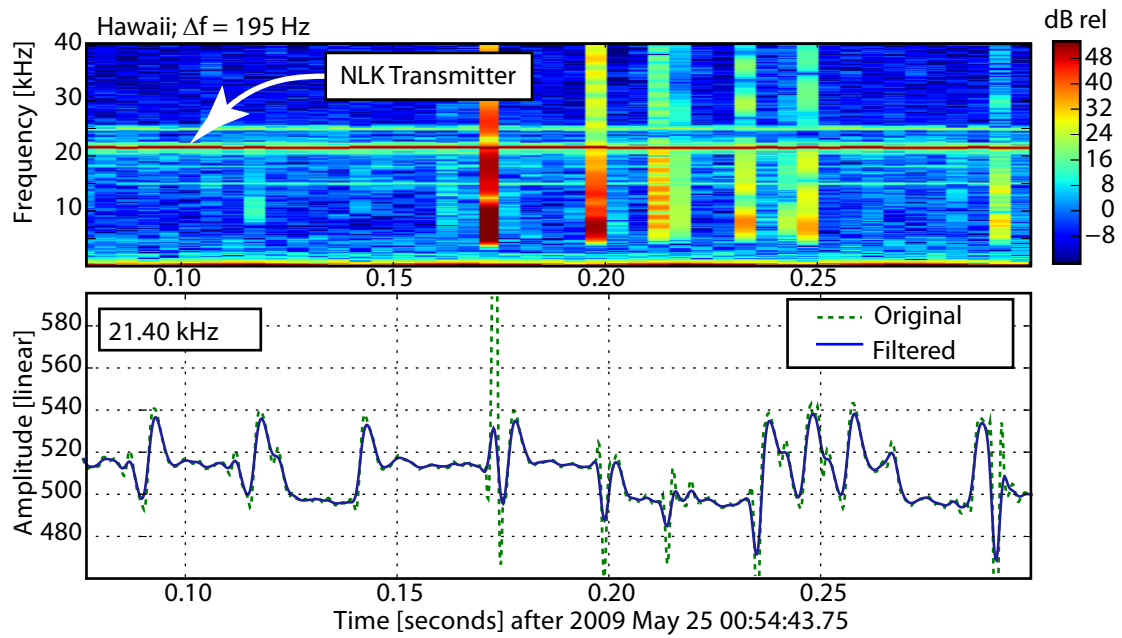
Figure 3.6: Illustration of the procedure for isolating the narrowband channel amplitude in the presence of sferics. The limited amplitude range of the transmitter signal allows us to partially clip the interference from the impulsive sferics.
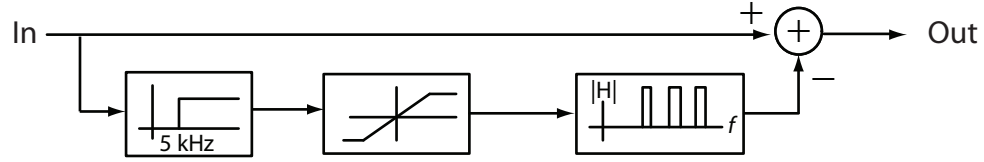
Figure 3.7: Block diagram of the broadband saturation method for reducing narrowband transmitter interference.

magnitude above and below the natural range of the transmitter and lowpass filtering the resulting amplitude. The natural amplitude range is set by calculating a specific percentile range; in this example, the amplitude is saturated at the 1st and 99th percentile levels. The baseband phase is also lowpass filtered to reduce the effects from the impulsive sferics; the two are then recombined, re-modulated, and the result is then coherently subtracted from the original broadband data.

This approach has an important drawback. During the mix-down and filtering step, strong sferics generate a ringing effect due to the bandwidth isolation filter. The baseband amplitude and phase filtering partially removes this ringing by filtering out impulsive deviations from the narrowband signal; however, these filtering operations cause ringing themselves.

Since the impulsive sferics, with a vastly wider bandwidth, tend to have a higher amplitude than the narrowband transmitters, the amplitude of each sferic can be well above the surrounding noise floor raised by the transmitter. Therefore, to mitigate the ringing effect from the narrowband isolation filter, we first high-pass filter the broadband data to remove any DC bias and powerline noise and then nonlinearly clip the impulsive sferics in the time domain. Clipping has the effect of reducing the energy content over the whole frequency range of the early-portion of the sferic. When the narrowband isolation filter is then applied, the impulse response from the sferics is correspondingly attenuated, thus improving our ability to extract the narrowband signal free from impulsive noise. Figure 3.7 illustrates this process, where the passband filtering (only non-zero for the identified interfering signals) may either involve a pure filtering operation or additional filtering on the baseband signal as in Figure 3.6.

The following test procedure was set up to evaluate this procedure for removing narrowband interference. A sferic was simulated as a broadband impulse between 5 and 45 kHz. A real-world transmitter signal was used as the interference source, using the NLK signal from Figure 3.6. In an otherwise low noise environment, this narrowband signal was shifted to one or more frequencies, where each frequency may have a different associated amplitude with respect to the height of the modeled sferic. The above narrowband removal procedures were then applied. The fidelity of the recovered sferic was measured by measuring the mean-squared-error of the idealized sferic to the recovered sferic with a 1 ms window centered at the impulse; the continuous-time equivalent is

$$\text{MSE} = \frac{1}{1 \text{ ms}} \int_{t_0 - 0.5 \text{ ms}}^{t_0 + 0.5 \text{ ms}} [\widehat{s}(t) - s_0(t)]^2 \, dt \qquad (3.8)$$

where $\widehat{s}(t)$ is the recovered sferic waveform and $s_0(t)$ is the original sferic waveform before adding the narrowband transmitters.

Figure 3.8 plots the MSE of the recovered sferic, with amplitude 1000, using one transmitter with a variable amplitude indicated on the x-axis. The MSE resulting from no filtering is indicated by the dashed black line; the MSE resulting from a simple notch filter centered around the transmitter is indicated by the horizontal solid gray line. Also shown are the MSE curves using three techniques. The first approach, which directly isolates the transmitter amplitude and phase and attempts to filter out the impulsive interference from the sferic on those baseband signals is shown in blue. The MSE is improved by a factor of ∼5 with respect to the notch filtering operation. For a transmitter amplitude level ≲7, however, the ringing effect due to the baseband filter causes the overall MSE to become worse compared to the MSE from no filtering.

The second approach involves highpass filtering the broadband signal above the ELF range, saturating the time-domain data to remove some of the energy in the sferics, and then isolating the transmitters by applying a series of passband filters, and finally coherently subtracting the result from the original signal. The MSE from this approach is plotted in green. Above a transmitter amplitude of ∼80, this approach
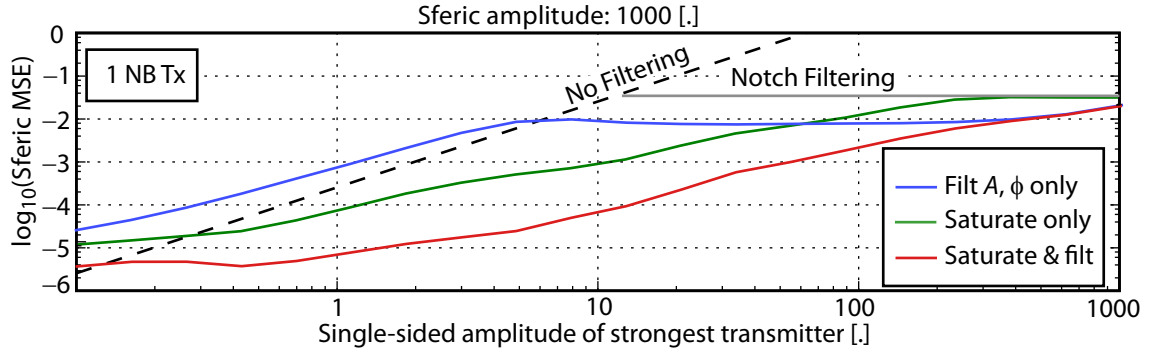
Figure 3.8: MMSE of the sferic after removing a narrowband channel using three techniques.

fares worse than the previous method since the standard deviation of the transmitter is high enough such that the saturation level is also high; correspondingly, not much energy of the sferic is removed and the MSE approaches that of the pure notch filter. Below this level, however, the MSE improves as more of the sferic energy is removed before the notch filter isolates the narrowband signal.

The technique of saturating the large, impulsive sferics in order to better isolate an MSK signal was also used in *Shafer* [1994, Ch. 3], though the stated goal was only to reduce ringing in the baseband filter. Sferic clipping was done after a narrowband filter was applied, so only energy in the sferic which overlapped with the energy band of the transmitter remained prior to the saturation step. This approach differs from the technique employed here, where a highpass filter is applied before saturation which only removes the ELF content. Shafer's approach was applied here as well in the context of Figure 3.8 using one transmitter, yielding results nearly identical to the green curve for a transmitter relative amplitude above 10. For transmitter amplitudes below 10, Schafer's method yields slightly worse results than the green curve, perhaps due to extra ringing effects in the initial bandpass filter. It should also be mentioned that *Shafer* [1994, Ch. 3] introduces a technique to reliably decode the phase trellis of an MSK signal in the presence of impulsive sferic noise, which is a more sophisticated approach to isolating the phase of the MSK signal than the technique used above, which applied a simple lowpass filter to the unwrapped baseband phase.
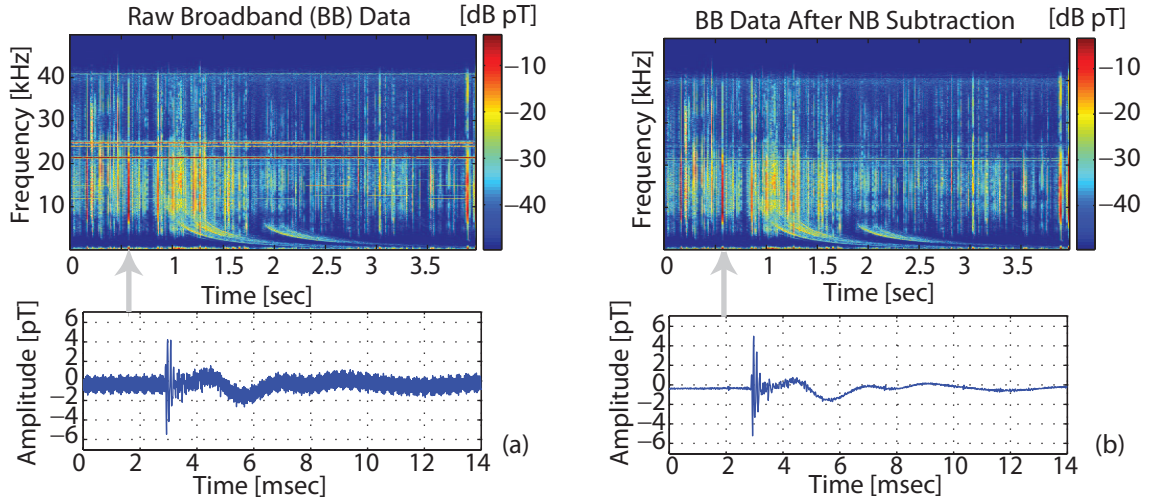
Figure 3.9: Spectrogram and time-domain waveform before and after the narrow-band transmitters are removed using the techniques described in this section.

Decoding and then re-generating the phase may improve the transmitter isolation performance for strong narrowband signals, but the performance gained using this technique likely degrades quickly as the transmitter signal level decreases and an accurate measurement of the phase trellis is no longer possible.

A hybrid approach whereby the broadband signal is first saturated above a certain level, and then the magnitude and phase of each transmitter is isolated and filtered again as in Figure 3.6 (with another saturation step on the transmitter amplitude preceding the lowpass filter) is shown in red. At high transmitter amplitudes, the time-domain saturation has less of an effect and this procedure approaches the results of the blue curve. For lower transmitter amplitudes, this hybrid approach does considerably better than the first two algorithms in isolation.

In a real-world situation, there are many sferic-to-transmitter strength ratios, such that the SNR improvement varies from one sferic to the next. As more transmitters are brought within range of a receiver, the performance also degrades as an attempt is made to filter out more interference sources within the bandwidth of the sferic. Figure 3.9 shows an example from Palmer station, where four separate transmitters were removed. A sample sferic before and after this cleaning process is shown along with a clear slow-tail response after the impulsive VLF component.

3.2   Arrival Azimuth

An integral part of the geo-location algorithm discussed in Chapter 5 is the determination of the arrival azimuth of individual sferic waveforms. This section derives the method used to determine the arrival azimuth using two cross-looped magnetic loop antennae.

Calculating the arrival azimuth from VLF broadband data has been considered in many applications. For example, in analyzing the arrival azimuth of broadband and frequency-varying subionospherically propagating whistlers, *Burgess* [1993] used a Fourier Goniometer. Consider Figure 3.10a, which shows the Poynting vector **S** of an incoming electromagnetic wave at an incident angle θ', measured in degrees East of geographic North. In a ray-hop perspective, the incident wave arrives at an elevation β with, in general, an elliptical polarization (not shown). At a particular frequency f_0, the resulting signals on the North-South and East-West channels form an ellipse when plotted in parametric form. The parameters of this ellipse for all frequencies are readily found by first forming the time harmonic complex signal $z(t) = \text{NS}(t) + j\text{EW}(t)$ and then decomposing the result into the positive and negative frequency Fourier components, as shown in Figure 3.10b. In practice, the Fourier decomposition for all frequencies may be done with a single complex Discrete Fourier Transform (DFT) [*Oppenheim and Schafer*, 1989, Ch. 8]. The major (M) and minor (m) axis lengths, as well as the angle θ and phase ϕ of the ellipse, directly result from the positive and negative frequency phasors $A_0 e^{j\phi_0}$ and $A_1 e^{j\phi_1}$:

$$M = A_0 + A_1; \qquad m = |A_0 - A_1|; \qquad \theta = \frac{\phi_0 + \phi_1}{2}; \qquad \phi = \frac{\phi_0 - \phi_1}{2}. \qquad (3.9)$$

The direction of rotation around the ellipse is determined by the ratio A_0/A_1

$$\text{Rotation direction} = \begin{cases} \text{Clockwise} & \text{if } \dfrac{A_0}{A_1} < 1 \\ \text{Counter-clockwise} & \text{if } \dfrac{A_0}{A_1} > 1 \end{cases}. \qquad (3.10)$$

For a signal consisting of multiple frequencies, an overall arrival angle is found
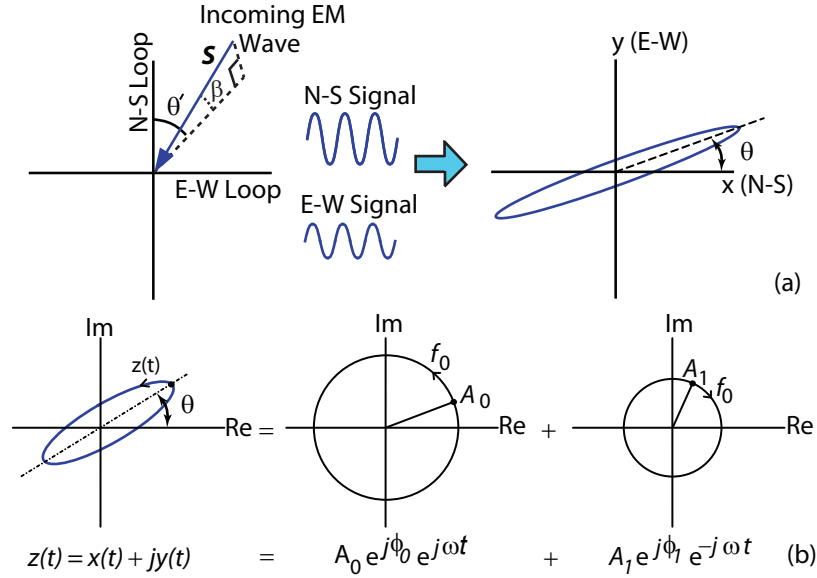
Figure 3.10: Illustration of the arrival azimuth calculation in the frequency domain. (a) Resulting ellipse from an incoming electromagnetic wave. (b) Decomposition of the ellipse into two counter-rotating phasors.

by fitting a line to a scatter polar plot wherein each point is plotted according to its magnitude M and angle θ. In this way, the more robust azimuth measurements that correspond to the highest magnitude are preferentially weighted. An additional artificial weighting may be applied if desired. If all frequencies are used, a best-fit line can alternatively be fitted to the points in a time-domain Lissajous parametric plot. The best-fit line minimizes the squared (Cartesian) distance between each point and the perpendicular distance to the fitted line. This method is discussed further in Appendix A.

Using (3.9), one can plot a time-frequency graph of the arrival azimuth [*Burgess*, 1993, Ch. 3], which is analogous to the magnitude spectrogram. Figure 3.11 shows a spectrogram from the North-South and East-West channels at Palmer, together with the azimuth plot, where the color axis is proportional to the arrival azimuth. The narrowband transmitters arrive from a constant azimuth, as expected. In this example, the impulsive sferics arrive in two angle clusters, one predominantly from the north, the other from the east/west (the lossy Antarctic ice sheet to the south allows us to rule out the presence of north-propagating sferics). One sferic from
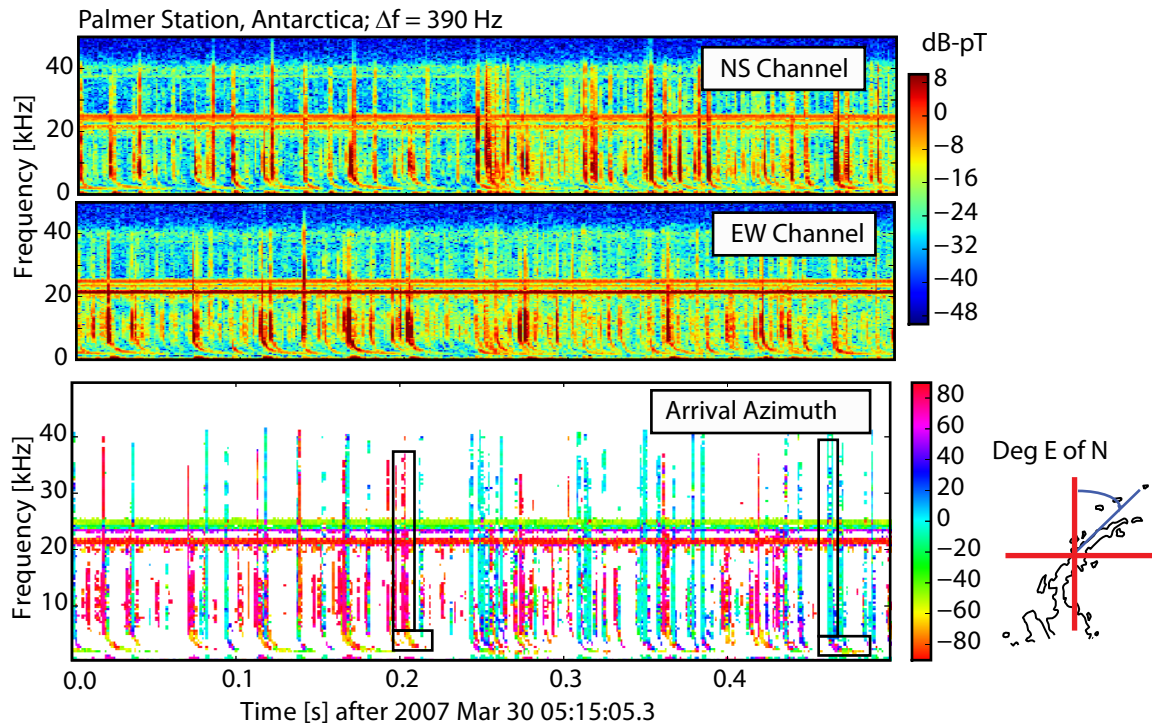
Figure 3.11: Magnitude spectrogram of each channel and a time-frequency display of the arrival azimuth generated using the technique outlined in Figure 3.10. This receiver does not have an electric field antenna, so the arrival azimuth has a 180° ambiguity.

each group is highlighted in the azimuth subplot. The Fourier Goniometric method for determining arrival azimuth assumes that the arrival angle, which is defined as the incidence angle of the Poynting vector **S**, is equal to the major axis angle of the resulting ellipse at each frequency. This is true for surface-propagating waves. Polarization error [*Horner*, 1954] refers to errors in the measured arrival azimuth caused by violations of this assumption. Below ∼5 kHz, the delayed tweeks are seen to arrive from a different angle due to polarization error.

Polarization errors have two interpretations, one based on a ray-hop and another based on a waveguide mode model. In the ray-hop paradigm, polarization error occurs when the incident plane wave has a nonzero elevation ($\beta > 0$) and is not linearly polarized with an electric field vector in the plane of incidence [*Leavitt et al.*, 1978]. In the waveguide paradigm, ellipticity may result if there is a TE component, where there is a component of the magnetic field in the direction of propagation.

Thus, the tweeks, which have endured many ionospheric reflections and therefore many opportunities for mode coupling/elliptically polarized ray hops, tend to introduce errors into the azimuth calculation. Figure 3.12 plots a more detailed example. The top two panels show a sferic after 975 km of nighttime propagation recorded at Taylor, Indiana. The two-channel data has been digitally rotated so that the blue curve corresponds to an antenna pointed at the source with an orthogonal antenna generating the green curve. In the first millisecond of the expanded view on the second panel, the signal is seen to be nearly "linear": there is very little contribution from the orthogonal component. As the signal progresses, however, the sferic becomes more elliptical. The later parts of the signal arrive from sky waves that propagate at more oblique angles, and hence create a progressively more elliptical signal. The bottom set of three panels provide an alternative view in the frequency domain. The first of these plots the magnitude from the two signal components, followed by the angle θ at each frequency and the group delay, $-\partial\phi/\partial\omega$. The horizontal frequency axis is plotted on a log scale to emphasize the lower frequency components. Below ∼1800 Hz, where only the QTEM mode propagates, the magnitude along the 'TM' channel is correspondingly larger and both components attenuate at larger frequencies. After the first waveguide cutoff at 1800 Hz there is a significant contribution
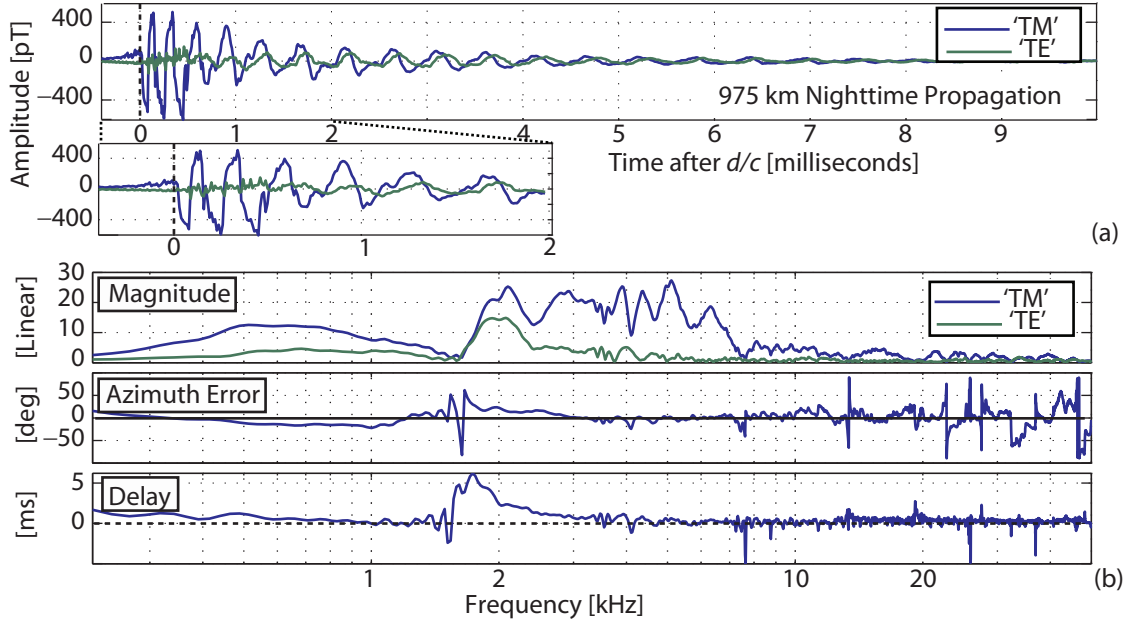
Figure 3.12: Waveform of a short propagation path sferic. (a) Time-domain waveform of a virtual antenna aligned (blue) and orthogonal (green) to the arrival angle. (b) Magnitude spectrum, frequency-dependent arrival azimuth error, and group delay beyond speed of light propagation speed.

from the 'TE' component, which proceeds to fall off at higher frequencies. The presence of the second mode at ~3600 Hz is evident from the more rapid oscillations in amplitude and phase above this frequency, which results from alternating constructive and destructive interference between the first and second modes.

Thus, below ~4 kHz there is a large contribution from the 'TE' orthogonal signal. The presence of this signal contributes to significant azimuth error at these frequencies, as seen in the second subplot in (b). As seen from the group delay panel (c), however, these frequency components arrive later in time. These results are also consistent with previous theoretical modeling. *Yamashita and Sao* [1974], for example, assumed a sharply bounded anisotropic nighttime ionosphere and two propagating modes and found significantly lower polarization errors for the 10 kHz band compared to the delayed 5 kHz band.

A similar plot for a long daytime propagation path is provided in Figure 3.13. In this case, there is very little 'TE' mode contribution, and the arrival azimuth remains
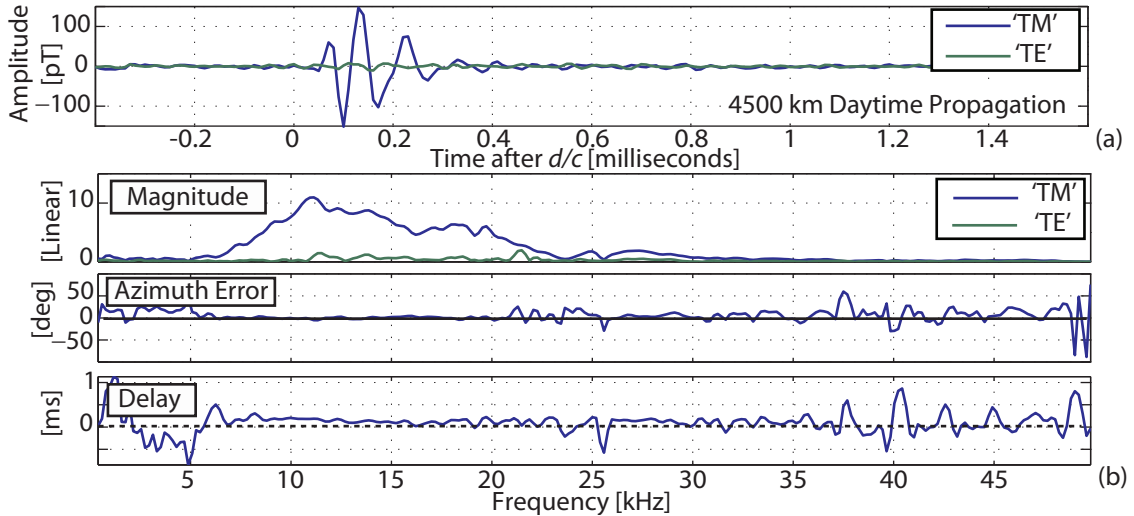
Figure 3.13: Waveform of a long propagation path sferic. (a) and (b) mirror the plots in Figure 3.12, but (b) uses a linear frequency scale to better show the spectral features.

nearly flat across the bandwidth of the signal.

This type of polarization error may be minimized in several ways. High-frequency direction-finding systems [e.g. *Horner*, 1954] minimize polarization errors at close range by using only the initial portion of the ground wave, which has the added benefit of being influenced primarily by the lower, mostly-vertical portion of the lightning strike [*Krider et al.*, 1976]. Nonetheless, with VLF receivers that are tuned to detect distant sferics, receiver saturation for these powerful ground waves out to ~500 km becomes a limiting factor and makes this approach impractical. Another approach [*Lewis et al.*, 1960] is to cross-correlate sferics from nearby receivers and use the arrival time difference to plot the ATD hyperboloid. If the sferic range is much greater than the receiver pair separation, then one of the "tails" of this hyperboloid points in the arrival azimuth direction. The disadvantage of this approach, of course, is the added cost and logistics of additional receivers. Finally, the ELF component, which will consist of only the QTEM component, may be used [*Yamashita and Sao*, 1974]. Not all sferics have a strong ELF component, so this method is not a practical general purpose alternative.

The pattern of a linear signal toward the onset of the sferic, with a more elliptical
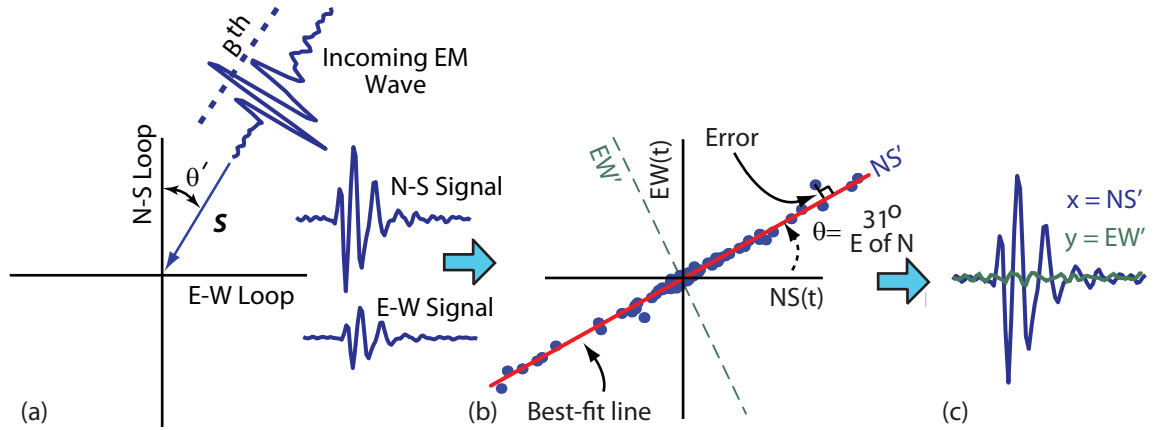
Figure 3.14: Illustration of the sferic arrival azimuth calculation and digital axes rotation. (a) An incoming sferic induces a signal on the North-South and East-West antennas. (b) These signals may be plotted parametrically to calculate the best-fit line and then the arrival azimuth. NS and EW can be time-domain signals $NS(t)$ and $EW(t)$ or frequency-domain signals $M(f)\cos(\theta(f))$ and $M(f)\sin(\theta(f))$, respectively. (c) After the axes are digitally rotated, most of the sferic energy is isolated on one channel.

signal as the sferic progresses, suggests that a determination of the arrival azimuth with only the leading portion of the sferic be considered. Accordingly, the arrival azimuth results presented in this work are calculated by first windowing the sferic around the maximum point, which occurs near the sferic onset, with a window width of 0.2 ms. The arrival azimuth is then found according to the method illustrated in Figure 3.14, either in the frequency or time domain. An incoming sferic that rises above a certain threshold is windowed near the beginning of the waveform and plotted as either a Lissajous or a polar plot. This procedure yields a collection of points that dictate a best-fit line, the angle of which determines the arrival azimuth. The same arrival azimuth calculation scheme is applicable for long propagation-path signals, though is not necessary. For simplicity, the same arrival azimuth algorithm is used on all sferics.

Chapter 4 plots and analyzes several sferic waveforms, with particular attention on their polarity. If an electric field measurement is available, the polarity of a particular feature on a measured wave is unambiguous assuming the grounding polarity of the electric field antenna is known. With only a magnetic field measurement, however, the

polarity depends on the incident quadrant and the winding direction of each magnetic loop antenna. If the north-south loop and east-west loop are perfectly aligned and are wound, following the positive terminal to the negative terminal, counterclockwise when viewed from the east and north, respectively, then sferics incident from the north have the correct polarity after the axes are digitally rotated and those from the south will be multiplied by -1 after this rotation.

To calibrate out this effect in the absence of an electric field measurement, we first digitally rotate the axes to allow us to isolate the sferic on one channel. When this sferic is used as part of a geo-location algorithm, an initial geo-location estimate provides a good approximation of its starting position and therefore the originating quadrant. Using this 'true' azimuth θ_{true}, the polarity can be corrected by flipping the polarity if the following condition is met:

$$\cos\left[\theta_{\mathrm{true}} - (\theta_{\mathrm{calc}} - \rho)\right] < 0 \tag{3.11}$$

where ρ is the misalignment of the north-south axes from geographic north.

Figure 3.15 shows a scatter of arrival azimuth errors for many strikes at various distances, with the 16th, 50th, and 84th percentile ranges calculated using 300 km-wide bins. The performance degrades for strikes closer than 1000 km due a combination of factors, some of which were discussed above. The sky waves are at a steeper incidence, for example, and so the arrival azimuth is more susceptible to a non-vertical polarization. At closer ranges, the saturation of the ground wave at a sensitive receiver may further add to the noise of the arrival azimuth measurement.

As mentioned above, with an additional electric field antenna, the azimuth ambiguity may be resolved. Results in this work were obtained using only a cross-looped antenna; however, introduction of an electric field antenna is perfectly consistent with the methods put forth in this thesis and is thus briefly described here.

With a perfectly conducting surface, only the three field components H_x, H_y, and E_z remain. In this case, the Poynting vector $\mathbf{S} = \mathbf{E} \times \mathbf{H}$ becomes

$$\mathbf{S} = -H_y E_z \widehat{\mathbf{x}} + H_x E_z \widehat{\mathbf{y}}. \tag{3.12}$$
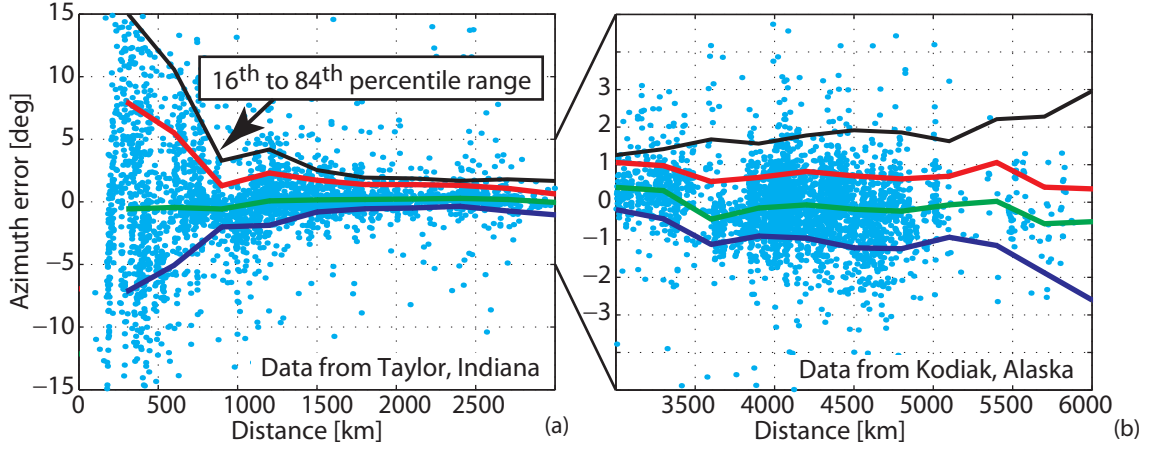
Figure 3.15: Arrival azimuth error versus distance for 0–3000 km range (a) and 3000–6000 km range (b). The 16th, 50th, and 84th percentile curves are shown in blue, green, and red, respectively. The 16th to 84th range is shown in black.

A robust measure of the arrival direction can then by found by evaluating the direction of the time-integrated Poynting vector: $\int \overline{\mathcal{S}} dt$ [*Rafalsky et al.*, 1995]. This technique also helps mitigate the effects of the TE component; integrating the horizontal magnetic field components against E_z brings out the coherent portion of these signals [*Shvets et al.*, 1997]; since E_z only contains the field from the TM modes, this projection preferentially excites those components in the magnetic field measurement.

Using a ray hop perspective, *Leavitt et al.* [1978] applied a vertical antenna with two magnetic loop antennas in the measurement of whistler azimuths. With just two magnetic loop antennas, the measured arrival azimuth θ equals the true angle of the Poynting vector θ' only for low elevation angles or linearly polarized waves. By using the time-averaged measurement of (3.12), *Leavitt et al.* [1978] shows that $\theta = \theta'$ for the same criteria as the magnetic-field-only measurement and also when the incident plane wave has a purely circular polarization or polarization orientations near 0° or 90° with respect to the plane of incidence.

3.2.1 Instrument and site errors

In the preceding section polarization errors were discussed. There are two other important sources of azimuth error, which include instrument and site errors.

Instrument errors are caused by antenna misalignment, improper channel calibration, and mutual coupling between the antennae and the amplifiers. These errors may be minimized with careful antenna alignment and hardware calibration. If there is a misaligned antenna or an unforeseen hardware calibration factor, then a simple equation may be used to recover the idealized measured azimuth. Let ρ be the angle the north-south antenna makes with geographic north and ξ be the extra skew beyond $90°$ of the east-west antenna. If in addition α is the ratio of the magnitude response of the north-south to the east-west antenna (even if the two hardware channels are properly calibrated, there may be a slight gain difference due to discrepancies in the antenna windings), then *Wood* [2004, p. 99] shows that the calibrated azimuth may be calculated from the raw, calculated azimuth by

$$\theta_{\text{cal}} = \tan^{-1}\left[\alpha\frac{\tan\theta_{\text{calc}}}{\cos\xi} - \tan\xi\right] + \rho. \tag{3.13}$$

Another source of error in the arrival azimuth is the so-called site error. A local object capable of scattering VLF radio waves, such as a powerline, fence, or a buried cable, creates an azimuth-dependent gain modification to each antenna. The result is a two-cycle error of the form [*Hiscox et al.*, 1986]

$$E(\theta) = A\sin(2\theta + \phi). \tag{3.14}$$

Any given site may have many such scatterers; the summation still is a two-cycle sinusoid of the form (3.14). However, higher-order harmonics have been measured [*Kawamura et al.*, 1988]. Once this error function is measured, the relation $\theta_{\text{measured}} = \theta_{\text{true}} + E(\theta_{\text{true}})$ needs to be inverted and solved for θ_{true}.

For lighting location systems that operate purely on direction-finding, minimizing site-errors was of primary importance, and so much has been published on measuring (3.14); see, for example, *Hiscox et al.* [1986]; *Kawamura et al.* [1988]; *Schulz and Diendorfer* [1996]. Without a reference network, site-error corrections must be estimated through an iterative process whereby the accumulated cost function from each triangulation is minimized. In *Hiscox et al.* [1986], for example, equation (1.5) from

Chapter 1 of this thesis contained only the azimuth and amplitude terms. Each of these chi-squared cost values then contributed to an overall cost function: with N events, the network chi-squared is $\chi_C^2 = \sum_{j=1}^{N} \chi_j^2$. This χ_C^2 was then minimized over the two variables A and ϕ in (3.14), where each χ_j^2 must be recalculated as these two parameters change. *Passi and López* [1989] expand on this technique to allow for several harmonics of (3.14) and to decouple the site error equations from the localization optimization. In this work, $E(\theta)$ was derived by using NLDN as a reference. It is noted that since both arrival time and azimuth are used here for geo-location, it would not be difficult to produce accurate "self-referenced" site corrections.

Chapter 4

Waveform Bank

As seen in Section 1.2, arrival time accuracy plays a dominant role in geo-location accuracy for long-range lightning location. Short range lightning location systems rely on the initial rising edge of the ground wave, which depends on the ground conductivity and the source term. After several hundreds of kilometers, this ground wave attenuates into the noise and one needs to rely on the sky wave(s) to determine arrival time. To avoid a much higher timing uncertainty σ_t, one needs to account for a widely varying channel response $h(t)$ between the strike and the receiver.

In order to compensate for this varying propagation channel, this chapter sets out to catalog how the received waveform changes under various propagation channels. Using measured sferic waveforms that have been correlated with known strike locations, it is shown that the majority of the received waveforms conform to a canonical shape that depends primarily on distance and the ionospheric profile.

The demonstration of the conformance of waveforms to such canonical shapes indeed constitutes the core innovation of this dissertation, and has formed the basis for International Patent Application 12/152,232. Once it is realized that waveforms conform to canonical shapes, a catalog of such shapes can then be used for a powerful new method of lightning geo-location, as demonstrated in Chapter 5.

4.1 Empirical Canonical Waveform

To investigate the sources of the variability in the received waveform, known strike locations from the NLDN are used to plot the digitally rotated sferic on a time axis that is adjusted for propagation delay at the speed of light.

An example of this procedure is shown in Figure 4.1. Figure 4.1a shows a receiver, indicated by a red 'x', with a storm cluster $\sim$4000 km from the receiver. Each received sferic from this cluster has propagated over a specific ground conductivity [from *Morgan*, 1968] and day/night path. Figure 4.1b indicates the ground conductivity profile along this propagation path, with a zero distance corresponding to the origin of the strike. Figure 4.1c plots in gray each recorded sferic from this storm cluster with a time axis adjusted such that the zero is referenced to the time of the strike plus the delay introduced by propagation at the speed of light. That is, $t_0 = t_{NLDN} + d/c$, where d is the great circle distance from the strike point to the receiver and c is the speed of light[1]. Additionally, each sferic is scaled in amplitude by the NLDN-reported peak current value $I_{\max}$. Given a collection of waveforms with the above time and amplitude adjustments, the three curves represent the 16$^{\text{th}}$ (blue), 50$^{\text{th}}$ (green), and 84$^{\text{th}}$ (red) percentile values of this collection at each time instant. Thus, the blue, green, and red lines capture the behavior of such a collection of sferics by parameterizing the spread at each time instant. Assuming a Gaussian distribution of points at each sample time, the blue and red bounds give the plus/minus standard deviation values; the spread between these two curves is thus indicative of the spread in waveform data contained in the given collection of waveforms at that time. The 50$^{\text{th}}$ percentile line defines an outlier-resistant average to capture the average shape of the received waveform from this storm cluster. We define this curve, referenced to the speed of light propagation line t_0, as the 'canonical waveform' for this specific propagation path.

Figure 4.1 highlights a central observation to this work that is analyzed throughout this chapter. For this specific source-receiver configuration and time, the 50

---

[1]Recall from Section 1.2 that if the arrival time is referenced to this speed-of-light line, the central processor is free to assume speed of light propagation to each receiver, thereby reducing the computational complexity at the central processor.
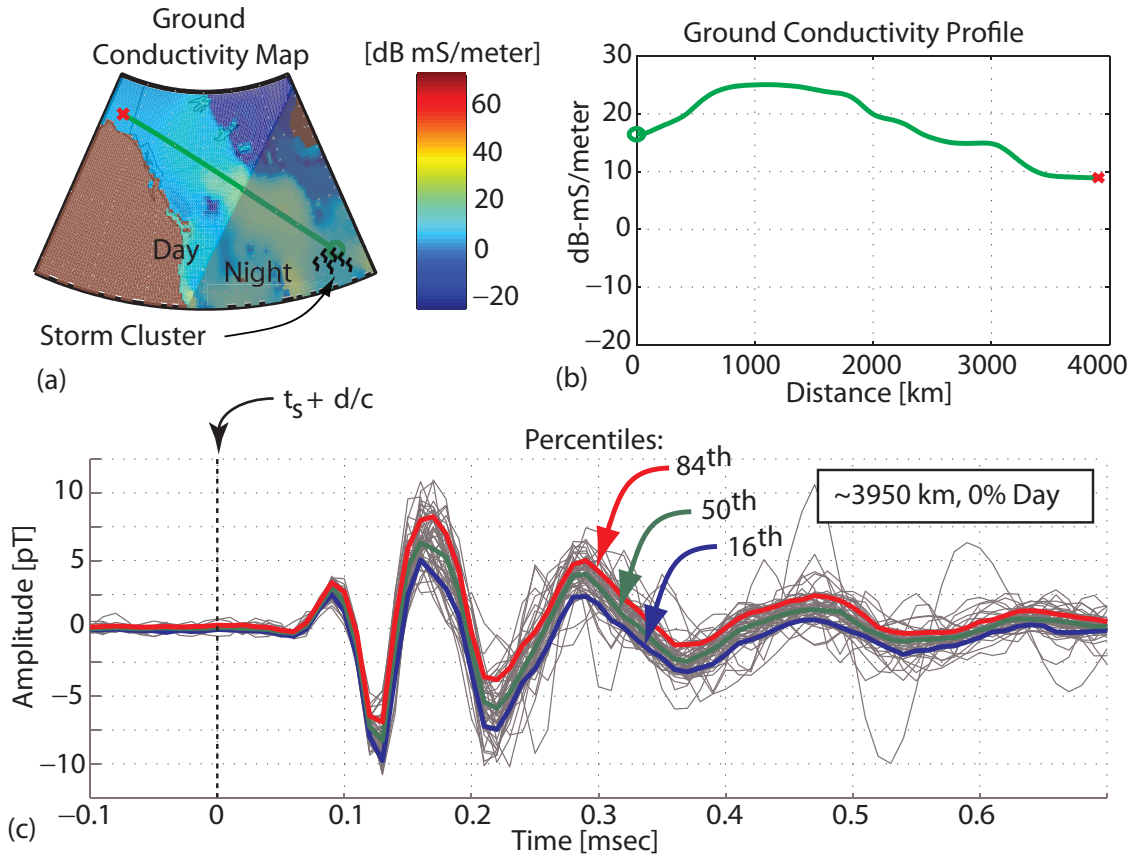
Figure 4.1: Process for measuring an averaged waveform. (a) Storm cluster with a specific propagation path profile to the receiver. (b) Ground conductivity profile between the storm cluster and the receiver. (c) The locus of received waveforms defines a canonical waveform. The shown sferics are from a 100% nighttime path.

polarity-adjusted recorded waveforms conform to a well-defined shape. Variation in the received waveform $r(t)$ depends on the variation in the source term $s(t)$ and the impulse response $h(t)$ of the channel between the source and the receiver per (2.1). By restricting the temporal and spatial location of the source lightning so that the ionosphere is roughly constant and the propagation distance difference between any two sferics in the collection varies by less than $\sim 1\%$, an attempt has been made to isolate the variability in the source term $s(t)$. The literature has relatively few surveys on the measured variability of the received waveform for a fixed path. *Pessi et al.* [2009] show a histogram of sensor delays, referenced to speed-of-light propagation, using NLDN-reported stroke location and times. Using propagation distances out to 4000 km and lumping the delays from all distances into one histogram, the results show a clear clustering at ~ 20 μs for the ground wave, at 52.9 μs (70.5 μs) for the first, inverted sky wave for daytime (nighttime) paths, and a third cluster centered at 90.0 μs (104.0 μs) for the second ionospheric hop for the daytime (nighttime) paths. Each cluster had a spread with $\sigma \simeq 5$ μs. While the relative occurence of each feature was plotted versus distance, with the ground wave providing the predominant trigger at short range and giving way to the first ionospheric hop and, farther still, the second ionospheric hop, the delay of each feature was not plotted versus distance. Nonetheless, the mean of the ground wave cluster is consistent with the delay points plotted for the ground wave in Figure 5.6a, and the cluster locations of the first and second ionosphereic hops are consistent with the zero-crossing delay times for both the daytime and nighttime paths plotted in Figure 5.7. *Lin et al.* [1979] takes a closer look at the spread of waveform features over a short propagation distance, measuring the magnetic field for 46 (77) first (subsequent) strokes and finding a first zero-crossing in the ground wave with a delay of 54 (36) μs and a standard deviation of 18 (17) μs from strikes 200 km away. These measurements are consistent with the first zero crossing seen in Figure 4.3a. In the same study, the rise time to the initial peak was measured at ~ 2.5 μs with a standard deviation of <1.5 μs. This increased consistency at the onset of the waveform is also consistent with the decreased spread in the measured waveshape variability towards the beginning of the waveform, as shown below in Figure 4.3.

The consistency in the received waveforms is not obvious, and this observation forms the bases for the receiver algorithm described in the next chapter. The initial impulse current of a return stroke can last $50-100$ μs [*Lin et al.*, 1979] while the waveguide impulse response can last $\sim$1000 μs. Given these intrinsic time scales with their associated variations in the source and waveguide terms, *Lee* [1986] argued, "With wide equipment bandwidth, the complexities of charge movement through a lightning discharge ensure that sferic waveforms vary quite widely in shape, giving each a unique signature ...". This variation formed part of the rationale for the direct correlation of waveforms in his ATD method. However, it appears that for a given specific propagation path one may indeed define a 'canonical' source term, with the caveat that the expected spread increases over the evolution of the waveform. As is shown later, by assuming a 'canonical' source shape and properly teasing out the dominant factors affecting the propagation path, the best cross-correlation results reported by [*Lee*, 1990] can be approached at each receiver, independently, without needing the source-receiver path to be similar between two stations. This feature allows us to more easily compensate for path-dependent distortions since we can correct data from each receiver independently (as opposed to calibrating pairs of sensors together, as is necessary in an ATD measurement).

Figure 4.2 shows four more such collections ranging from 200 km (a) to 4400 km (d). The number of waveforms and the mean distance of each collection to the receiver is indicated on the plot; all collections are from daytime ionospheric conditions.

The ground wave, first ionospheric reflection, and second ionospheric reflections are indicated by the labels G, 1, and 2, respectively. Our goal is to identify a reliable method to measure an arrival time on each individual sferic. At short range, the ground wave provides an easily identifiable and consistent feature. The onset of this initial rise slowly delays with respect to the d/c line as the ground wave attenuates with distance. Eventually, the ground wave disappears into the noise, so that at larger distances another feature must be utilized to extract an arrival time measurement.

As the ground wave attenuates into the noise floor with increasing distance, the shape of the rest of the waveform changes as well. At larger distances, as the path difference between the ground wave and ionospheric reflections reduces, successive
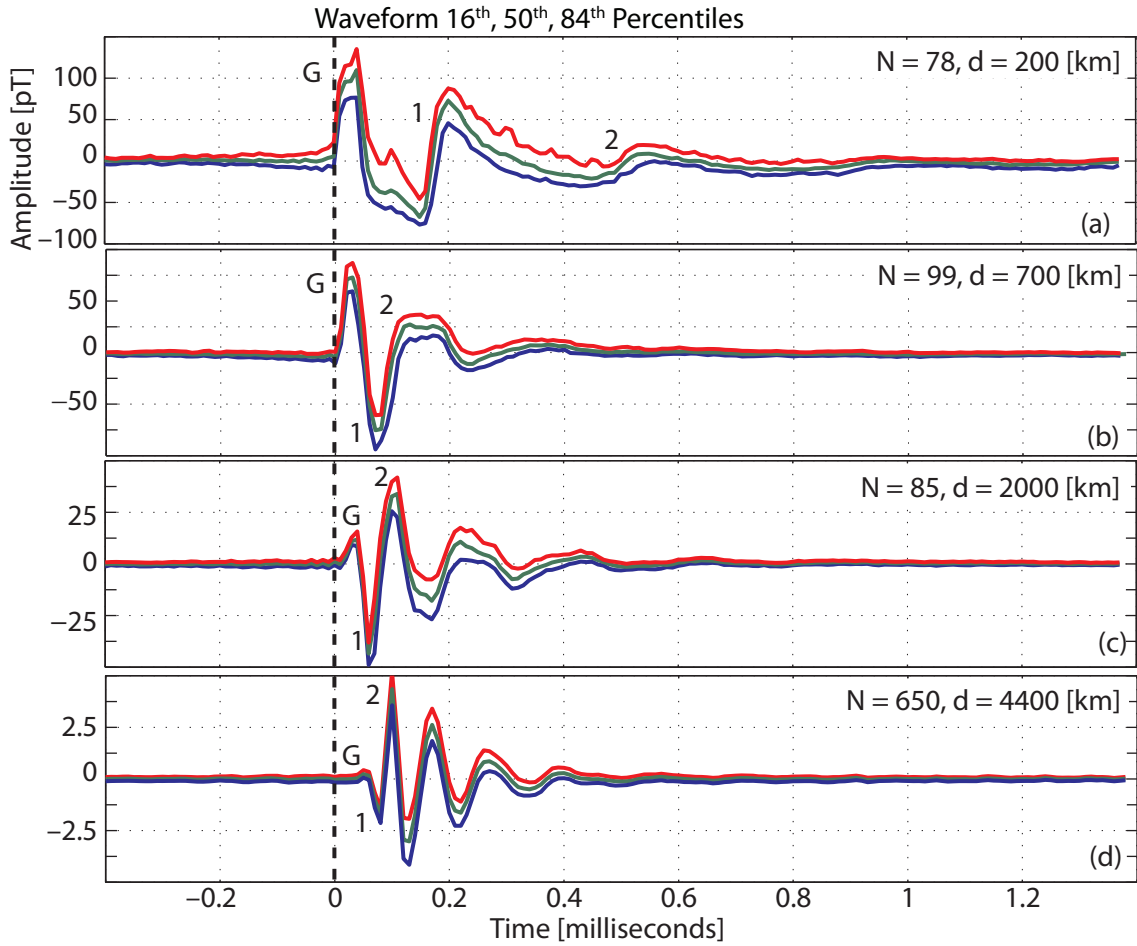
Figure 4.2: Measured daytime canonical waveforms at four distances: (a) 200 km; (b) 700 km; (c) 2000 km; (d) 4400 km.

ionospheric reflections are seen to move in toward the d/c line.

Correct identification of each feature is therefore necessary in order to extract a reliable arrival time measurement. For example, if the point labeled '1' in the 4400 km waveform is identified as the ground wave, the arrival time measurement would be in error by $\sim$100 μs. Furthermore, the non-constant onset of each feature, such as the gradual retardation of the ground wave and the advance of the ionospheric reflections, adds a distance-dependent bias to the timing measurement if not properly accounted for.

Assuming that a particular feature within a received sferic has been correctly

identified, the nonuniform spread in waveforms in Figure 4.2 suggests that some features provide a more reliable timing measurement than others. Figure 4.3, which is an expanded view of Figure 4.2 for two of the distances, illustrates this point more clearly. The spread is seen to increase as the waveform progresses in time, and to be larger for low-slope portions of the waveform. In the 200 km plot, the onset of the ground wave has the lowest feature variance with time. At larger distances, the ground wave has attenuated beyond the point of providing a reliable timing measurement; the double arrows in the 2000 km waveform plot indicate a possible location with a low time variance to look for a consistent arrival time measurement.

As mentioned above, nearby measurements of the radiated field indicate a lower variance for the initial rise time as compared to the fall time. This feature accounts for the minimum spread at the leading edge of the ground wave. At farther distances, when the more consistent early portion of the ground wave has attenuated, extracting timing information from the early portion of the sferic may also help guard against small variability in the path profile. From a ray hop perspective, the early portion has suffered fewer ground and ionospheric reflections and so has fewer chances to deviate from the canonical shape.

It is thus clear from Figures 4.2 and 4.3, that if one knows a priori the propagation distance and the d/c line, one can identify a low-variance portion of the waveform and extract a reliable arrival time estimate. At each receiver, prior knowledge of the propagation distance is not known. However, given the consistency of the received waveform, one alternatively can keep a local record of canonical waveforms, called a waveform bank, and compare each measured sferic to this waveform bank. If the match to one of these locally stored waveforms is strong, then one can estimate the propagation distance and may identify a low-variance feature–the rising edge of the ground wave for nearby lightning strikes; the zero-crossing after the ground wave or the first ionospheric reflection for sferics that have propagated a longer distance.

The above plots demonstrate a heavy distance dependence in the received waveform and suggest that one can mitigate the timing spread due to variability in the source waveform (and possibly the path) by extracting timing information from the onset of the sferic. In the procedure outlined in the next chapter, in order to identify
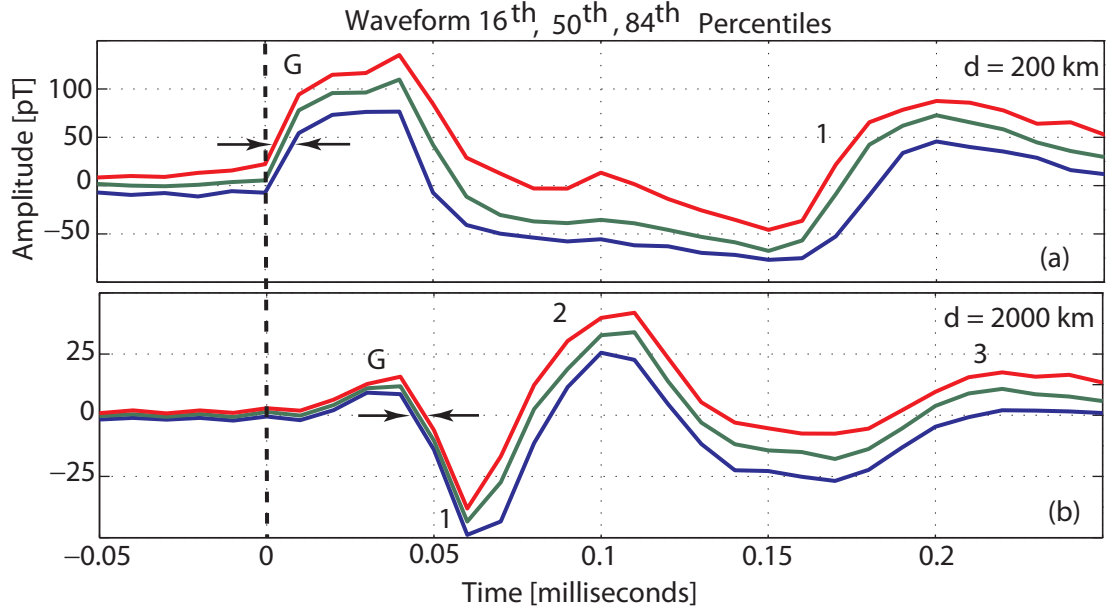
Figure 4.3: Expanded view of the initial portion of the measured daytime canonical waveform distribution at (a) 200 km and (b) 2000 km.

specific waveform features, each sferic is first compared to a locally stored waveform bank. For computational efficiency, it is beneficial to isolate additional factors beyond distance that warrant a separate waveform bank entry. Thus, the following sections explore the effects that the boundaries of the propagation channel have on the received waveform shape, starting with the effects of day versus night. Measured sferics are once again used to build up a canonical waveform for each path profile.

4.1.1 Day/night

The importance of resolving the day/night profile is seen clearly in Figure 4.4. The first subplot of part (a) shows the waveform spread for a collection of 50 sferics from a mixture of day/night percentages, followed by three examples from a fixed day/night percentage, where each subplot is also calculated using a collection of 50 sferics. The top plot of Figure 4.4b plots the delay after the d/c line of the first two zero-crossing points of each sferic that contributed to the collections in part (a). The yellow points correspond to the circled yellow marks; the blue points correspond to the blue x's.
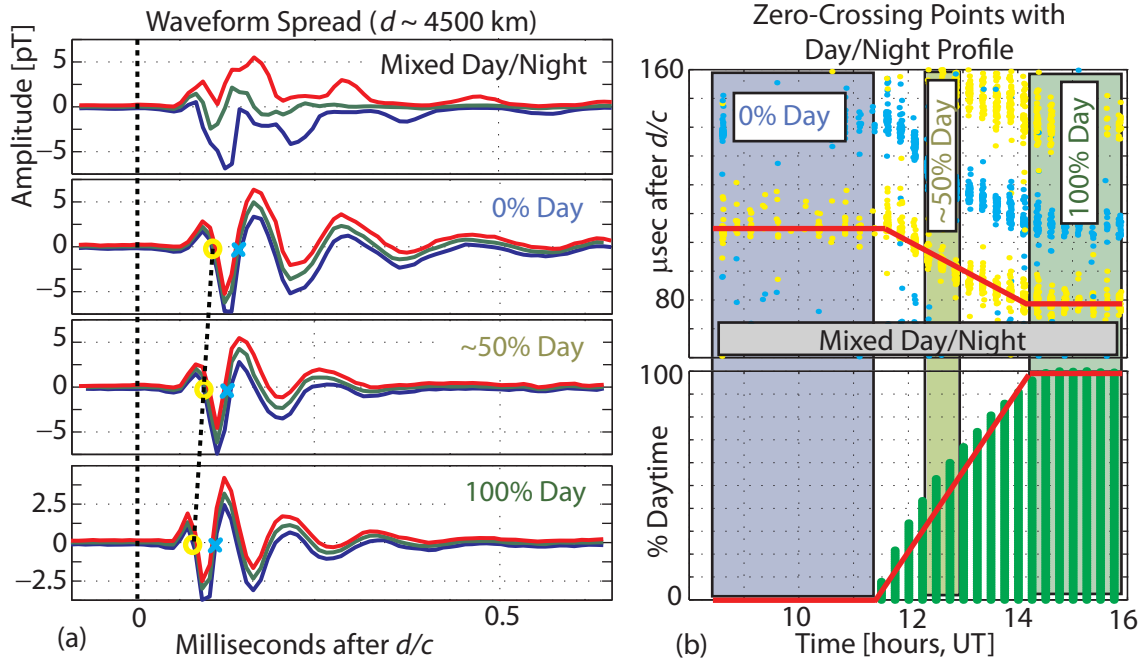
Figure 4.4: Waveform dependence on the ionospheric profile. (a) Waveform averages for a 4500 km propagation distance without resolving the day/night path percentage and with three different day/night path conditions. (b) First two zero-crossing time delays over an eight hour time span, showing the dependence on the day/night path profile.

The bottom plot of (b) shows the percent daytime propagation along the path. From 8–10 UT, the propagation path was all nighttime; after 14 UT, the propagation path was all daytime. Between 12 and 14 UT, the day/night terminator sweeps across the path between the storm and the receiver.

If the day/night percentage is held fixed, for this specific path on the globe the waveform resolves into a regular shape, with a small spread in the zero-crossing points. The delay of these points is seen to vary smoothly from one value to the next as the day/night terminator sweeps across the propagation path.

The difference between day and night propagation is more pronounced at shorter distances. Figure 4.5 plots only the 50th percentile points for day, plotted in green, and night, plotted in blue, at 700 km (a) and 2100 km (b). The ground wave and ionospheric reflections are labeled as before. The ground waves for both daytime
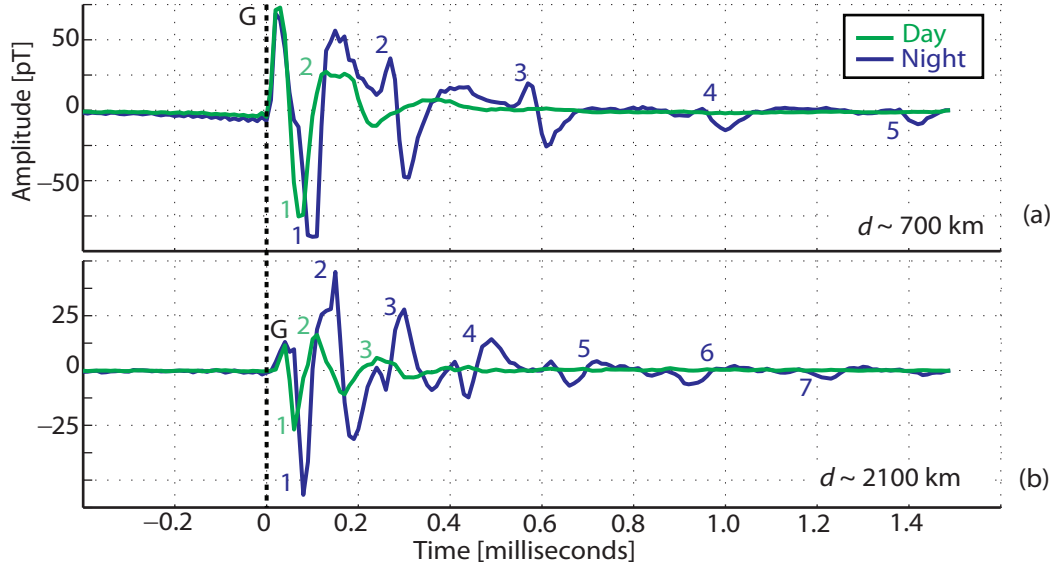
Figure 4.5: Daytime and nighttime measured canonical waveforms at (a) 700 km and (b) 2100 km.

and nighttime paths are time-aligned since this component is not dependent on the ionosphere. The lower attenuation and higher reflection height is clearly evident in the nighttime waveform compared to the daytime waveform, resulting in the tweeks that persist past 1.4 milliseconds from the ground wave. The earlier arrival of the first hops during the day are due to the lower reflection height.

These examples show that distance alone is not a sufficient index to capture the dominant variation in the received waveforms. Therefore we propose a family of waveform banks, where each bank is indexed by propagation distance and has a specific daytime versus nighttime propagation path profile. The results in this work were obtained by using two waveform banks, one for an all-day path, one for an all-night path.

4.1.2   Effects of Ground Conductivity and $B_{\mathbf{Earth}}$

Having established the heavy dependence of the received waveform on the day/night profile, this section proceeds to investigate the dependence on the ground conductivity and the relative direction of earth's magnetic field to the direction of propagation.

Figure 4.6 plots six averaged waveforms from six different path profiles. The specific paths between the storm centers and the receivers is shown in (a), superimposed on a conductivity map for the upper VLF range (10–30 kHz). The conductivity profiles along each path are shown in (b); not shown are the earth's magnetic field profiles. Subplot (c) shows the measured canonical waveforms, which cluster into two groups. The solid curves show the three daytime paths; the dotted curves show the nighttime paths. Figure 4.6 (d) lists the cross-correlation values among the daytime-path waveforms, nighttime-path waveforms, and a 3×3 cross-correlation matrix between the day and night waveforms.

The results summarized in Figure 4.6 allow us to refine our understanding of the dependence of the received waveform on the propagation path and the ground conductivity profile. As can be seen visually in subplot (c) and numerically with the cross-correlation matrices, there is a strong correlation among the daytime curves and nighttime curves and weak correlation between the two, which is consistent with the findings of the previous section. Under a specific day/night profile, however, there is less variation in the waveform due to the nuances of the conductivity profile of each path. The blue path between Florida and California presents a relatively high conductivity path compared to the red path between north-east Canada and Alaska. As expected, the phase of the lower conductivity path is delayed with respect to the higher conductivity path and the amplitude is much lower. Even with these large ground conductivity profile differences, the cross-correlation value of 0.90 is still relatively high compared to the correlation of the waveform from either path to any of the nighttime waveforms.

A similar case is seen between the nighttime pink versus blue path. The lower conductivity path yields an attenuated waveform, yet the correlation between the two is still much higher (a minimum of 0.90 among the three curves) compared to the cross-correlation with a daytime waveform (the highest is 0.82).

The influence of earth's magnetic field is also lumped into these results, as path profiles that had a variation in the magnetic field direction and strength were chosen, though an eastward-propagation path with a similar propagation distance was not available with our current constellation of receivers. Based on the plots in *Wait and*
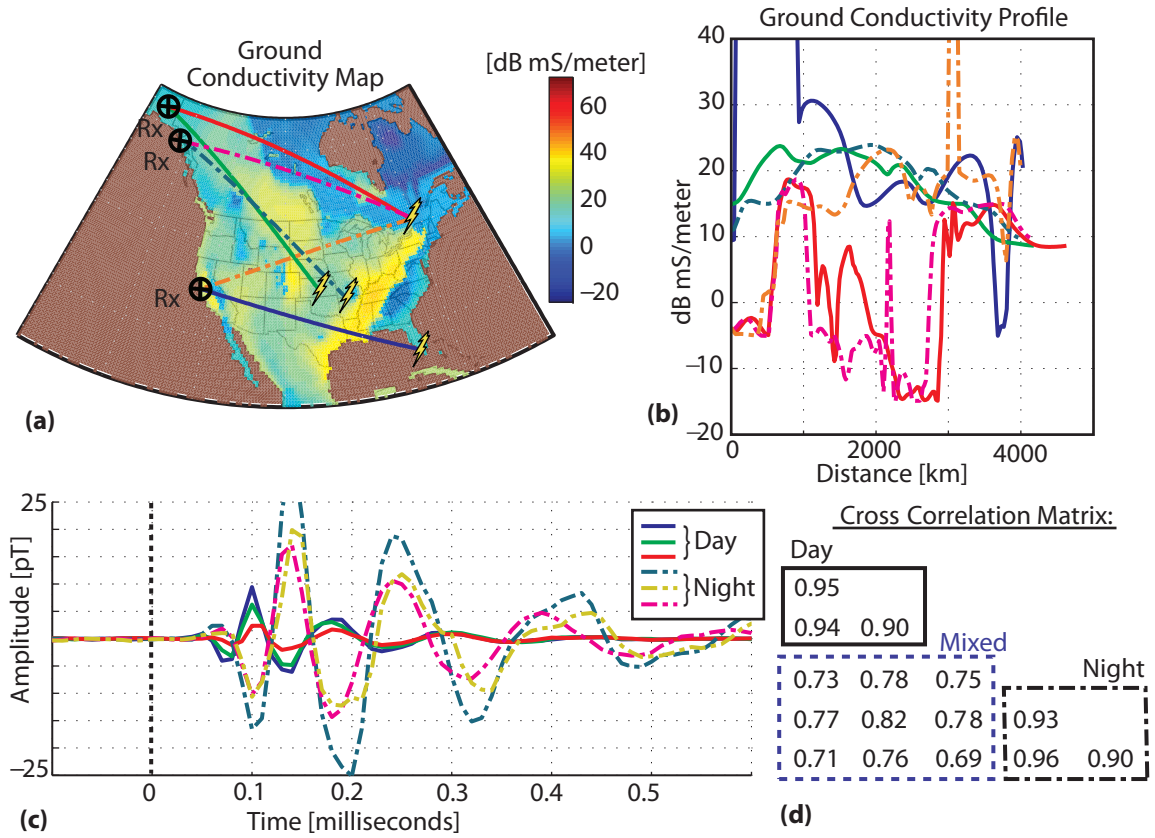
Figure 4.6: Waveform dependence on ground conductivity and magnetic field configuration along the propagation path. (a) Three different nighttime and daytime paths at ∼4500 km. (b) Ground conductivity profile for each path. (c) Measured average waveform for each path. (d) Cross-correlation matrix for the six averaged waveforms, where the value indicates the peak of the normalized cross-correlation result (an autocorrelation would give a result of 1.00). The cross-correlation values among waveforms from all-day and all-night paths are boxed and differentiated from cross-correlation between an all-day and an all-night path.

Spies [1964], one can expect that the phase difference due to the magnetic field direction is on the order of the phase differences that arise from these different conductivity profiles.

Given the similarity, for a fixed day/night profile, among waveforms over large variations in ground conductivities and magnetic field configurations, we conclude that it is possible to capture the dominant variation in waveform shape with canonical waveform banks indexed only by day/night profile and distance[2]. Given these two parameters, a sferic from a 'typical' lightning strike can be cross-correlated with a high degree of confidence to a stored entry in the waveform bank. Smaller perturbations due to the specific path ground conductivity profile (which may be seasonal, due to snow accumulation, for example) and the ambient magnetic field can be accounted for by location-indexed correction matrices at the central processor. These specific path profiles may delay and attenuate the waveform with respect to a canonical shape but have a smaller effect on the phase structure and therefore overall shape of the waveform.

The results outlined above also make intuitive sense. The phase structure is dominated by the propagation distance and the effective height of the ionosphere, with smaller perturbations introduced by the phase variations in the reflection coefficient on each boundary. While there is a large range in the magnitude of these reflection coefficients, the phase variation is small compared to the additional phase introduced by a variable ionospheric height.

4.2 Properties and plots

In the previous section, it was found that one can capture the dominant waveform features by indexing canonical waveforms by distance and the day/night propagation

---

[2]As discussed further in Chapter 6, if there are exceptions to this generalized conclusion, for example for extremely low conductivity paths or perhaps under certain magnetic field profiles, then a supplementary set of waveform bank entries may need to be supplied to certain receivers

path profile, resulting in "bands" of waveforms. This section presents several summary plots and highlights the dominant features of both measurement- and model-generated waveforms. Each bank has a specific day/night path profile percentage; for example, 100% daytime or 100% nighttime, and each waveform entry corresponds to a specific propagation distance. Each waveform entry represents the 50th percentile level using the averaging technique presented in Figure 4.1. Four receiver stations are used in acquiring the data: Taylor, Indiana; Santa Cruz, California; Juneau, Alaska; and Midway Island. NLDN data was used as a reference over the US, and each station gave a semi-unique distance range with which to measure the averaged waveforms. Each station is calibrated (see Appendix B) so that the measured waveforms are referenced to the magnetic field strength in picoteslas (pT) at the magnetic loop antenna. Path profiles with reasonably high conductivity are chosen so that the magnitudes aligned with reasonable consistency, although a combination of eastward versus westward profiles was unavailable.

4.2.1 Empirical Waveforms

Figure 4.7 illustrates how each waveform bank is displayed in the subsequent plots. This figure shows a top-down view of a daytime waveform bank where each waveform has been normalized so that the maximum amplitude is unity. The time axis (abscissa) is referenced with 0 corresponding to a speed-of-light propagation; the distance index is indicated on the ordinate. Each row indicates a specific waveform bank entry. The color scale indicates the polarity, as illustrated by the waveform on the right. The various features may be readily identified: the ground wave dies out beyond about 1500 km, and the sky waves move in with respect to the d/c line as the ionospheric grazing angles become more shallow. The distance is plotted on a log scale to best capture the rate of feature changes.

Figure 4.8a shows the day and night waveform banks on a linear vertical scale from a 3-D perspective. On this scale the exponential decay prevents us from seeing the features beyond 3000 km. These plots are repeated in Figure 4.8b with the vertical scale normalized independently on each waveform so that the maximum absolute
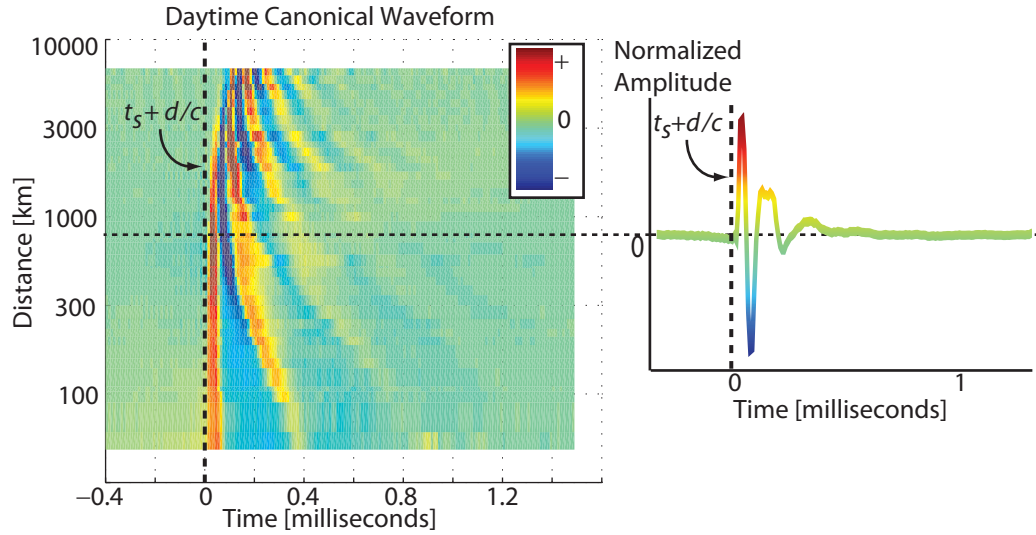
Figure 4.7: Top-down view of a normalized daytime waveform bank with a sample entry illustrating the color mapping.

value is 1.0. Comparing the daytime to the nighttime waveforms, the subsequent hops from the nighttime path arrive later due to the higher reflection height. Also, the subsequent hops are more pronounced at night due to the lower attenuation from each reflection.

Figure 4.9a plots the spectrum of each entry from the waveform banks in Figure 4.8 using the magnitude of the DFT. The distance axis is again plotted on a log scale to highlight the salient features. In the time domain, the successive ionospheric reflections contract in time with distance as the differential distances decrease. In the frequency domain, the inverse happens: the deep nulls from the interference between the ground wave and subsequent reflections are seen to increase their separation with frequency at greater distances.

During daytime conditions the amplitude spectra smooths out considerably beyond 1000 km. One can conclude that past this point there is one dominant mode. The sharp falloff in amplitude at 3000 km in both banks is partly due to the logarithmic distance scale since the amplitude follows a roughly linear dB attenuation with distance. At close range, the peak amplitude occurs near 3 kHz, though at this distance the amplitude spectrum may be artificially skewed to lower frequencies due
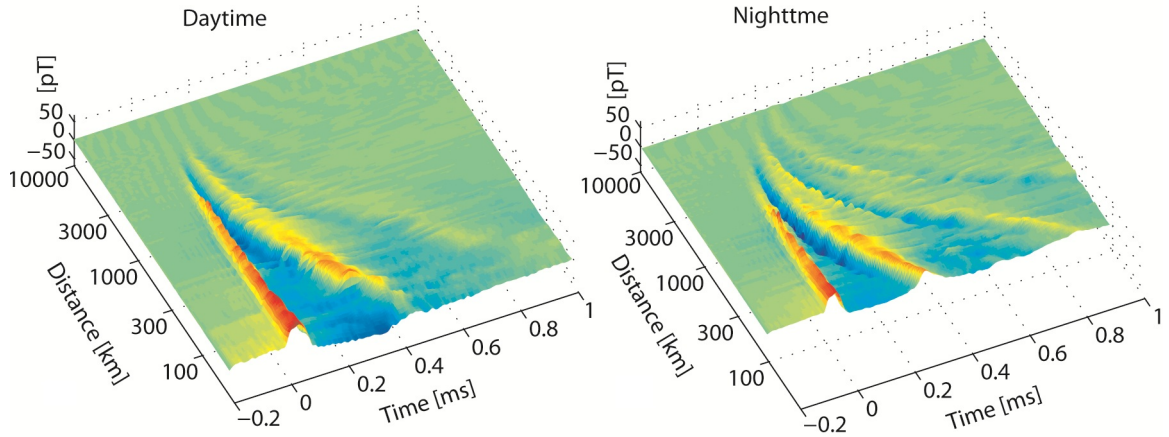
(a) Calibrated measured day and night time-domain waveform banks.
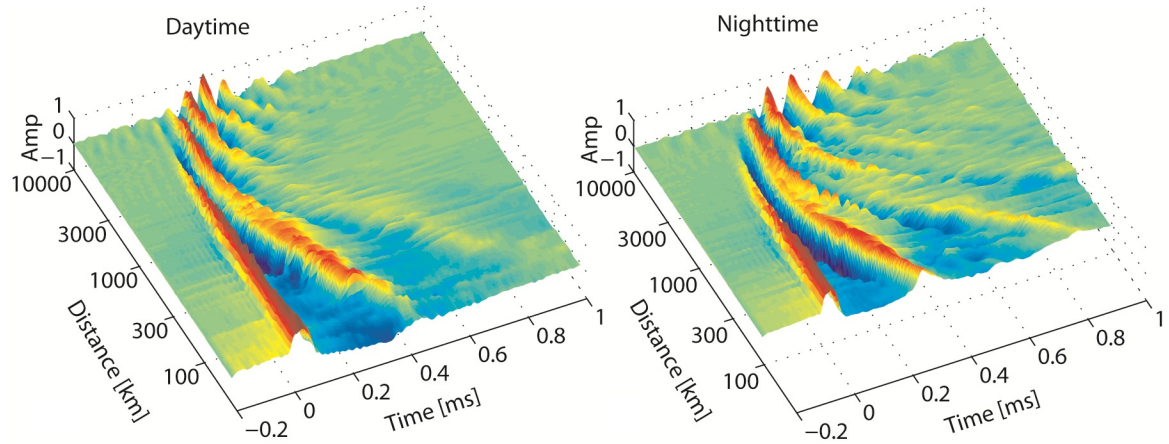
(b) Normalized measured waveform banks.

Figure 4.8: Calibrated (a) and normalized (b) time-domain waveform banks for daytime (left column) and nighttime (right column) propagation paths.

to saturation at the receiver. At 1000 km, the spectral peak has shifted up to ∼6 kHz as the lower frequencies suffer high attenuation (see Section 2.5). At 10,000 km, the spectral peak is near ∼13 kHz, shifting toward the minimum attenuation frequency of ∼18 kHz calculated in *Wait and Spies* [1964] (the measured spectrum will be the product of the radiated spectrum from the discharge and the channel response of the path).

Features similar to those observed in the daytime waveform bank spectra can be observed in the nighttime waveform bank spectra. These spectra were taken from only the first 1 millisecond of each waveform, so the interference pattern from the late-time tweeks is not seen. As with the daytime waveforms, the spectral peak is near 3 kHz at 100 km, rising to ∼6 kHz at 1000 km. At 10,000 km, the spectral peak is slightly lower compared to the daytime waveforms at ∼9 kHz, which is consistent with a lower minimum attenuation frequency for nighttime paths (∼12 kHz) predicted in *Wait and Spies* [1964]. The 'lumpiness' in the nighttime waveform bank spectrum reflects the more variable nighttime conditions compared to the daytime waveform bank. These waveform banks reflect measured sferics from several stations over many different days and path profiles. The correlation techniques presented in the next chapter do not depend on a steady received amplitude but rather a consistent phase profile; therefore, the inconsistent amplitude in the nighttime (and up to a point also in the daytime profile) has little effect on the accuracy of the ultimate lightning geo-location result.

Using the same DFT but now utilizing the phase, Figure 4.9b plots the group delay of the daytime and nighttime waveform banks, $\delta(\omega) = -\partial\phi/\partial\omega$, where ϕ is the phase of each DFT index. To minimize the noise shown on each plot, only group delay points corresponding to magnitudes of a sufficient level are shown. As seen before, the modal interference pattern is seen out to a certain distance, after which point the profile becomes smoothed out. After ∼3000 km, the group delay profile follows the canonical pattern of a parallel plate waveguide: the lower frequencies are dispersed and arrive later in time than the higher frequencies.

These empirically-measured waveform banks anchor the lightning geo-location algorithm discussed in the following chapter. For a particular day-night path profile,
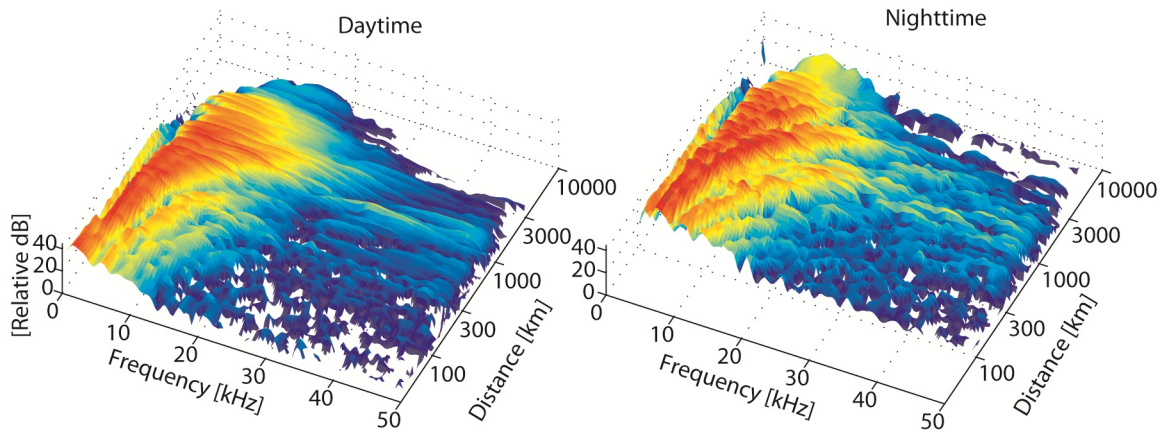
(a) Spectrum of the measured daytime and nighttime waveform banks.
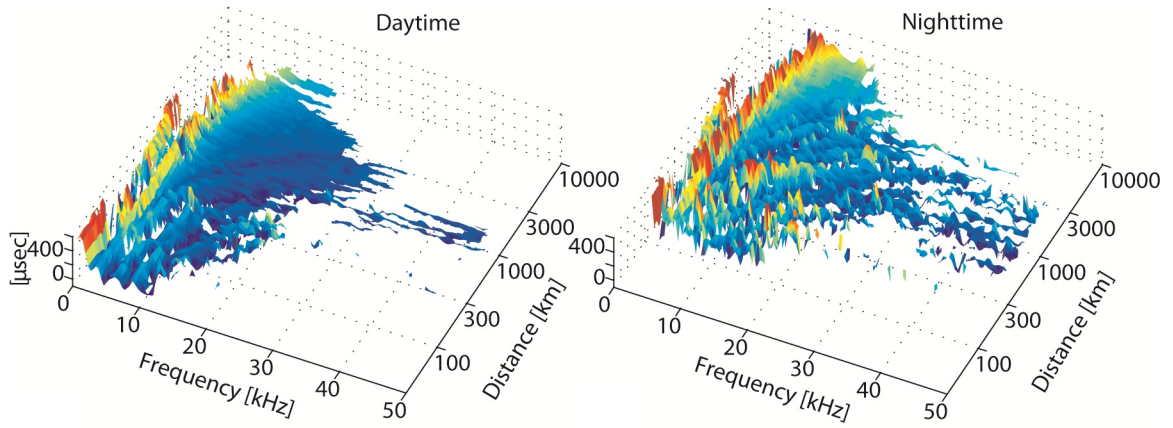
(b) Group delay of the measured daytime and nighttime waveform banks, referenced to the d/c line.

Figure 4.9: Magnitude and group delay of the measured canonical waveform banks.

the average sferic waveform exhibits a very predictable pattern with increasing distance, with more minor modifications due to variable ground conductivity changes and magnetic field configurations. We use this information to our benefit to form a phase-coherent measurement at each geographically separated receiver, which allows for a far greater location accuracy over varied propagation paths than other long-range networks in existence today.

While we use the empirical waveforms for the geo-location results presented in this work, it is both practical and instructive to derive the waveforms through numerical modeling. After dominant propagation effects introduced by distance and the day/night profile have been corrected for, there are still perturbations in the measured arrival time (and amplitude) introduced by variations in the waveguide due to the specific ground conductivity and the Earth's magnetic field profile. In the geo-location algorithm outlined in the next chapter, these extra, path-specific variations are corrected for using empirical correction matrices, which are built using NLDN-reported strike locations. These final correction matrices may also be constructed using a redundant (>3 stations) receiver constellation, where redundant arrival time measurements are used to estimate location-indexed biases at each receiver by minimizing a global network chi-squared value, in much the same way as ATD biases are removed on a per-region basis in *Lee* [1990]. These empirical methods both require sufficient lightning data in each region and either an existing medium-range network or at least four sensors within range of the regions in question. A more expedient approach would be to use a numerical model to predict the offsets at each receiver using an estimate of the specific path profile. Also, modeled waveforms could be used to fill out the waveform bank for paths that have a waveguide profile that is sufficiently different from the source-receiver profiles available in the empirical waveform bank. For example, insufficient data may be available to fill in certain mixed day-night paths using measured waveforms. Additionally, if there is an unusual path profile, such as a path over a high icy mountain range, the received sferics may not conform sufficiently well to the canonical shape measured over a high conductivity profile, even if the day/night profile and the distance are the same. In such cases, a

special waveform bank, derived with a robust model, could be used for that particular receiver for sferics whose arrival angles indicate propagation over this particular low-conductivity path.

A detailed and comprehensive waveform bank model is beyond the scope of this thesis, but, using some simplifying assumptions, one can reproduce many of the dominant features seen in Figures 4.8, 4.9a, and 4.9b. These modeled results show a basic analytical understanding of the features in the waveforms, and we may therefore be able to use the model in future work to help fill in the gaps in the waveform bank.

4.2.2 Ray-Hop Model

In this section, a homogeneous ground and a vertically inhomogeneous, highly collisional, isotropic ionosphere is assumed. The reflection coefficient is calculated using the iterative method from *Wait* [1970, p. 406], which matches the impedance at successive homogeneous slabs to propagate the reflection coefficient from the top of the ionosphere to a region below the ionosphere. In addition to assuming the isotropic condition $Y = 0$, $Z \gg 1$ is assumed so that $n^2 \simeq 1 - j\omega_r/\omega$ (see Section 2.1). As in *Wait and Spies* [1964], the ionospheric conductivity profile used here is

$$\sigma = \epsilon_0 \omega_r = \epsilon_0 (2.5 \times 10^5 s^{-1}) e^{\beta(z-h)} \tag{4.1}$$

where β describes the slope of the D region profile and h provides a reference reflection height. A shallow slope (low β) provides a smooth transition between free space and the lossy ionosphere, creating a quasi-WKB condition which draws the wave in and, thus, gives a lower reflection coefficient (suffering more loss before the wave is reflected). A steeper slope presents a sharper boundary and so leads to less loss upon reflection. The former condition represents a daytime profile, with $\beta \sim 0.3$ km^{-1}; the latter a nighttime profile with $\beta \sim 0.5-0.7$ km^{-1}.

Figure 4.10 shows the resulting reflection coefficient for the parallel polarization (representing the TM mode, with a vertical electric field), adapted from *Wait* [1970, pp. 410,412]. In this figure, θ is measured from the vertical. Near vertical incidence the reflection coefficient is much lower for the lower β values. Also shown is a sharp
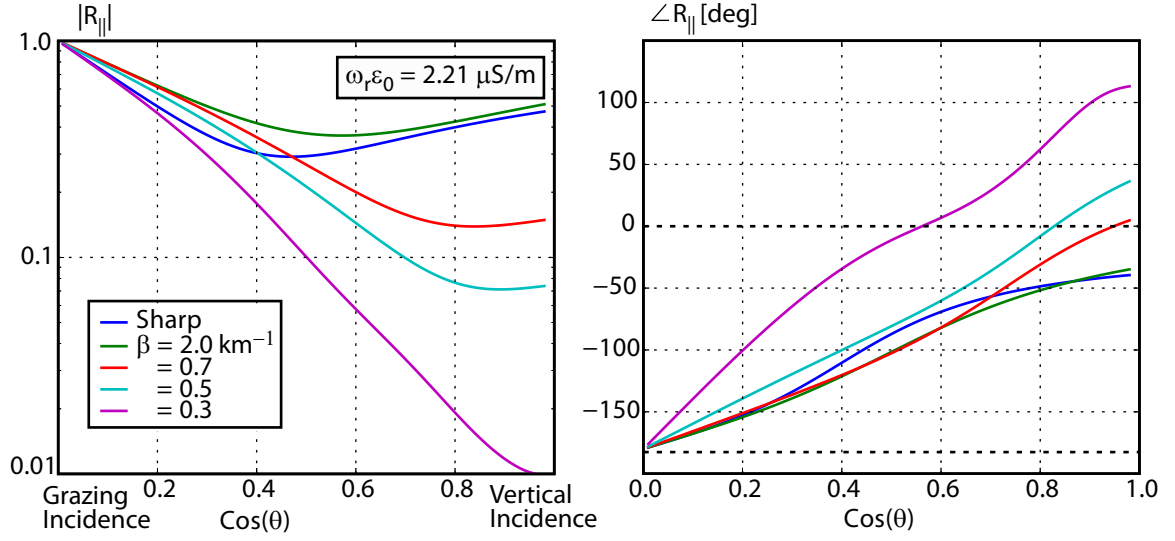
Figure 4.10: Magnitude and phase of the reflection coefficient for an isotropic iono-sphere for various exponential slopes. Adapted from *Wait* [1970, pp. 410,412].

reflection boundary calculated using (2.15), which follows closely the high β value of 2. A quasi Brewster angle is seen with a minimum in the reflection coefficient for the sharp transition regions and less visibly for the more gradual transitions.

For all ionospheric profiles, the reflection coefficient changes sign from vertical to grazing incidence. For the lower-order modes, where $\cos\theta$ is small, the reflection coefficient is ~-1; hence, the ionosphere approximates a magnetic conductor for lower order VLF modes.

Equation (2.33) is used for the source current and velocity. The channel length profile $l(t) = \int_0^t v(\tau)d\tau$, current profile, and the resulting current-moment are plotted in Figure 4.11. The resulting ground wave radiation electric field term, assuming a perfectly conducting flat earth and the *Bruce and Golde* source model, is, from (2.28),

$$\frac{\mu_0}{2\pi d}\frac{\partial\left[i(t)l(t)\right]}{\partial t} \tag{4.2}$$

and is plotted in Figure 4.11 for the this current-moment profile. Note the logarithmic time scale which captures both the quick rise and slow decay that is characteristic of
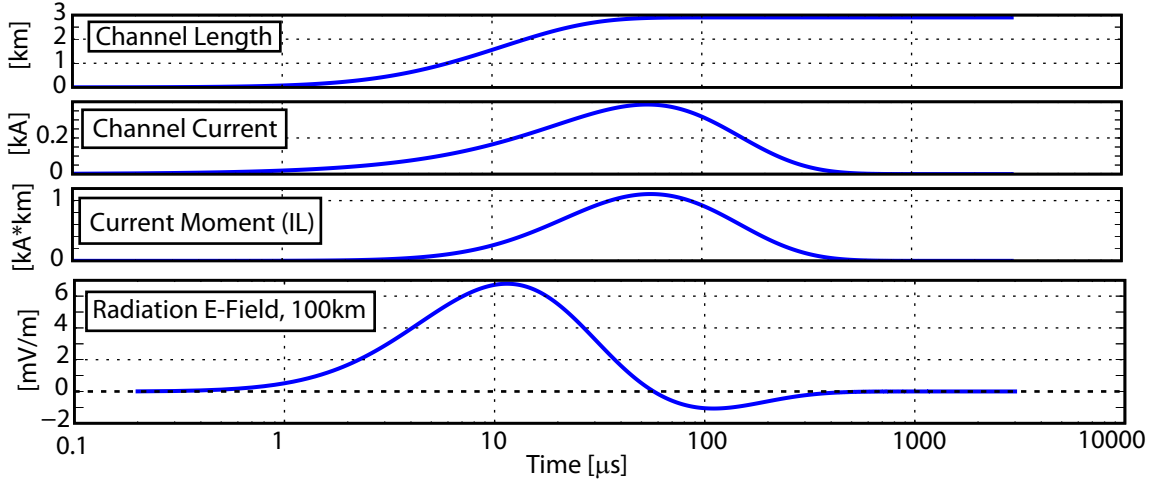
Figure 4.11: Radiation source term. (a) Channel length; (b) channel current; (c) current-moment; (d) resultant radiated electric field normalized to 100 km.

the *Bruce and Golde* model.

The model parameters used here are velocity parameter $a =$9e4 s^{-1} and current parameters $\alpha =$1e4 s^{-1}, $\beta =$3e4 s^{-1}, which give a radiation-term zero-crossing delay value of $\sim$60 μs, consistent with the measurements of *Lin et al.* [1979]. The 10%–90% rise time of the radiated field using the above parameters is longer than the rise time previously reported in the literature. This slower rise time was chosen to improve the fit between the modeled results and the measured waveforms. This discrepancy between the faster rise time reported in *Lin et al.* [1979] and the rise time in the model used here may result from the simplifying assumptions inherent in the *Bruce and Golde* model and perhaps an insufficient dispersion in the channel response, especially for the ground wave component.

The radiation electric field has an initial rise that corresponds to an upward current. Physically, for a cloud-to-ground downward stroke, this initial rise corresponds to lowering negative charge to the ground; hence, the results and waveform banks in this chapter are all referenced to a negative cloud-to-ground stroke.

Using this source term and the reflection coefficients plotted in Figure 4.10, we build a simple ray-hop model using (2.30) and (2.34). In the former equation, we assume a ground conductivity and relative permittivity $\sigma =$0.4 mS/meter and $\epsilon_r =$15,
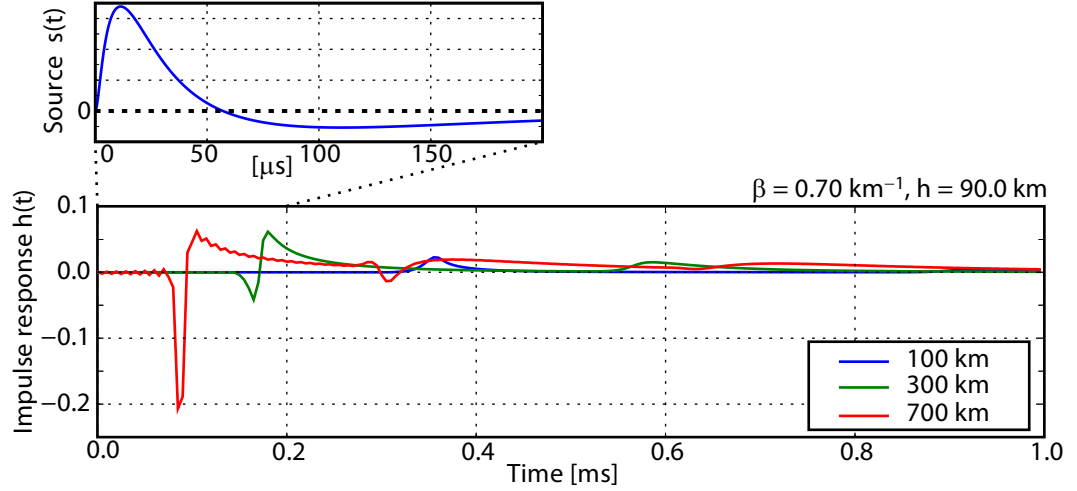
Figure 4.12: Radiation source term and impulse response at three distances for a nighttime ionosphere.

respectively. In the latter equation, to simplify the model, we set $F_t = F_r = 1$, retain only the spherical spreading term from c_m, and assume a perfectly conducting earth $[R_g(\psi) = 1]$. The measured horizontal field for the ground wave is $\mathcal{B}_{\phi,\text{gnd}} = B_0(t) * \mathcal{F}^{-1}\{W\}$, where W is given in (2.30), and the m$^{\text{th}}$ ionospheric hop is expressed as

$$\mathcal{B}_{\phi,m} = B_0 * \mathcal{F}^{-1}\left\{\cos\psi\,\frac{d}{R_{\text{E}}\sin(d/R_{\text{E}})}R_i^m(\theta)\frac{\mathrm{e}^{-jk\Delta d}}{1 + \Delta d/d}\right\} \qquad (4.3)$$

where $\Delta d_m = d_{s,m} - d$ with $d_{s,m}$ given by (2.36) and

$$B_0(t) = \frac{\mu_0}{2\pi dc}\frac{\partial(il)}{\partial t}. \qquad (4.4)$$

Here, $B_0(t)$ is defined as the source term which convolves with the impulse response $h(t)$ to give the measured field. The impulse response for the ray hop terms (excluding the ground wave) for three distances under a nighttime ionosphere, with $\beta = 0.7$ km^{-1}, $h = 90$ km is shown in Figure 4.12. As one would expect from the antenna pattern of a vertical dipole, which has a maximum in the horizontal direction and a minimum directly above the source, coupled with the increasing value of the reflection coefficient with shallower reflection angles, the ray hop amplitudes increase

with increasing distance (note that $s(t)$ contains $B_0(t)$, which absorbs the loss introduced by spherical spreading; the impulse response contains only the correction factor introduced by the spherical waveguide structure). Also, as the phase of $R_\parallel$ transitions from ~ 0 in the limit of vertical incidence to $\sim 180°$ in the limit of grazing incidence, the polarity of the ionospheric hop flips from positive to negative. This polarity reversal at these two extremes accounts for the labeling of the large first negative peak as the first ionospheric hop in the waveforms throughout this chapter.

Convolving the impulse response with the source term yields the results shown in Figures 4.13a,b. For comparison, the modeled waveforms are plotted against the measured 50$^\text{th}$ percentile waveforms. The labels on the ground wave indicate an artificially imposed scaling on just the ground wave to account for saturation at the receiver. There is a strong similarity in waveform shapes and zero-crossing values between this simple model and the measured waveforms. At large distances, the model over-attenuates the low frequencies in the ground wave, but the sky waves continue to be faithfully reproduced. At 1600 km under the nighttime propagation, the successive sky waves in the measured result are likely stronger due to the static magnetic field. The overall agreement solidifies our understanding of the evolution of the waveform shape out to $\gtrsim$1200 km and provides a starting point for more elaborate models in the future. Due to the simplifying assumptions inherent in the *Bruce and Golde* model and because the model parameters were adjusted to match properties of the radiated term to previously recorded waveforms, this agreement between the model and recorded sferics does not strengthen our understanding of the physical lightning discharge process.

4.2.3   Modal Attenuation and Group Velocity

At great distances from the receiver, a modal model which was outline in Chapter 2.1 is more appropriate. Figure 4.14 shows calculated group velocity and attenuation curves under an isotropic daytime ionosphere along with measured values using a differential distance between 4000 and 5000 km. Both the group velocity and attenuation oscillate slightly and are clustered close to the first mode, indicating that
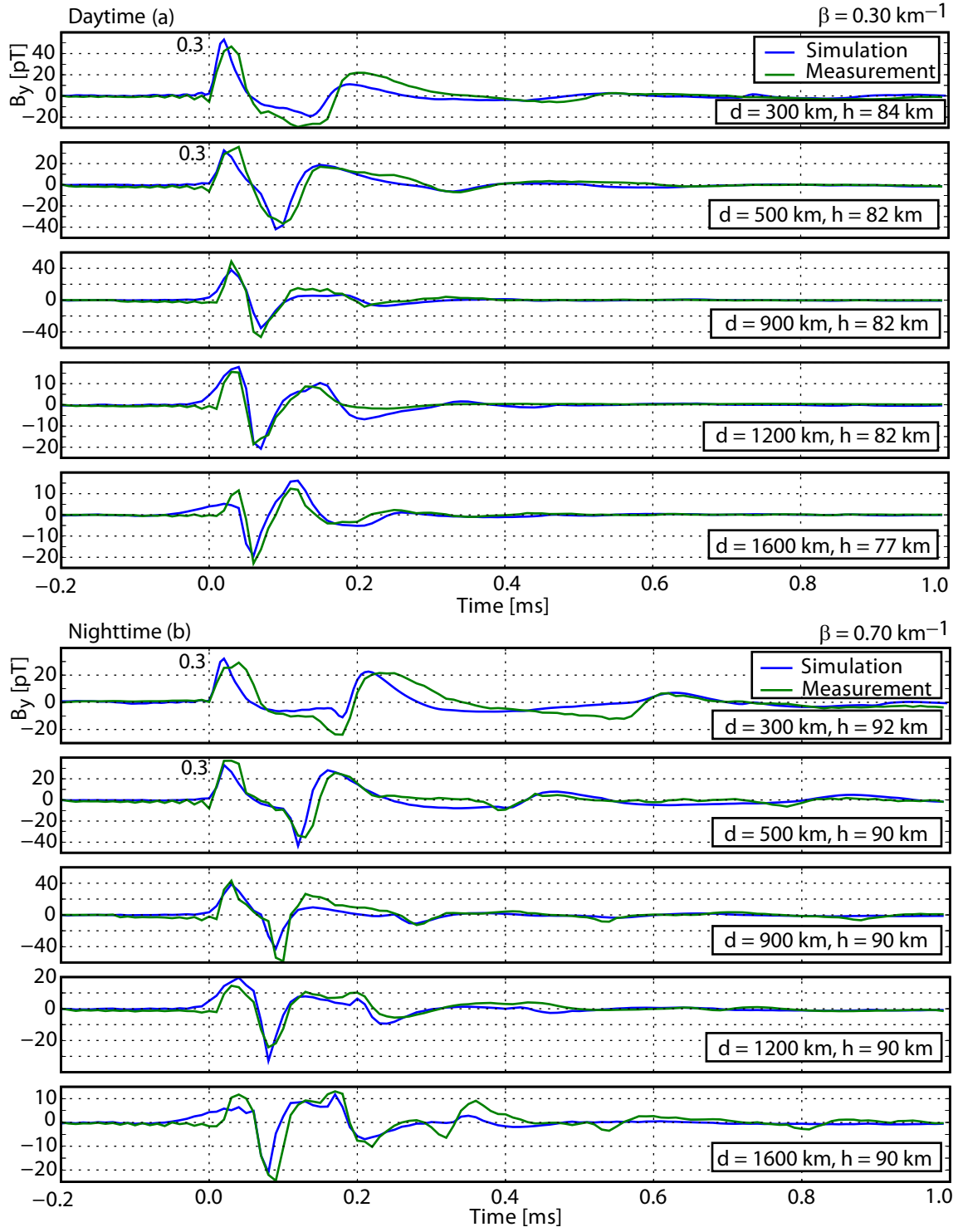
Figure 4.13: Simulated versus measured waveforms for (a) daytime and (b) nighttime propagation paths at five distances.
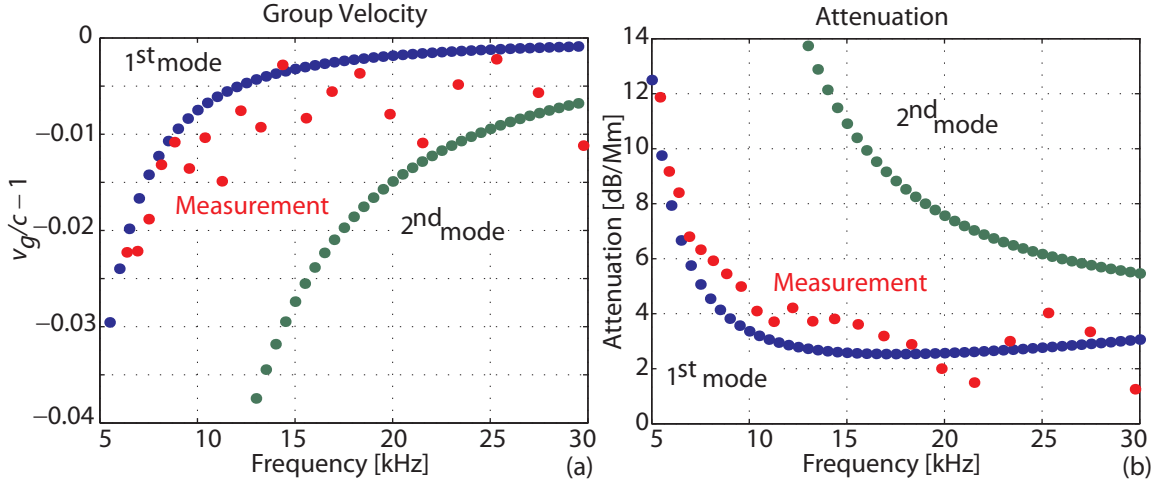
Figure 4.14: Daytime measurement and numerical calculation for the first two modes of (a) group velocity and (b) attenuation in the earth-ionosphere waveguide.

propagation at these distances is dominated by the characteristics of this first mode.

A full modal model which calculates the excitation factors has been applied to the ELF portion of a sferic waveform [*Cummer and Inan*, 2000] and the VLF portion [*Cummer et al.*, 1998] using the Long-Wavelength Propagation Capability (LWPC) code [*Ferguson*, 1998]. The latter paper by Cummer focused on nighttime propagation using a long time duration (∼20 ms) to investigate the average ionospheric profile between the source and receiver. Application of this technique to the early-portion of the sferic using a typical ionospheric profile is left as future work (see Section 6.2).

4.2.4   Comparison to FDTD simulation

Finally, a finite-difference time-domain (FDTD) simulation [*Chevalier and Inan*, 2006] was used to derive the daytime waveform bank, using the same source term as above, out to 1000 km. An FDTD simulation was run over a 1000 km path using a daytime ionosphere to obtain an impulse response of the channel at varying distances along the path. The results in Figure 4.15 were obtained after a convolution of this impulse response with the source radiation term, with a direct comparison in the lower subplot to a measured average sferic. The FDTD code used for this simulation assumed a flat
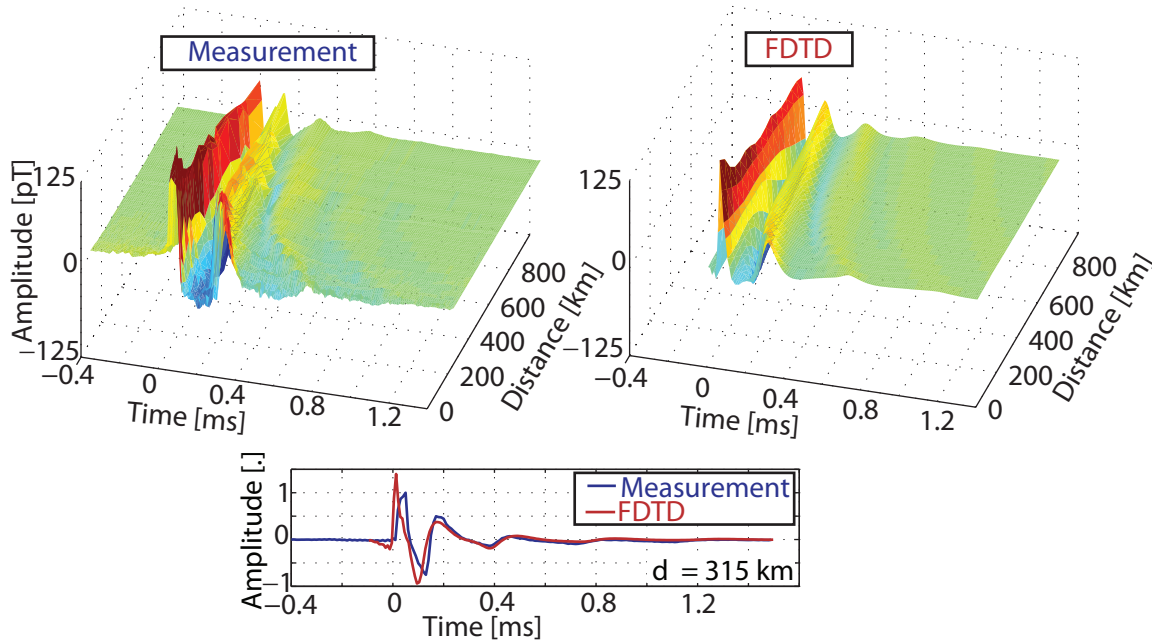
Figure 4.15: Comparison between measured and FDTD-modeled waveform bank for a daytime ionosphere out to 1000 km.

earth and a perfectly conducting ground. As is evident from the overlapping time-domain waveforms in the bottom plot of Figure 4.15, the ground wave is insufficiently attenuated as a result of this simplified profile. However, at least out to the end of the simulation space, the subsequent ionospheric reflections are well-reproduced by the FDTD run.

This section has shown that the dominant features of the waveform bank may be readily reproduced through relatively simple modeling techniques. Both the empirical and modeled banks capture a large percentage of expected waveform behavior through an ordered list of waveforms, indexed by distance and grouped according to the ionospheric profile along the path. In the next chapter, this insight is used to develop a new long-range lightning detection and geo-location network.

Chapter 5

Lightning Geo-Location

Armed with the observation that one can define a canonical waveform with reasonably few parameters, the next section seeks to create an algorithm for reliably determining the arrival time of an incoming sferic, and, by extension, to arrive at a new geo-location algorithm.

5.1 Arrival Time Determination

The arrival time determination follows a two-step procedure. First, each sferic is compared against a bank of locally stored canonical waveforms, in order to estimate the polarity and the approximate distance to the stroke. Second, the low-variance timing features are isolated and sent back to the central processor for comparison with other receiver stations. These two steps are described in the following two subsections.

5.1.1 Cross-Correlation Step

The initial cross-correlation step proceeds as follows:

1. Given data from a two- or three-channel receiver, the first step is to clean each channel according to the procedure outlined in Chapter 3 to minimize noise from local powerline hum and narrowband VLF transmitters, maximizing the

SNR between the impulsive broadband sferics and all other electromagnetic radiation in the recorded bandwidth.

2. The next step is to identify the sferics in the broadband data. Without prior knowledge of the propagation distances and to maximize the SNR for the distant and therefore weaker signals, each channel is filtered in order to isolate the frequency range with the greatest contribution from distant sferics. For large propagation distances, the sferic energy is concentrated near 9–13 kHz. Therefore a filter between 5–15 kHz is applied to maximize the received SNR for these distant and weak signals. Then, a composite magnitude signal from the two magnetic field channels is formed:

$$m(t) = \sqrt{[B_{\mathrm{NS}}(t)]^2 + [B_{\mathrm{EW}}(t)]^2} \tag{5.1}$$

where $B_{\mathrm{NS}}(t)$ and $B_{\mathrm{EW}}(t)$ denote the sampled and filtered time-domain signals from the North-South and East-West channels, respectively. The occurrence of a received sferic is flagged when the magnitude value rises above a given (pre-determined) threshold. The specific threshold value is in turn governed by the local noise floor and the desired sensitivity of the receiver for a given lightning peak field value, which is in turn a function of the lightning stroke's peak current and the path distance and profile. A threshold detection algorithm is more easily implemented if $m(t)$ is first passed through a lowpass filter, with cutoff frequency $f_c \simeq 3$ kHz.

3. Once the occurrence of a sferic has been identified, a segment of data containing the sferic from each channel is isolated. Most waveforms from daytime paths and long-range nighttime paths have a practical waveform length of less than 1 ms, as seen in Figure 4.8. Nearby lightning with nighttime propagation paths may have many milliseconds of enduring energy near the modal cutoff frequencies, but for the purpose of a speed-of-light referenced arrival time, the initial portion of the received waveform is the only region of interest. Thus, a window is chosen starting 0.2 ms before and 1.0 ms after the threshold crossing point.

4. The arrival azimuth from the windowed data segment is determined next. To minimize polarization errors, the arrival azimuth is determined using data from the leading part of the sferic. If an E-field channel is available in addition to the two magnetic channels, the arrival azimuth can be determined uniquely. Otherwise, there is a 180° ambiguity. This azimuth is then used to digitally rotate the magnetic field data so that one channel is along the measured arrival azimuth. This rotation forms the time domain signal $r(t)$ of the received sferic.

5. Based on this measured arrival azimuth, a locally stored bank of canonical waveforms is then retrieved. If there is a 180° azimuth ambiguity, two waveform banks need to be loaded–one for each possible direction. The waveform bank consists of a series of canonical sferic waveforms at discrete distances, such as those found in Figure 4.8. The waveform bank selection depends on the projected day/night path at each distance, which is a function of the local time of day and arrival azimuth. The local receiver stores a daytime, nighttime, and ideally a series of mixed day/night banks that can be loaded into memory as needed.

6. The received sferic is then compared with each waveform in the canonical waveform bank. The comparison between the received sferic and each entry in the waveform bank involves:

   (a) Normalizing the received waveform and the bank entry by the Euclidean norm, thereby removing any bias based on the as yet unknown peak current-dependent amplitude.

   (b) Performing a cross-correlation type procedure between the normalized received sferic and the canonical waveform, whereby the measured signal is delayed by discrete amounts and compared to the reference waveform. This procedure is only the first step in determining an arrival time, so its precise nature is less important. One possible method is to iteratively cross-correlate the two waveforms and search for the maximum correlation coefficient, as is done in *Lee* [1986]. Another option is to minimize
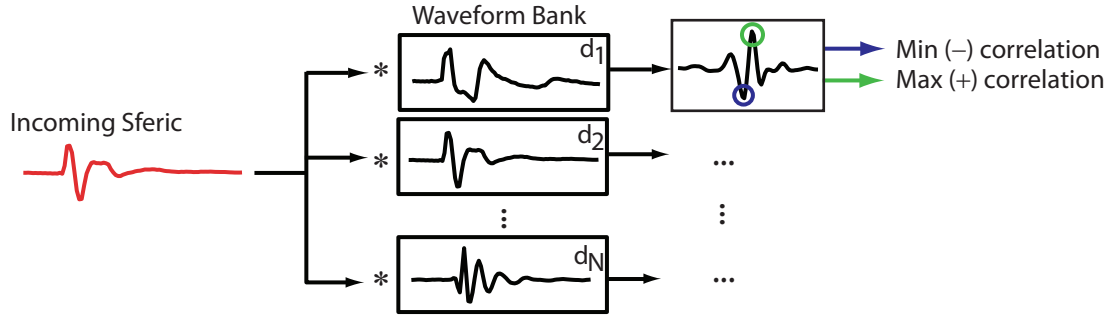
Figure 5.1: Illustration of a sferic cross-correlation with a waveform bank.

a specific (say, L^2) norm distance between the two waveforms by varying the time delay of one of them, employing an iterative procedure to achieve subsample accuracy.

(c) Repeat (6b) with the negative of the received waveform. The waveform bank entries are all normalized to a specific polarity (in this thesis, a negative cloud-to-ground stroke), so the peak of the cross-correlation with each polarity yields a different time offset. This cross-correlation procedure, which is repeated for both polarities with each waveform bank entry, is illustrated in Figure 5.1.

7. At this step the 'best' correlation across all entries of the waveform bank for both the positive (step 6b) and negative (step 6c) is chosen. If a cross-correlation method is used, the 'best' results are the maximum across all bank entries; if the norm distance is used, the best results are the minimum across all distances.

8. Thus for both the positive and negative assumed polarity, there is a particular optimum result based on the comparison with each entry in the waveform bank. Ideally, one polarity yields a much better match than the other. In this case, this sensor's polarity estimation is identified by the polarity that yields this better match and a range estimation is approximated by the indexed distance of the waveform bank entry used in this match. At the central processor, this polarity estimation will be compared to the same estimation made by other sensors to make a final polarity determination, as described in more detail in Section 5.2.1.

When both the negative and positive polarities yield approximately the same best correlation value, then this sensor's polarity estimation is ambiguous. For either case, both polarity results are sent back to the central processor and the final polarity estimation is determined after incorporating all of the available sensor data for that event.

Figure 5.2a illustrates this procedure. The sferic in question originated from a -16.7 kA strike 660 km from the receiver. The recorded and rotated sferic is drawn in red in the bottom graph. The top graph plots $R(\delta)$ from equation (1.8) from cross-correlating the received sferic with every waveform in the canonical waveform bank. The dotted blue $(-)$ correlation line corresponds to the correlation result with the received waveform set to the correct polarity (in this case, negative, but operationally this is not known a priori); the green $(+)$ correlation line corresponds to the result from correlating the wrong polarity with the waveform bank. The distribution of distances represented in the waveform bank is set to be approximately logarithmic, as this best captures the rate of change of the features, as was shown in Figure 4.8. The red line indicates the true distance of the strike (660 km), and the red square, obtained via quadratic interpolation from the neighboring points, indicates the inferred distance (676 km). The red '$\times$' indicates the inferred distance based on the incorrect polarity correlation result. In this case, with a short propagation distance, the correct polarity yields a much better correlation compared to the incorrect polarity.

The second subplot in Figure 5.2a plots the normalized measured sferic with the chosen entry of the waveform bank. The abscissa of the measured sferic waveform (red curve) is referenced such that the extra offset needed to achieve the maximum cross-correlation is zero. The correct polarity peak correlation waveform from the waveform bank, indicated in blue, closely resembles the received waveform. The offset of this cross-correlation procedure used to align the measured sferic with the reference waveform immediately provides a first estimate for the arrival time referenced to speed-of-light propagation: the time t in the received sferic waveform which aligns with the d/c line represents this estimate.

Figure 5.2b repeats this procedure for the same sferic at a receiver 4510 km from the source. Due to the more slowly changing waveform at larger distances (which

(a) Waveform bank cross-correlation with a nearby lightning strike.

(b) Waveform bank cross-correlation with a distant lightning strike.
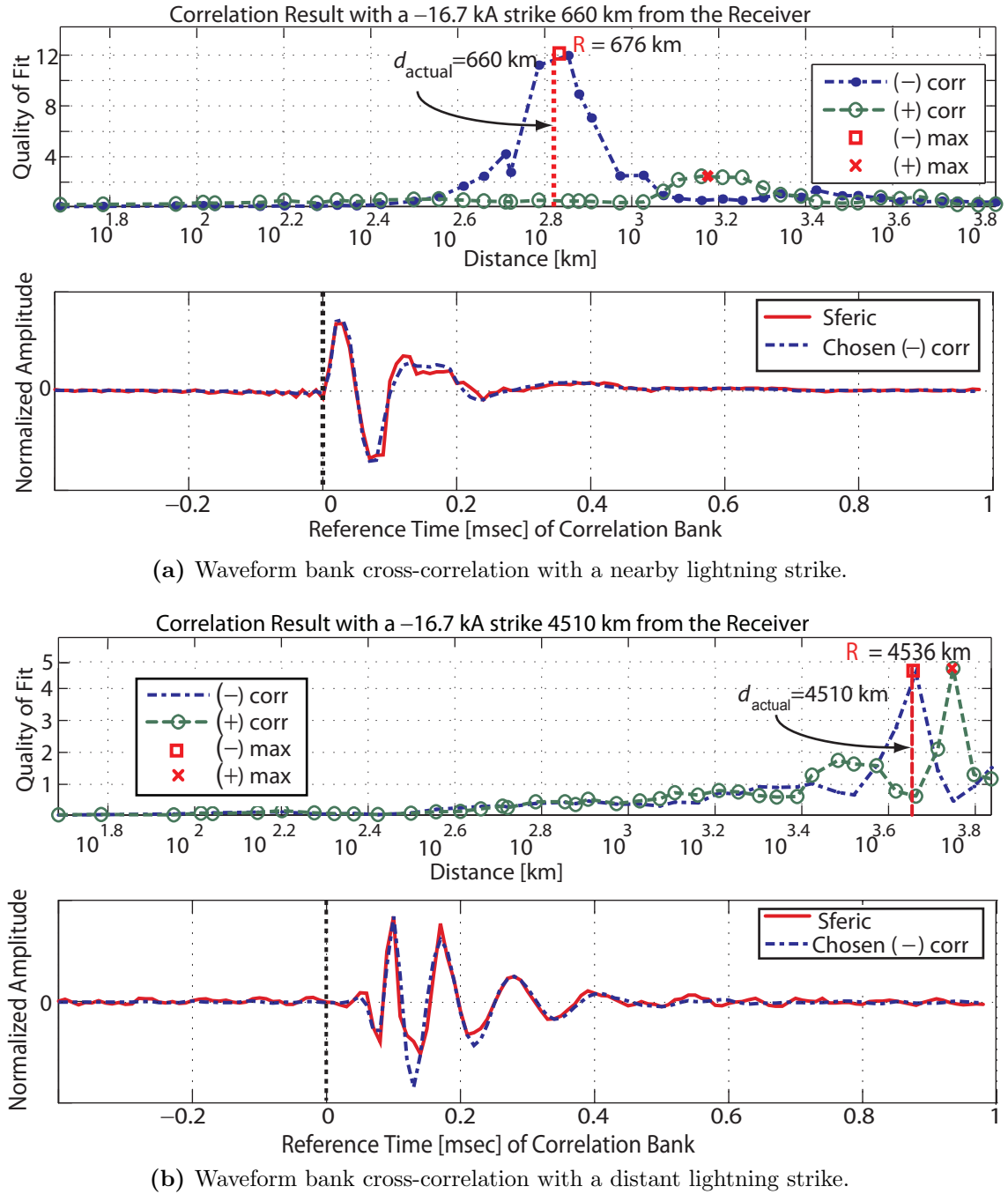
Figure 5.2: Illustration of the cross-correlation procedure with a waveform bank for two example source-receiver distances: (a) 660 km; (b) 4510 km. The top panel of (a) and (b) shows the mean power SNR (1.8) resulting from the cross-correlation of each waveform bank entry with the sample sferic and its negative. The bottom panel overlaps the sample sferic and the waveform bank entry corresponding to the distance with the highest cross-correlation result.

again can be seen from Figure 4.8), there is greater ambiguity between the correct and incorrect polarity. Indeed, in this case the incorrect polarity yielded a higher quality of fit. Both correlation offsets are transmitted back to the central processor. During the geo-location procedure, described below, meta-data from this sferic is to be correlated with sferics from other stations. For a given lightning strike, the inferred polarities must match from each site. For three stations, if just one site is close enough to make a definite polarity determination, then the other two sites may pick their polarity accordingly. If all three stations are distant enough to yield ambiguous polarity estimates, then a fourth station may be used to resolve the ambiguity. In this latter case, the triangulation procedure is performed using the four stations for both polarities; since there is a redundant timing measurement, the polarity which yields the lower χ^2 estimate is the chosen solution. If sferic data from only three stations is available, the weighted accuracy of the range estimates associated with each polarity can also be used, with nearby sensor results weighted more heavily. This procedure is described in more detail below.

Thus, if the incorrect polarity is chosen, this issue can be resolved at the central processor under many circumstances. If, however, the peak correlation for the correct polarity is off by a full cycle ($\sim$60 μs), then the arrival time refinement in the next section is also off by roughly this amount and introduces an offset of $(c)(\sim$60 μs$) =$18 km in the triangulated strike position. When such an offset occurs, a fourth station is needed to provide the redundant measurement to flag a mismatch with a high χ^2 value. This mismatch is the cost for a phase coherent detection scheme. A frequency-domain approach where the whole sferic is used for timing, such as the TOGA measurement as used in the WWLLN network, does not jump to a different arrival time. Such a measurement will not, however, take advantage of the lower timing spread toward the beginning of the waveform and must contend with multi-mode propagation at close range, issues that the arrival time approach presented in this thesis deals with naturally. If a 60 μs jump is detected or suspected based on the SNR and propagation distance, it may be advantageous to instead use something like an averaged group delay with a separate set of correction grids, or to average the positive and negative correlation offsets using the correlation weighting to help reduce the consistent error
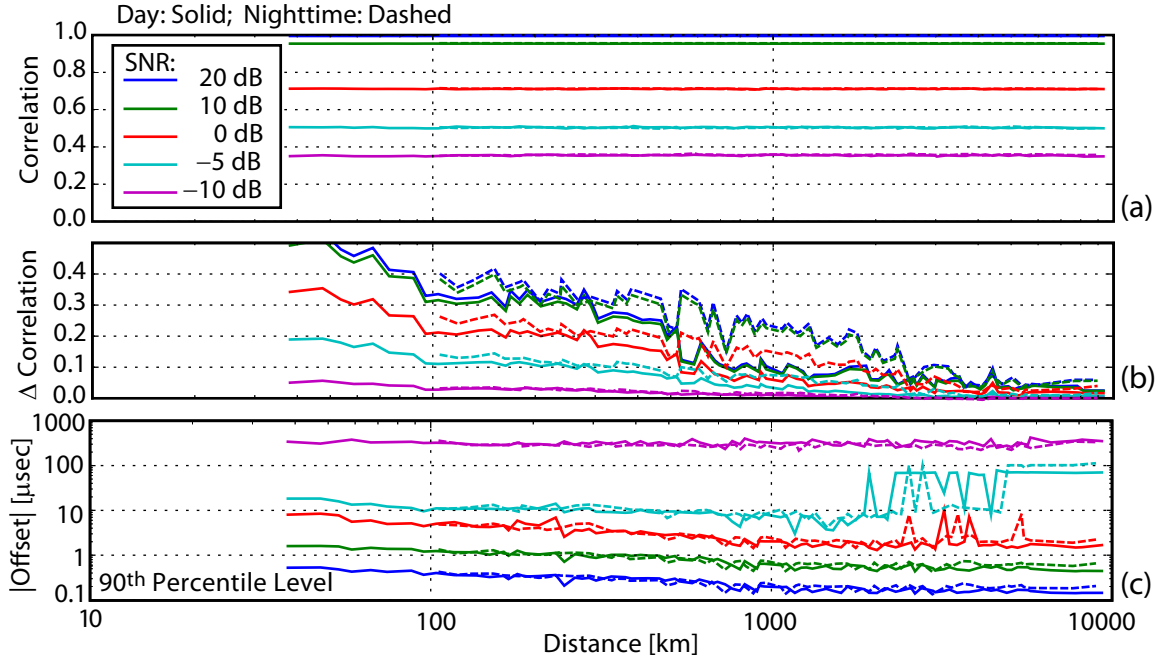
Figure 5.3: Simulated cross-correlation results for varying SNR levels: (a) normalized peak cross-correlation; (b) difference between the normalized peak cross-correlation of the correct and incorrect polarity; (c) resulting offset from the maximum cross-correlation. Results from the daytime (nighttime) bank are shown in solid (dashed) lines.

offset.

Figure 5.3 illustrates this problem in more detail. For both the daytime and nighttime measured waveform banks of Figure 4.8, each waveform bank entry is cross-correlated with the entire waveform bank of the same ionospheric profile with varying levels of SNR. The first subplot (a) shows the maximum correlation value with the correct polarity, where the correlation is on a linear scale from 0 to 1, with 1 representing a perfect match. As expected, as the SNR decreases, the correlation correspondingly decreases. The second subplot (b) shows the peak correlation difference between the correct and incorrect polarities. Two trends are evident from this plot. First, the correlation difference is highest for the shorter range sferics which have a higher modal content and so a more distinct feature set. This trend parallels our observation from Figure 5.2. It is also notable that the correlation difference is higher

for the nighttime waveform bank where the modal attenuation is lower. Second, as the SNR decreases, the difference between the positive and negative peak correlation values also decreases, lowering our ability to differentiate between the positive and negative polarity at any distance. Finally, Figure 5.3c shows the 90th percentile of the magnitude of the correlation delay offset. For SNR levels including 0 dB and above, 90% of the correlation offsets are below 10 μs, indicating that the correct offset was usually chosen. At −5 dB, beyond ∼2000 km, the 90th percentile level jumps to the full-cycle offset level, which is ∼70 μs for the daytime waveform bank and ∼100 μs for the nighttime waveform bank (the increased separation is clearly seen in Figure 4.8). At an SNR of −10 dB, the offsets are uniformly greater than 100 μs, indicating a total breakdown of this technique. An operational SNR threshold of ∼0 dB sets a minimum detection threshold based on the local and site noise at that receiver (see Appendix B), which in turn sets a maximum detection range for a given peak amplitude. In this context, application of the lightning geo-location techniques described in this dissertation requires the utilization of sensitive receivers with high dynamic range deployed at sites with low (or removable) local electromagnetic interference (EMI).

Figure 5.4 shows a scatter plot with the 16th and 84th percentile levels of the inferred range estimate for the correct polarity from this cross-correlation procedure. These results suggest an overall range estimation accuracy of ∼20%. In the results that follow, the range uncertainty is therefore set at $\sigma_R = 0.2R$ in (1.4).

There are many other methods that have been used to estimate the source-receiver range. While our method gives a range estimate for all distances and ionospheric profiles, each of the previously published methods works for only a subset of strike types and propagation distances. *Sao and Jindoh* [1974] use a formula introduced in *Wait* [1970, p. 314] which estimates the correlation between the delay between the impulsive VLF sferic and the lower-frequency slow tail and the propagation distance to estimate the range. This technique will only work, of course, on the subset of strokes which generate a measurable ELF response. *Gopalakrishna et al.* [1984] introduces a method which measures the relative group delay and spectral content between multiple frequencies, but this technique assumes single-mode propagation and so is
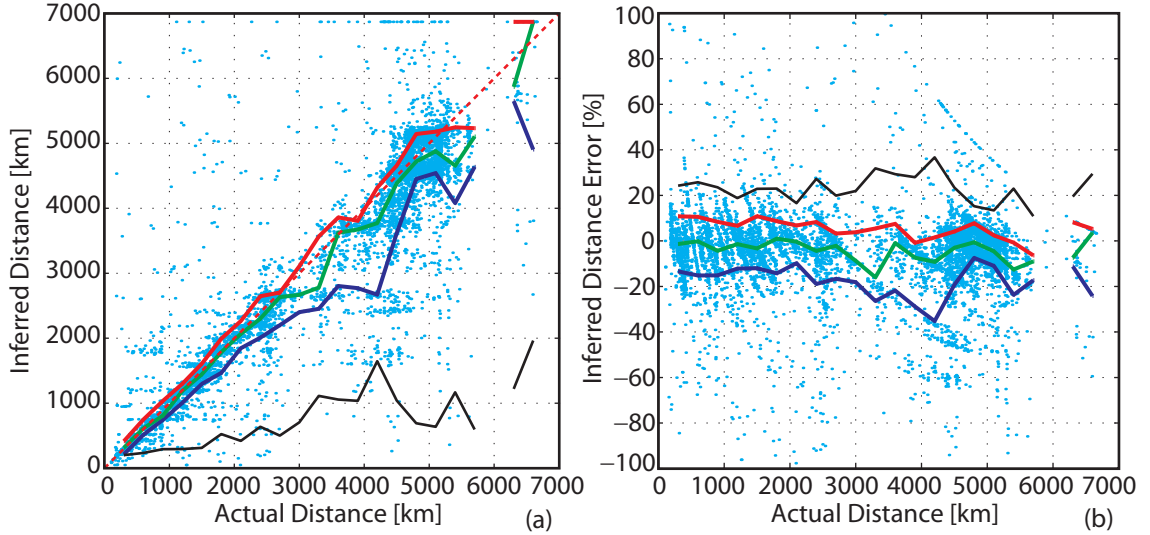
Figure 5.4: Daytime range estimation error resulting from the cross-correlation procedure.

ill-suited for short ($\lesssim$1000 km daytime, $\lesssim$2000 km nighttime) propagation paths. *Rafalsky et al.* [1995] uses a technique that attempts to isolate the dispersion profile of the TE_1 mode to estimate distance. This method yielded range estimates of $\lesssim$7%, but relies on an electric field measurement in addition to two magnetic field channels and an appreciable amount of energy in the TE_1 mode, and was only evaluated out to 3500 km.

At the heart of the location accuracy of the system are the arrival time estimates. An arrival time estimate based on the correlation offset alone is not optimal. The canonical waveform database may be derived from measured averages of waveforms which may present a smoothing of the waveforms (as a result of taking the median at each point in the waveform from hundreds of sample sferics) and a skew based on the actual propagation paths used to build the waveform bank. Also, in the previous chapter it was shown that the range of expected waveforms increases over the duration of the sferic. Finally, the waveform bank only has entries at discrete distances, so that a sferic originating from an intermediate distance 'snaps' to the nearest waveform, thereby causing the arrival time correction beyond the d/c line to assume discrete values that may not represent the correct offset for the intermediate distance. The
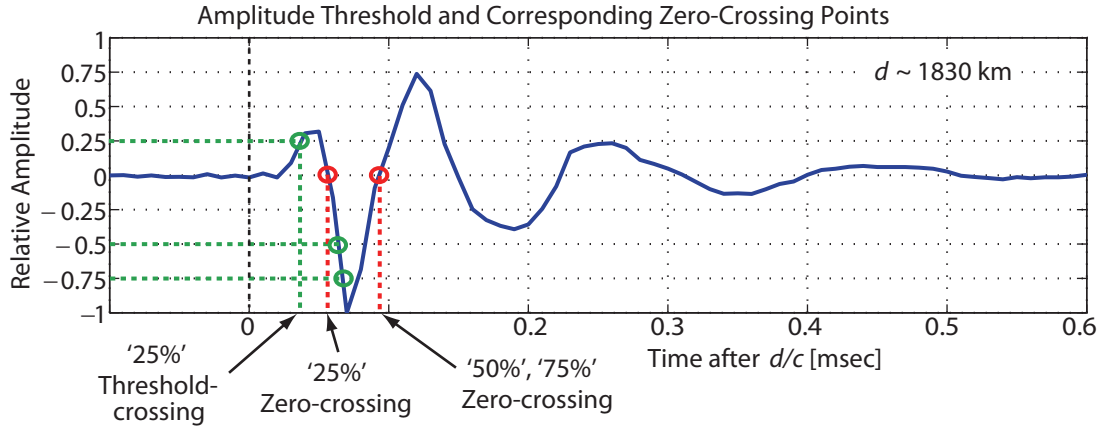
Figure 5.5: Sample sferic with highlighted threshold and zero-crossing features.

reduced spread of the waveform near the onset of the sferic may be partly exploited by introducing a second correlation pass that windows toward the start of canonical waveform, but this method still does not address the problem of the discrete and possibly empirical nature of the waveform bank. The next section shows how the low variance features of the sferic are used to further refine our timing estimate at the central processor.

5.1.2 Timing refinement

The cross-correlation sets the arrival time using the entire waveform. Currently, the minimum spread towards the beginning of the waveform is utilized by refining the arrival time using either the 50% threshold-crossing point on the leading edge of the ground wave, for strikes closer than a pre-defined distance d_0, or an early zero-crossing point for strikes at distances greater than d_0. The distance d_0 is chosen so that most of the sferics out to d_0 have a ground wave amplitude at least 50% large as the peak of the whole waveform, ensuring that the threshold measurement is reliably triggered by the ground wave and not by the first ionospheric reflection. For most conductivity paths, d_0 may be set between 800–1000 km. Figure 5.5 illustrates the relevant waveform features. The time delays of the threshold crossing and the zero-crossing points are linearly interpreted to achieve sub-sample accuracy.

For each waveform bank entry, the first zero-crossing after the 25% threshold

crossing point is identified and used in the timing refinement procedure described below. Figure 5.6b plots the delay after the d/c line of these zero-crossing points for the daytime waveform bank. Also shown are the zero-crossing points after the 50% and 75% threshold crossing (some of which map to the same zero-crossing point, as is the case in Figure 5.5) to aid with the regression curves also plotted here and described below. Figure 5.6a plots the 50% threshold-crossing point for the same waveform bank. As seen here, the threshold points corresponding to the ground wave (those points whose delay is $\lesssim 50$ μs) delay in time as the ground wave attenuates with distance. While the amplitude threshold points slowly delay with distance until they merge with the next feature, the zero-crossing points maintain more discrete levels, where each level is fitted here with a second-order polynomial. The variation in the delay points is due to variation in the propagation paths and slight biases in the source waveforms for the specific collection of sferics used to generate each point.

Using the zero-crossing points in the waveform bank entries, an early zero-crossing time of a measured sferic at each receiver is determined in the following way. After the cross-correlation step described in the previous section, the measured sferic is time-aligned to the chosen entry of the waveform bank with an offset determined by the peak in the cross-correlation. The '25%' zero-crossing is then identified as that zero-crossing which is nearest to the corresponding pre-computed 25% zero-crossing in the chosen waveform bank entry. Since the polarity of the slope of the appropriate zero-crossing must also match (for the 'negative' cross-correlation, the polarity of the zero-crossings of the measured sferic are reversed), the cross-correlation need only provide a rough timing estimate for the measured sferic: the polarity of the zero-crossings in the waveform bank flips between one discrete level and the next, providing a roughly 60 μs buffer between one zero-crossing level and the next.

Besides providing a range estimation for each possible polarity, the cross-correlation step is therefore used only to identify the appropriate zero-crossing in the measured sferic. Along with the 50% threshold timing value, which is used as the timing reference at the central processor if the sferic is determined to be closer than d_0 from the receiver, the absolute, GPS-referenced time of this zero-crossing is sent back to the central processor. These timing estimates need to be referenced back to the d/c
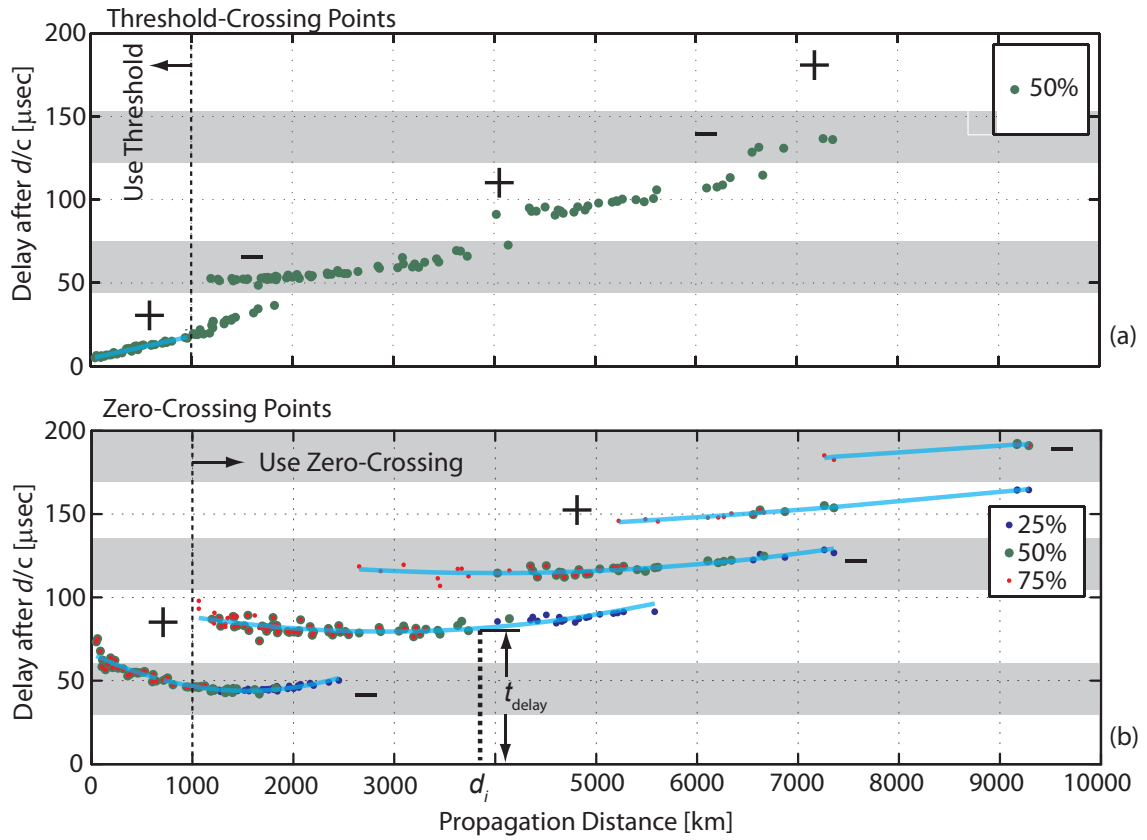
Figure 5.6: (a) Threshold and (b) zero-crossing delay values, referenced to the speed of light propagation time, for the measured daytime waveform bank. In this figure, d_0 is set at 1000 km.

line to recover our speed-of-light assumption in the geo-location algorithm. The central processor computes smooth regression curves for the threshold point and the zero-crossing clusters, as illustrated in Figure 5.6. The threshold crossing point is unambiguous since it is only used where the ground wave is clearly visible. However, the zero-crossing point may be on any one of the curves in the second subplot of Figure 5.6. To resolve this ambiguity, each receiver also sends back an index corresponding to the level of the zero-crossing from the chosen waveform bank entry to identify which curve should be used when compensating for the extra delay after the d/c line. At distances $\lesssim$10,000 km, only five curves are necessary to capture all of the zero-crossing levels; hence, only a 3-bit index is needed for each sferic and each polarity.

This procedure for finding the zero-crossing time is repeated for both polarities. Along with the single threshold-crossing timing point, these two zero-crossing points are sent back to the central processor to be used in the algorithm described in the next section. To save on the number of bits per sferic needed to send back to the central processor, a full time stamp for the threshold crossing may be used with a differential time measurement between the threshold and the two zero-crossing points.

While the choice of the threshold and zero-crossing levels are somewhat arbitrary, as indicated above they are motivated by the features in the waveform bank. The 50% threshold is chosen to allow sufficient SNR while ensuring that the ground wave is correctly identified out to d_0, and the zero crossing is chosen after the 25% threshold point to take full advantage of the lower-variance nature of the beginning portion of the waveform.

Using the zero-crossing values is qualitatively similar to directly cross-correlating sferics in an ATD network, except that this calculation effectively weights the cross-correlation to the initial part of the sferic and using a local waveform bank instead of a waveform measured at another site. Both methods enjoy the extra accuracy afforded by a phase-coherent timing measurement, but our method extends this ability to arbitrary differential sferic paths, and reduced the required communications bandwidth.

5.2 Geo-Location

5.2.1 Location Estimate

The central processor must take a continual stream of arrival time estimates from each station, correlate the appropriate sferics, and find location estimates using (1.6). This section first focuses on the determination of the speed-of-light arrival times, based on the correct identification of the same recorded sferic from two or more sites (three sites are necessary if the location is to be primarily determined by the more accurate arrival time estimates), and using the threshold- and zero-crossing times sent back from each receiver.

Suppose that the distance between the strike and a receiver is known by the central processor, and then first consider the case where the distance is less than d_0. In this scenario the central processor would take the 50% threshold-crossing time and subtract a pre-determined distance-dependent delay value which corrects for an attenuated and dispersed ground wave. Figure 5.6a shows such a correction curve in cyan color, which is derived by fitting a linear curve to the measured waveform bank threshold crossing points out to d_0.

This correction curve is the same for daytime and nighttime conditions since it only needs to correct for ground wave propagation. While this correction curve was derived from the measured waveform bank, a similar curve may be derived from numerical modeling. This theoretical curve will vary based on the exact conductivity profile between the source and the receiver. Extra perturbations due to the this varying conductivity are corrected for through empirical correction matrices (see Section 5.2.3). One could have taken the speed-of-light arrival time to be the threshold crossing point and delegate all of the correction to this latitude- and longitude-indexed empirical correction matrix. Nonetheless, removing this linear distance-dependent delay shifts the arrival time so that the extra perturbations are small and centered around zero. This shift confers an obvious benefit: in the absence of any empirically-derived correction grids, the errors are relatively small, having already corrected for the dominant time offset. Also, if redundancy measurements are used to calibrate the correction grids the necessary additional correction factor is smaller, which leads

to a more expedient convergence to the correct value.

The case where the distance is greater than d_0 and the 50% threshold crossing point no longer provides a reliable time estimate is now considered. In this scenario, the zero-crossing point is used instead. It is also assumed for now that the correct polarity has been presumed. The zero-crossing group identifier identifies which cluster in Figure 5.6b to index. Each of these clusters has a second-order polynomial cyan curve fitted in a MMSE sense. We take this smooth curve as the 'average' time delay for the respective zero-crossing level and then make any further refinements using an emprically-derived correction grid (Section 5.2.3). The relevant regression curve is indexed at the actual strike-receiver distance, not the initial estimate given by the receiver. As an example, the correction to the zero-crossing value of a sferic measured 3800 km from the source is indicated in Figure 5.6b. In this case the specific waveform bank entry which yielded the highest correlation coefficient had a 25% zero crossing delay within cluster 2, i.e., this reference delay after d/c was in between 60 μs and 105 μs.

As mentioned above, the fitted curves have no particular significance beyond providing a correction term for the dominant delay beyond the d/c line. If the canonical database were derived from FDTD simulations, for example, the cluster of points around these curves would be much tighter than seen in Figure 5.5 and the polynomial fit may be slightly different. The correction grids depend on the specific polynomial fit and so it is imperative that only one curve is used for each cluster. If a different curve is used, the final correction grids must be re-calculated as described in Section 5.2.3.

Figure 5.7 shows the 25% zero-crossing points now for three different waveform banks, where the polarity is indicated by a '•' (positive slope) and a '×' (negative slope). Due to the higher reflection height, the delay points for the nighttime ionosphere are later than for the daytime ionosphere. Notably, the 50% day/night waveform bank yields delay curves that are roughly halfway between the all-day and all-night curves. These delay curves suggest that for long-distance propagation paths, in the absence of a profile with the specific day/night percentage, a linear interpolation may be used between the all-day and all-night zero-crossing profiles. It is
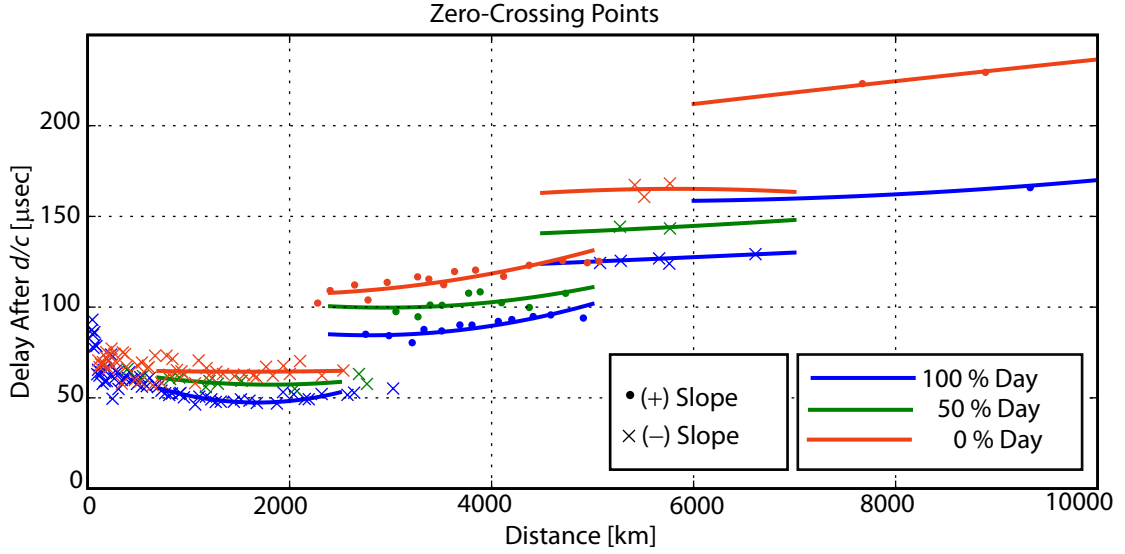
Figure 5.7: Zero-crossing points and polynomial regression curves for three waveform banks.

important to note, however, that the waveforms for the intermediate paths may not be simply derived from the all-day and all-night waveform banks, as is evident by the large difference between the waveform shapes in Figure 4.5. Results in this work were derived using only an all-day and all-night waveform bank; a detailed analysis of the benefits from this type of interpolation is outside the scope of this thesis.

The above procedure of subtracting the delay from the appropriate regression curve to recover the speed-of-light arrival time presumes that both the polarity and the propagation distance are known. Suppose instead that we have only correctly identified matching sferics from multiple stations and have yet to determine the correct polarity. Each station reports (i) a 50% threshold crossing time, (ii) a range estimate for both polarities, (iii) zero-crossing times for both polarities, and (iv) an arrival angle estimate. Using the threshold arrival time and a propagation speed $v \lesssim c$, (1.6) may be used to get an initial location estimate which is independent of the polarity. This initial estimate provides a better distance estimate d_i to each station, since the individual-sensor timing uncertainty (~ 60 μs) should result in $\lesssim 20$ km location error. With this estimate, we may now choose to use either the threshold crossing or the zero-crossing for the arrival time, making the corrections based on

the regression curves as described above. An initial polarity estimate may also be made at this point. Each polarity yields its own best correlation with the associated propagation distance. This intitial strike location allows us to determine the accuracy of the range estimate for each polarity. Evaluating the range contribution to the χ^2 in (1.6), with $\sigma_{R,i} \propto d_i$, for both polarities, we choose that polarity which yields the lowest cost. For strike distances greater than d_0, we now choose the zero-crossing time and associated correction curve corresponding to the chosen polarity.

With a chosen polarity and corrected arrival times referenced to speed-of-light propagation, we now calculate the location position again using (1.6) but with $v = c$. If there is a fourth station and the range estimate χ^2 was sufficiently high to cause concern, we can repeat the triangulation refinement with the opposite polarity and choose that polarity which yields an overall lower χ^2. If both polarities yield an unacceptably high χ^2, then it may be that (i) one or more of the sensor measurements corresponds to a different sferic and/or (ii) one or more of the sensors reported a zero-crossing time which has a full-cycle error, as was discussed in conjunction with Figure 5.3.

We now need to consider the manner in which we take continuous sferic reporting streams from multiple sensors and tease out which records to correlate from each sensor to geo-locate individual lightning strikes. This identification is a problem that can easily grow cumbersome. Suppose each station has a detection radius of 6000 km. Then a pair of stations may monitor a region of the size $\sim$6000 km. If we assume a nominal rate in this domain of 100 strokes/second, assuming a uniform temporal distribution of strokes each pair would have to correlate $\sim(100)(6000/c) = 2$ events from each sensor, which is 2^2 combinations. With n stations, the base changes from 2 to n and there are n^2 permutations to consider. If the stroke occurrence were more tightly clustered, causing the exponent to rise above 2, the number of possible combinations rises even more sharply. A cumbersome approach is to try all permutations and pick the configuration which yields the lowest overall χ^2. Having range, azimuth, and amplitude (discussed in the next subsection) estimates provide valuable redundancy to evaluate the merit of one combination over the rest, especially in the critical case where only three stations are available (the azimuth estimate only allows

for crude two-station geo-location). However, if this "crude" auxiliary information is used beforehand, computation time can be saved by limiting the number of station combinations one needs to consider.

Assuming each range estimate is accurate to within 20%, we can constrict the absolute time-coincidence window of each sferic to

$$\text{stroke time window } = t_{\text{thr}} - (1 \pm 0.2)R_{\text{p,n}}/v \tag{5.2}$$

where t_{thr} is the GPS-synchronized time of the threshold crossing point, $v \lesssim c$ is the approximate speed matched to the threshold crossing points, and $R_{\text{p,n}}$ is the range estimate of either the positive or negative polarity. Given the range estimation (for each polarity) of a sferic, (5.2) gives the range of possible times the causative stroke may have occurred. The search interval may be expanded by softening the range accuracy; the results used in this thesis employed a range estimation factor of 0.9 to capture many possible combinations at this initial step.

In the geo-location algorithm described in the next paragraph, a sferic reported by a particular station, referred to here as a reference sferic from a reference station, is used as a starting point, and then other candidate sferics from different stations are considered and tested to see if they may correspond to the same causative stroke. The reference station gives an initial possible time window of the causative stroke through (5.2). This reference interval width is proportional to $R_{\text{p,n}}$ of the causative sferic, so it is best to use a reference station (reporting the reference sferic) which is closest to the stroke.

Given a reference sferic that maps to a window of possible stroke times using (5.2) for each polarity, a reduced set from each of the other stations is chosen based on which reported sferics map to somewhere in this time range, again using (5.2) for each sferic (and for each polarity). The arrival time difference between the reference sferic and any other candidate sferic may also not exceed the baseline travel time between the reference station and the station reporting this candidate sferic. This set may therefore be reduced further by throwing out candidate sferics which violate

this baseline travel time[1]. From the resultant list of candidate strikes from each receiver (the lists generated using each polarity of the reference sferic are combined), clusters of three sferics at a time from three different stations, including one from the reference station, are used to map a strike location based on arrival time only. This procedure is used to generate the initial round of location estimates because it is an $O(1)$ calculation and therefore uses minimum computation time. Based on these results, preliminary χ^2 values are determined from the arrival azimuth contributions of each station. The three-station combination which yields the lowest χ^2 based on the azimuth information is then selected as the most likely candidate and the triangulation refinement proceeds as described above. If candidate sferics from stations other than the three contributing to this lowest χ^2 value are available, they may now be incorporated into the overall solution assuming the newly calculated χ^2 does not increase above a certain threshold.

The above procedure attempts to geo-locate strokes using at least three sensor measurements. Given a set of sensor thresholds, a subset of strokes will only trigger two stations. Should there be an interest, two-station pairings using the remaining sferics generated by these strokes may be pursued. A two-station geo-location must rely on the iterative procedure of minimizing the cost surface to obtain an initial estimate and is therefore much more computationally costly. Two-station combinations should only be pursued, therefore, if sufficient computational resources are available and the lower location accuracy associated with the results is acceptable. With fewer redundant measurements, two-station pairings also have a higher chance of miss-correlations, so there may be a higher rate of spurious events.

5.2.2   Peak Current Estimate

The peak current estimate follows an analogous procedure to the arrival time estimate. The peak amplitude value from each sferic is sent back to the central processor. After the strike position is determined, the measured amplitude is divided by a reference amplitude at the appropriate distance. If this reference amplitude is calibrated to

---

[1]The baseline time, representing the maximum possible arrival time difference between two sferics from the same stroke, is precomputed for each station pair and is kept in a lookup matrix.

refer to a 1 kA peak current, then the result will give the station's estimate of the peak current amplitude. The results from each station may be averaged to derive an overall peak current estimate.

The peak amplitude point corresponds to the peak of the ground wave and, beyond a certain distance, transitions to correspond to the peak of the first ionospheric reflection. Further still, the peak may correspond to a middle point in a dispersed wave packet representing a single propagating mode.

The NLDN calculates peak current assuming the transmission line model [*Cummins et al.*, 1998a], which predicts that the peak current will be linearly proportional to the peak in the radiated field [*Uman*, 2001, p. 333]. Recall that for the *Bruce and Golde* [1941] model the peak amplitude should correspond to the maximum of the derivative of the current-moment. However, for long-range sferic propagation, the peak measured field has been found to be highly correlated (correlation coefficients from 0.90–0.95) to the NLDN-reported peak current [*Wood*, 2004, p. 52]. In our network, we therefore assume that the peak current is linearly proportional to the peak of the measured magnitude of the magnetic field. The results in this work were calibrated against the peak current reported by the NLDN.

Figure 5.8 plots the normalized peak amplitude, multiplied by the inverse of the spherical spreading factor $(R_E \sin(d/R_E))^{-1/2}$, versus distance for the daytime waveform bank (thus each point is the result of averaging over hundreds of sferics within a small distance range). The flattened region at short distance likely represents a saturation at the receiver, though this phenomenon has not been investigated closely for this work[2]. To account for this effect, peak current estimates are weighted toward sferics with larger propagation distances. Beyond this point, the sferic amplitude exhibits a roughly exponential decay, consistent with single mode propagation (see Section 2.4.1).

---

[2]At close range, the amplitude dependence on distance may not be approximated using a single-mode propagation model as is the case for long paths. The ground wave attenuation is proportional to d^{-1} (and not $d^{-1/2}$), for example, and at an intermediate distance the peak amplitude feature transitions to the first ionospheric hop.
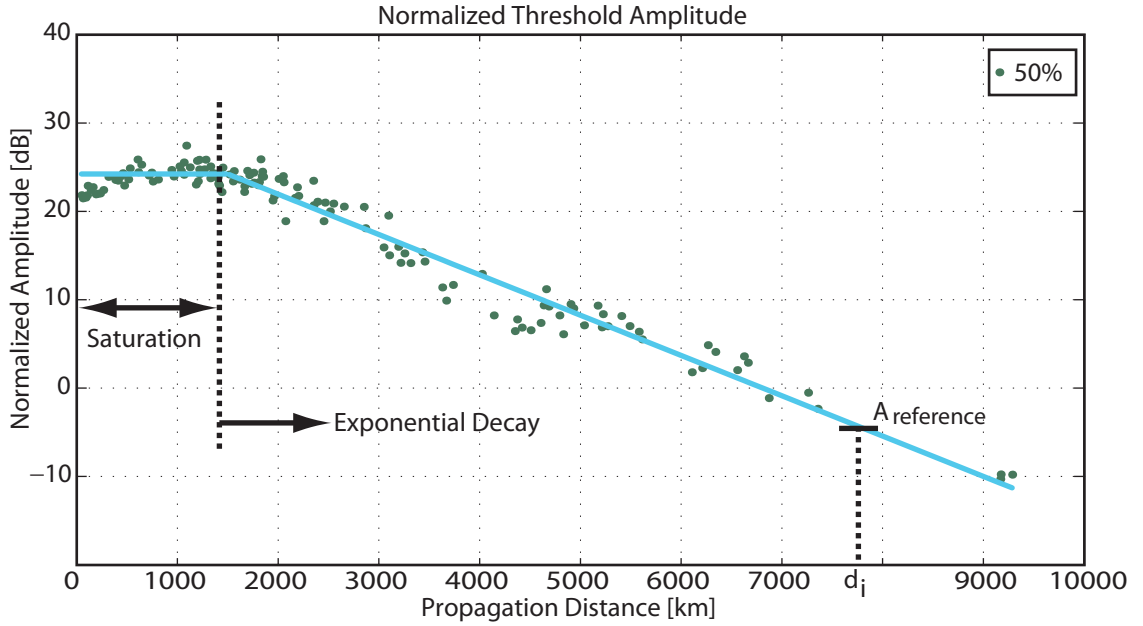
Figure 5.8: Peak amplitude versus distance from the daytime measured waveform bank.

5.2.3 Correction Grids

The above procedures for measuring the arrival time and sferic amplitude compensate for the dominant propagation factors that introduce offsets in the relevant measured quantities. Nonetheless, there are still consistent arrival time and amplitude offsets due to variations in the specific propagation path profile of each sferic for each receiver station. After the initial position is determined and the corresponding corrections are made, a final refinement is made at the central processor.

The final refinement is achieved by first constructing an empirical calibration map for both the delay and amplitude. If there is a reference network, this refinement may be done directly. In the absence of such a network, one must use either modeling or redundant measurements to construct this grid. In this work, each station has an associated correction grid discretized into $1° \times 1°$ pixels. After the regression curve calibrations were applied, the remaining delay and amplitude errors were collected in each bin. The median of the remaining error (the amplitude errors are given in dB) becomes the final correction for all lightning strikes originating in that pixel and is
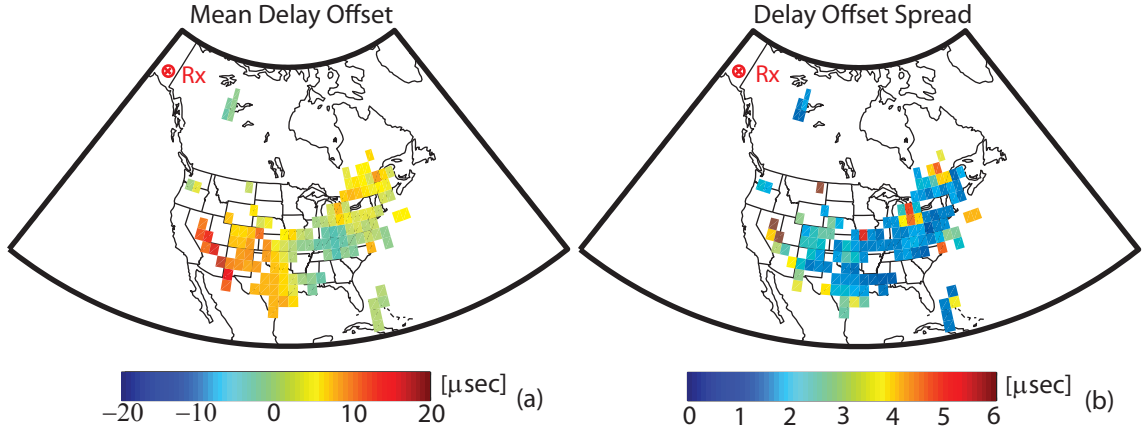
Figure 5.9: Final delay correction grid for Chistochina, Alaska, using NLDN data as a reference. (a) Mean Delay offset and (b) delay offset spread ($84^{\text{th}} - 16^{\text{th}}$ percentile difference) for each pixel.

subtracted before the final minimization step.

An example of a delay correction grid for propagation from a region in the US to an Alaskan receiver is shown in Figure 5.9. The median offset is seen to vary by up to 20 μs, reaching a maximum for paths propagating over the Rocky Mountains. By contrast, the delay spread (defined here as the difference between the 84th and 16th percentile of the delay offsets for a collection of strikes within a given pixel) for most regions is $\lesssim 3$ μs.

5.3   Results

This section analyzes geo-location results for September 6, 2007 using data from three stations: Juneau Alaska, Taylor Indiana, and Santa Cruz California. Synoptic data were recorded for one minute out of every fifteen at each station. Due to mismatched down times, synoptic data was available for the time periods 00–07 UT, 10–16 UT, and 20–24 UT, which covers several day/night profiles to each receiver.

The following sections compare the results to high-resolution stroke data from the NLDN. To allow for easy labeling, this section refers to this trial network as the

Stanford Lightning Detection Network[3] (SLDN).

As mentioned earlier, a loose range accuracy factor of 0.9 was used in the first step of the correlation of lightning strikes from different sensors. The following σ values were used in calculating χ^2 in (1.5)

$$\begin{aligned} \sigma_R &= \quad 0.2R \\ \sigma_t &= \quad 5 \ \mu s \\ \sigma_\theta &= \quad 3° \end{aligned}$$

To account for the increased azimuth uncertainty at close range seen in Figure 3.15a, σ_θ was increased linearly from 3° at 1000 km to 10° at 100 km.

Per-station limits on χ^2 were also used to minimize the number of spurious geo-location events. On average, the contribution from each station to the respective term in χ^2 was limited as follows:

$$\begin{aligned} \chi_R^2 &: \quad 2.0 \\ \chi_t^2 &: \quad 1.0 \\ \chi_\theta^2 &: \quad 1.0 \end{aligned}$$

The larger allowance for the range uncertainty deemphasizes the range estimate as a selection criterion; alternatively, σ_R may be increased. This relaxed constraint boosts the detection efficiency (defined below) for strokes which propagate over low-conductivity paths or strokes that lead to sferics that may for another reason produce sferics that do not yield an accurate cross-correlation result. With the above range, time, and azimuth uncertainty values, the location estimate is primarily determined by the arrival time measurements: due to the large baselines between the sensors, the minimization algorithm is free to minimize χ_t^2 with little penalty to χ_R^2 and χ_θ^2. Also, because there are only three contributing receivers in this test network, χ_t^2 is usually very small ($\ll 1$) since there is no redundant arrival time measurement. Finally, the average minimum linear correlation for each station to the waveform bank was set to 0.8 (out of a scale from 0 to 1).

The next two sections investigate the location accuracy and detection efficiency by comparing SLDN to NLDN stroke location results. For this comparison, individual

---

[3]Per the request of the author's parents, this network may also be referred to as the 'Said Lightning Detection Network'.
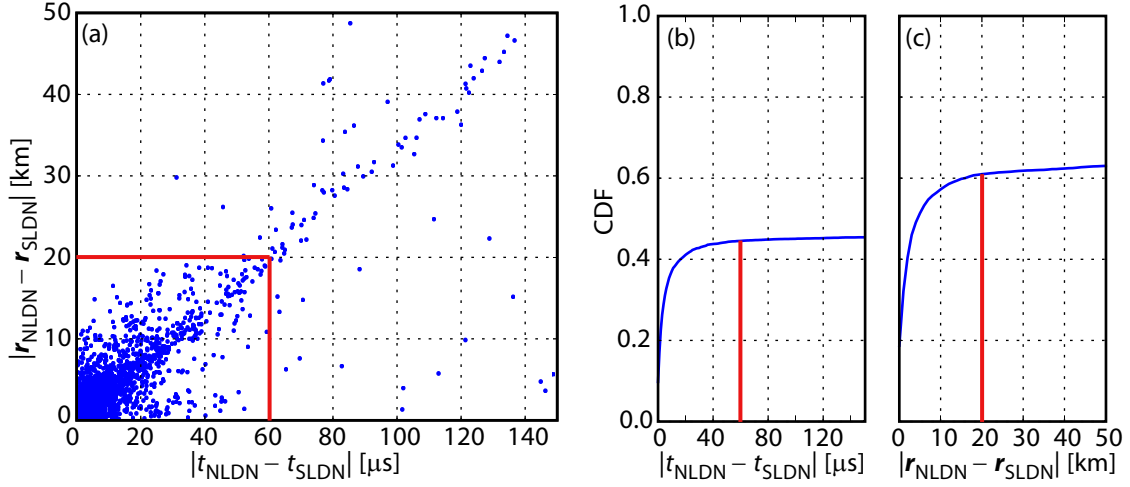
Figure 5.10: (a) Scatter plot of distance versus time error between SLDN and NLDN strokes. (b) CDF of the time error. (c) CDF of the distance error. The CDF's plotted ranges are truncated at 150 μs and 50 km in (b) and (c), respectively, hiding the long tails that slowly ramp up to 1.0. In all three plots, the spatial and temporal match criteria are indicated by the red lines.

events were matched between the two datasets if the spatial- and temporal-coincidence were within a certain tolerance. Figure 5.10a shows a scattergram of the distance and time error between the NLDN and the SLDN datasets for the 24-hour period, where for each NLDN stroke, the SLDN event with the closest time proximity (within 250 ms) was chosen as a possible match. Using the cumulative distribution functions (CDF's) of the time and distance errors in Figures 5.10b and c, the match criteria (indicated by the red lines) were set to be beyond the knee in both CDF curves to a simultaneous temporal and spatial coincidence of <60 μs and <20 km, respectively. The abscissa in each CDF plot was truncated to highlight the knee in the curve. The time outliers extend out to 250,000 μs and the distance outliers are only constrained by the size of the globe. The long tails not shown therefore capture the events reported by NLDN which have not time-correlated to SLDN events. As discussed further below, the CDF for the distance errors flattens out at a higher percentile value, indicating that many of the events that were not time-correlated do correlate to a nearby SLDN-reported stroke at some other time (perhaps a subsequent stroke in the same flash, for example).
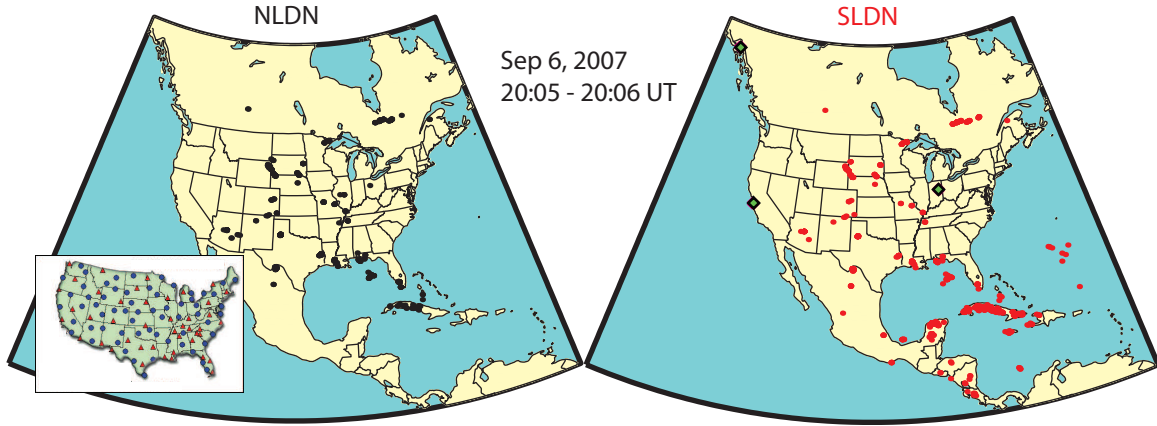
Figure 5.11: Visual comparison between NLDN and SLDN strike locations for a one-minute period. Inset adapted from *Cummins et al.* [1998b].

5.3.1 Location Accuracy

Figure 5.11 illustrates a preliminary qualitative check on the spatial coincidence of the two datasets. One minute of data is plotted from both NLDN (left subplot) and SLDN (right subplot). Most of the major storm clusters in NLDN are identified by the SLDN. The NLDN data clusters over the landmasses, with a few additional strikes occurring in the ocean. SLDN additionally detects many strikes beyond the NLDN network coverage. The inset of the NLDN map shows a map of the NLDN sensor network; this map is to be compared to the three sensors used in the SLDN, indicated by green diamonds.

Figure 5.12 plots a minute of stroke data from NLDN and SLDN from a time period two hours after that shown in Figure 5.11. The two data sets are now overlapped, with the NLDN data set plotted on top of and therefore partly obscuring the SLDN data set at this resolution. Two maps are shown: the top map shows strike locations from the two datasets that do not match; the bottom map shows the data sets which have a time- and spatial-match of $<60~\mu$s and <20 km. The Venn diagram summarizes the color coding in each plot. There are several unmatched strike locations outside the range of NLDN, as seen in the previous figure. Within the NLDN coverage area, the storm cluster locations match up with the exception of a few strike locations in eastern Colorado. The expanded window shows a region of storm clusters
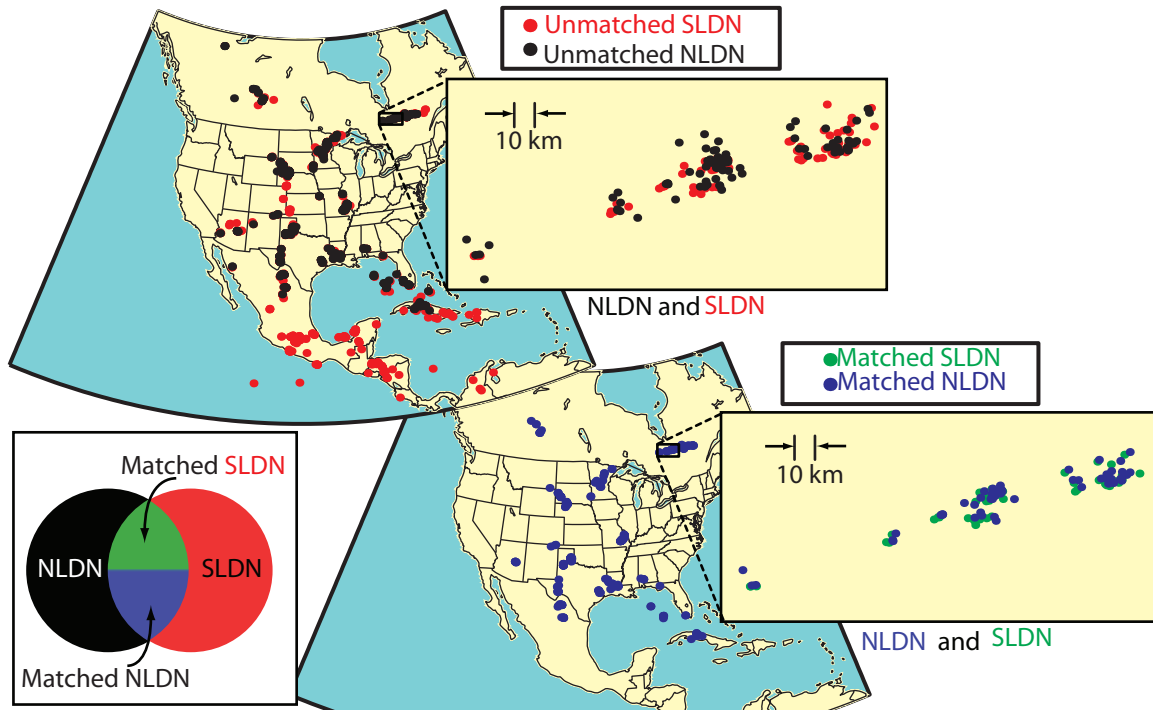
Figure 5.12: Matched and unmatched strike locations for the NLDN and SLDN networks. Data taken from a one-minute period between 22:05 and 22:06 UT.

in eastern Canada with the two datasets clustering in the same regions. This clustering suggests that the strike locations reported by SLDN are not spurious events and indeed correspond to lightning activity. A more detailed discussion of the unmatched events is given in the next subsection.

Figure 5.13 shows a third qualitative plot using lightning data from SLDN only. Two 45-minute synoptic segments of SLDN data are superimposed on IR (temperature) data from the Geostationary Operational Environmental Satellite (GOES)-12. The data on the left plot is at least 45-minutes delayed from the time of the satellite photograph; the data on the right plot straddles the time of the photograph. In both instances, the storm clusters coincide with cold cloud tops, further suggesting a minimum amount of spurious data. In particular, the lightning activity in the Atlantic ocean coincides with very cold cloud tops. The minimum detection thresholds set at each receiver combine to limit the range of the network for a particular stroke peak current and time of day. The regions with cold cloud tops that do not show lightning
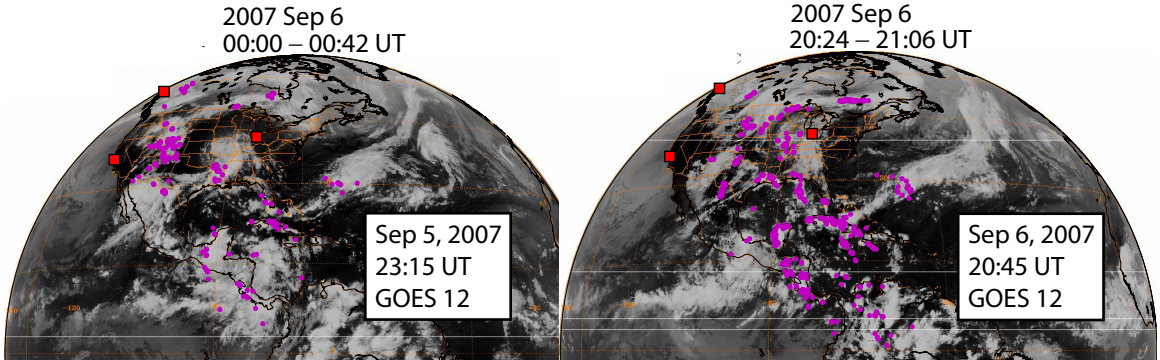
Figure 5.13: SLDN stroke data superimposed on GOES IR satellite images.

activity may have been cirrus clouds, or may have lightning strokes which produce sferics below the threshold set at one or more of the receivers. Also, the lightning activity may have been sufficiently low such that no sferics were produced that made it through the quality criteria of the network.

The following discussion seeks to quantitatively evaluate the SLDN location accuracy using the NLDN data set as a reference. The same match criteria is used as above: a temporal- and spatial-coincidence of <60 μs and <20 km, respectively. Figure 5.14a shows a scatter plot of the SLDN location error for the matched strikes for a small storm cluster in Louisiana under a daytime ionosphere, using NLDN data (which has a location accuracy of $\lesssim 0.5$ km) as a reference. An ellipse corresponding to the standard deviation of the error distributions along the major and minor axes is shown in red, along with these distributions on the right. Figure 5.14b shows a simulated error scatter plot using a Monte Carlo simulation where the arrival time error is normally distributed at each of the three receivers with $\sigma_t = 2$ μs. The resulting distributions along the major and minor axes are consistent with the measured scatter plot, suggesting, for daytime paths, a receiver time error of $\lesssim 2$ μs.

The tilt of the error ellipse is due to the network geometry and the location of this particular storm. A derivation of the expected ellipticity and angle of the error ellipse is given in Appendix C.

For the following global location accuracy analysis, the SLDN dataset was limited to the latitude range [25,55] and the longitude range [−120,−75] so that the statistics
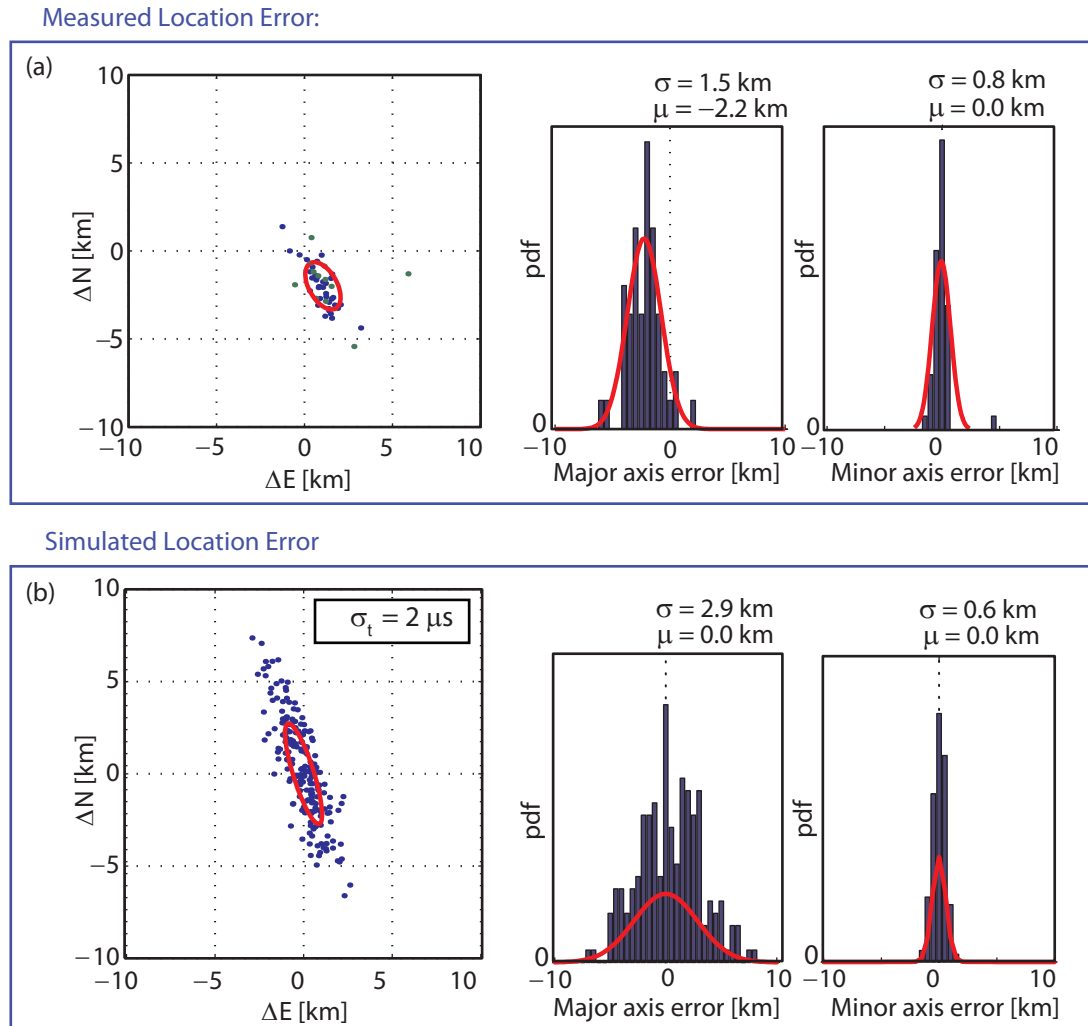
Figure 5.14: Measured and simulated location error scatter for an isolated storm in Louisiana under a daytime ionosphere.
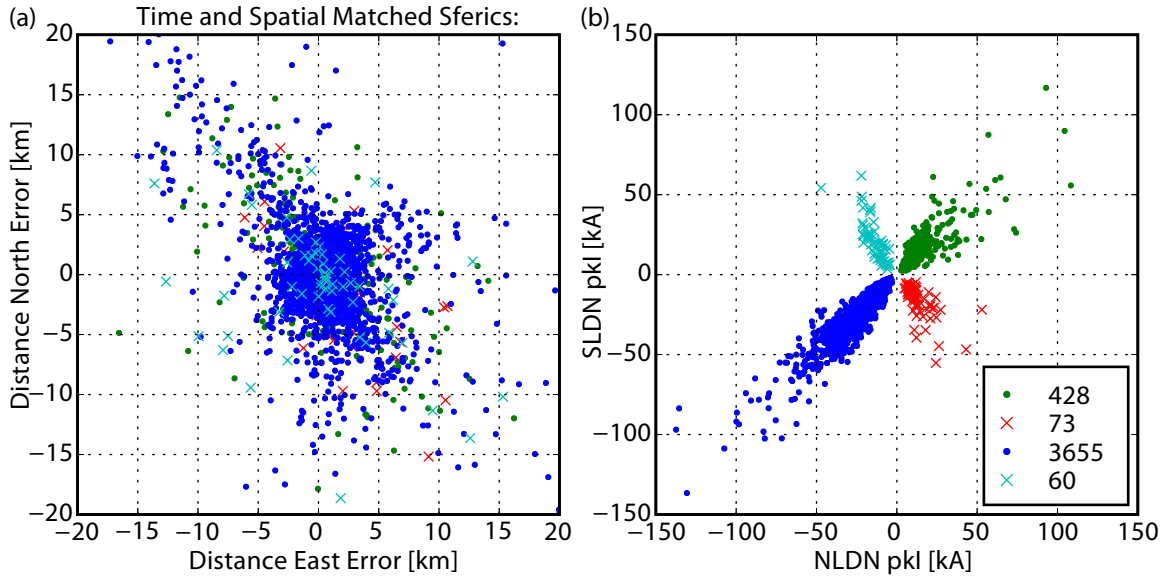
Figure 5.15: Location error and peak current scatter plots for all matched SLDN events during September 6, 2007 over most of the United States (see text).

are not skewed due to the lack of NLDN's coverage far beyond the continental US and Canada and to minimize the geometric effects of SLDN. Figure 5.15b shows the SLDN-determined peak current against the NLDN peak current for the entire dataset available on September 6, 2007. The blue points correspond to strikes that were identified as negative strokes (either ground or cloud) by both networks. Similarly, the green points were identified as positive strokes in both networks. The red (cyan) '×' events are identified as positive (negative) strokes by NLDN but as negative (positive) strokes by SLDN. Out of 4216 matched strokes, 4083 (96.8%) were identfied with the same polarity by both networks. In each quadrant, the magnitude of the slope is close to unity, suggesting the amplitude calibration procedure described in Section 5.2.2 is effective. Figure 5.15a shows the corresponding scatter plot of NLDN-referenced distance errors for the same dataset. This scatter plot contains four color-coded types of strikes, which correspond to the same events and color codes in Figure 5.15b. This scatter plot shows no constant bias, indicating no consistent offset at any of the three receivers. There is also a slight elongation in the North-South direction due to SLDN's geometry, consistent with Figure C.2c.

The scatter plot in Figure 5.15a contains strokes throughout the entire 24-hour period. Figures 5.16a,b show CDF's that better quantify this information. Subplot (a) shows the CDF of the distance magnitude error; subplot (b) shows the CDF of the angle, measured in degrees East of North, of each of the distance errors. The angle distribution is more tightly clustered near 0%, consistent with the network geometry (see Appendix C). The 50$^{\text{th}}$ percentile overall strike distance error for correctly-identified negative strokes is ~1 km; for positive strokes, ~1.5 km. The incorrectly matched polarities have a worse location median error of ~2 km.

The poorer performance for positive strokes is not surprising. Since negative ground strokes dominate the summer activity in North America, the empirical waveform bank and the corresponding polynomial and grid correction factors are tuned to this polarity. Also, these correction factors are all tuned to ground strokes (as reported by NLDN) exclusively. Figure 5.16c displays the CDF for ground strokes; the error curves for both polarities align much more closely. Subplot (e), which shows the CDF for cloud strokes, indeed displays a lower overall performance compared to (c).

Figures 5.16d and f show the CDF distributions for all daytime and all nighttime paths, respectively. Due to a consistent ionosphere, the daytime results enjoy a lower median error of $\lesssim$1 km with a relatively quickly decaying tail. The nighttime paths fare considerably worse, with a median strike distance error of ~2 km and a slowly decaying tail. This result is to be expected since the nighttime paths are more variable [*Thomson et al.*, 2007]. Also, only one set of correction factors are employed for all nighttime paths. More nuanced corrections for the nighttime ionosphere, one that considers the time since sunset, for example, may improve the location estimates.

5.3.2 Detection Efficiency

In addition to the location accuracy, the detection efficiency (DE) and an estimate of the number of spurious events are important performance metrics of a lightning geo-location network. The DE may be defined as a flash or stroke DE. A stroke (flash) DE gives the percent of lightning strokes (flashes) detected by a network. The
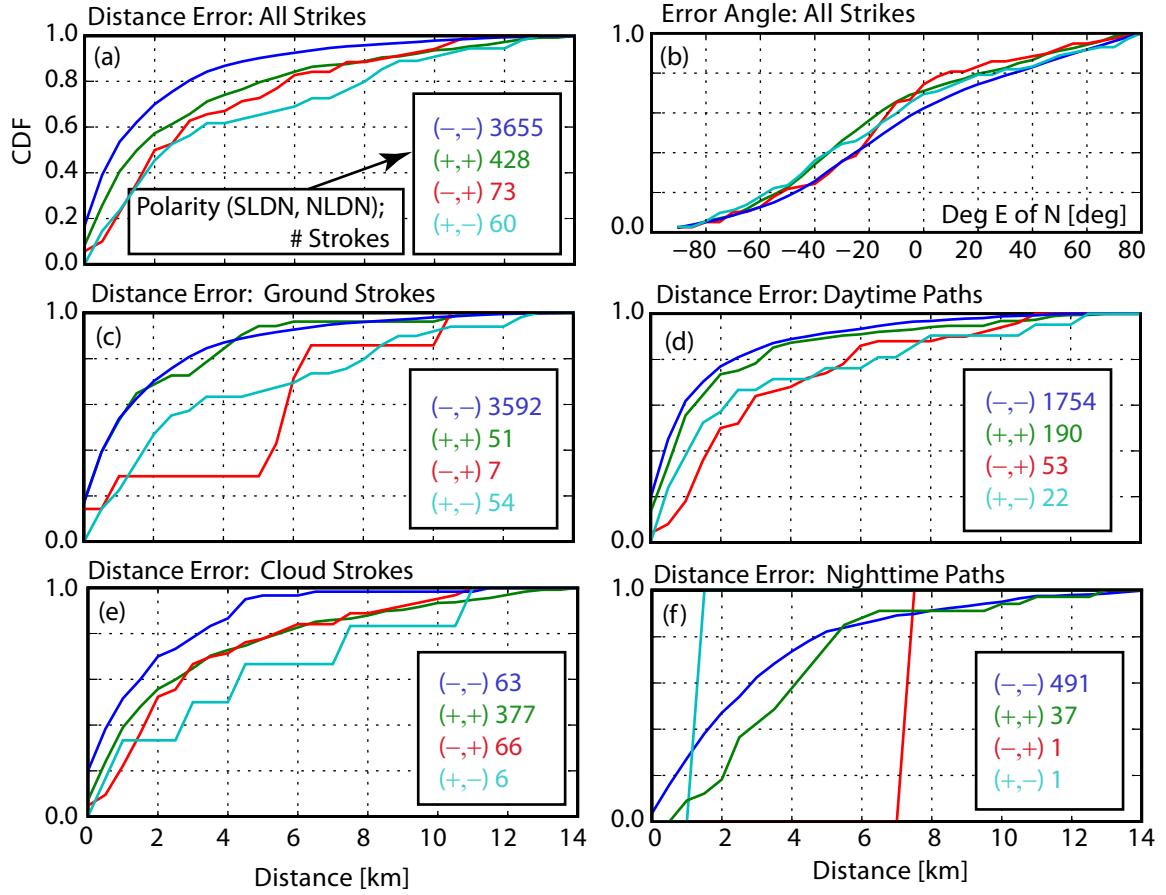
Figure 5.16: Location error cumulative distribution functions for (a) all strokes, (c) ground strokes, (e) cloud strokes, (d) daytime paths, and (f) nighttime paths. (b) Cumulative distribution function for the angle of the polar distance for all events. The legends indicate the number of events in each category, using the same color code for the NLDN- and SLDN-reported peak current polarities used in Figure 5.15.

flash DE will be higher, since a flash is detected if any of the constituent strokes are detected. The DE is also sometimes partitioned into a (stroke or flash) ground or a cloud DE. Some authors weight the peak-current-dependent DE by a peak current distribution function to arrive at a single value [e.g., *Pessi et al.*, 2009]; this section instead only evaluates the stroke DE as a function of peak current. Here the stroke DE is defined as the number of strokes detected by SLDN divided by the number of matched strokes detected by the NLDN in a given NLDN-reported peak current range. Strokes are matched if the time and spatial coincidence are within 60 μs and 20 km, as above.

The top plot in Figure 5.17a shows a histogram of all NLDN strokes during the synoptic recording periods, distributed according to the NLDN-reported peak current in 2.5-kA bins. The negative events are shown in blue and the positive events in green, where the majority of the low-current (<20 kA) positive strokes are cloud discharges [*Biagi et al.*, 2007]. The bottom subplot shows the percentage of strokes in each peak current bin that were matched to SLDN strokes, giving an estimate of SLDN's overall stroke detection efficiency relative to the NLDN. The dip in detection efficiency at higher peak currents is likely due to saturation at the Taylor, Indiana receiver. The dip in DE for the lower peak current values is expected as the SNR degrades at the Juneau, Alaska site.

Figure 5.17b plots the converse: a peak current distribution of SLDN strokes on the top plot with the percent in each bin that were matched to NLDN on the bottom plot. The high correlation rate for large ($\gtrsim$20 kA) strikes coupled with the 90% detection efficiency of NLDN suggests very few spurious events in the SLDN. These results were obtained using three stations only, without the benefit of a redundant time measurement. Good matches were therefore judged only by their range, azimuth, and amplitude consistency. These measurements are useful factors for evaluating the merit of a geo-location, but not as powerful as a redundant time measurement, especially at large distances.

Figure 5.18 emphasizes the results shown in Figure 5.17. Figure 5.18a shows the peak current distribution for the matched (top plot) and unmatched (bottom plot) NLDN strokes, and Figure 5.18b shows the same distributions for the SLDN dataset.
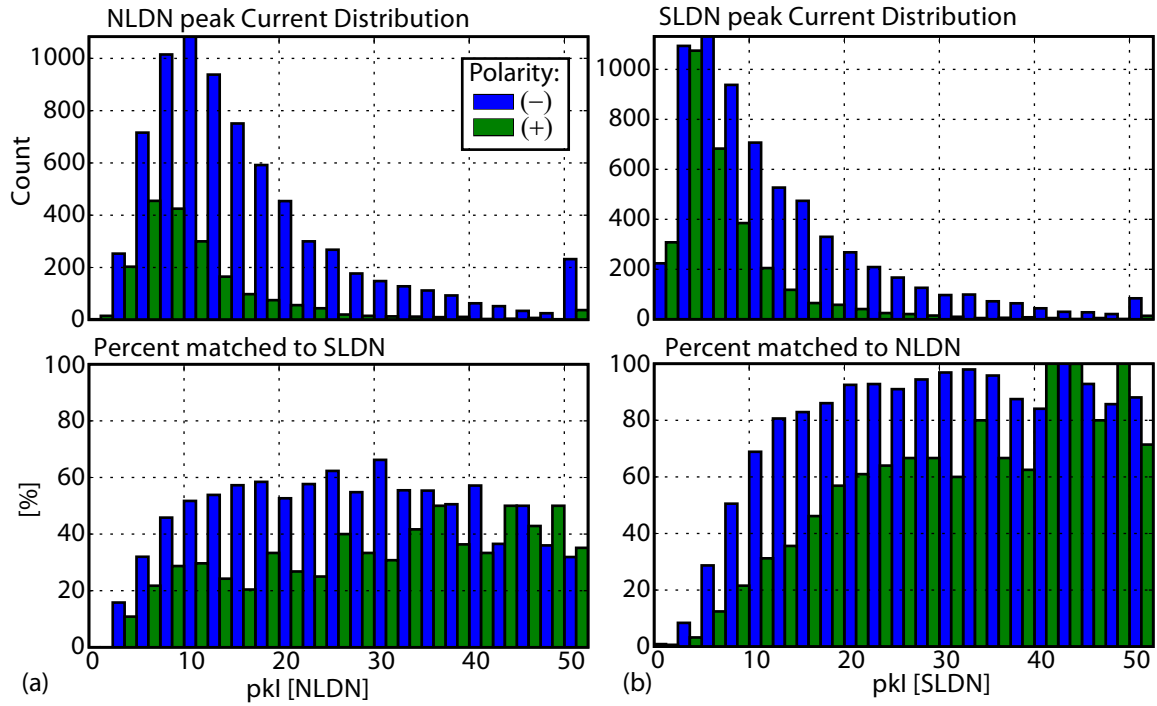
Figure 5.17: (a) NLDN peak current distribution (top plot) and SLDN detection efficiency (bottom plot) for each peak current bin using NLDN as a reference. (b) SLDN peak current distribution and NLDN detection efficiency for each peak current bin using SLDN as a reference. Each bin is 2.5 kA wide, but is drawn with a 1.25 kA width to plot the positive and negative events on the same axis. The bin covering 50–52.5 kA contains all strokes with an absolute peak current >50 kA.
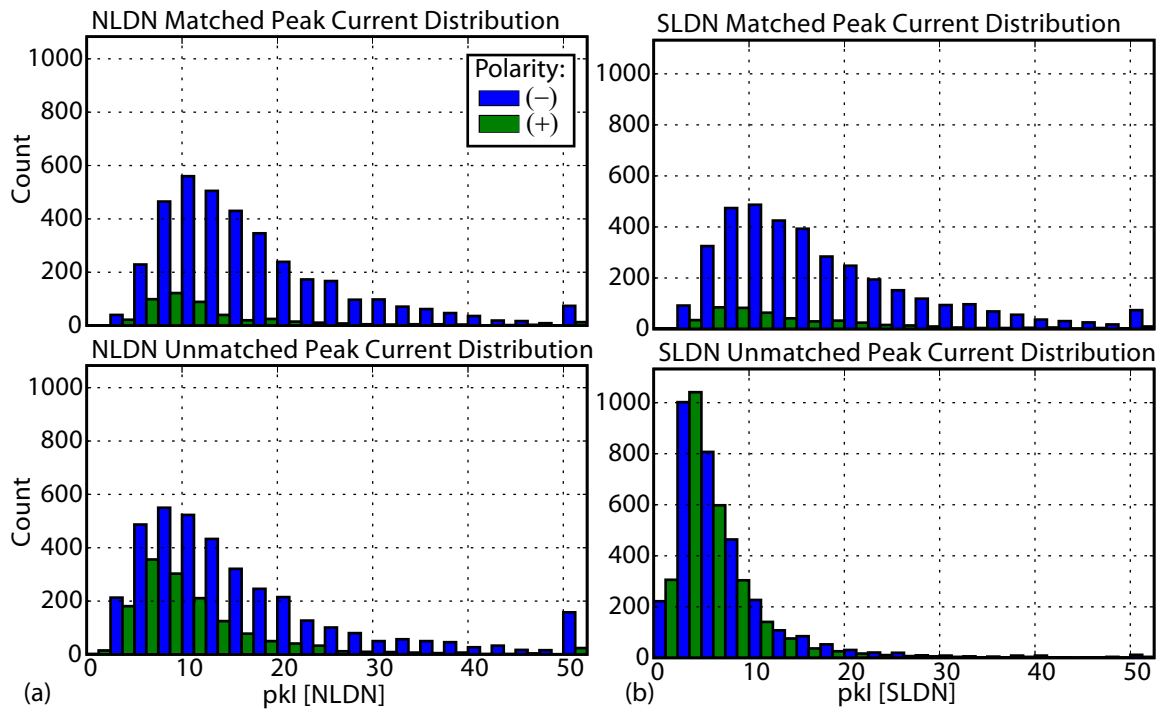
Figure 5.18: Matched (top panel) and unmatched (bottom panel) peak current distributions for the NLDN (a) and SLDN (b) datasets. The distribution bins are allocated in the same way as in Figure 5.17.

The matched (unmatched) NLDN distribution is slightly skewed to higher (lower) peak current values, reflecting the lower stroke detection efficiency for the weaker peak current events. The SLDN distribution has a more pronounced bias toward smaller peak current for the unmatched events.

A likely possibility is that many of these unmatched strikes are cloud discharges or weak peak current events (possibly weaker subsequent strokes in a multi-stroke flash) not registered by the NLDN. During the time of this dataset, the NLDN was only reporting $\sim$10% of cloud flashes and the 50%-detection efficiency threshold was $\sim$5 kA [*Cummins and Murphy*, 2009]. Given the geographic clustering of unmatched events in Figure 5.12 from both networks and the low DE of NLDN for cloud and weak CG strokes, it is plausible that many of the unmatched SLDN strokes coincide to real lightning events not registered by NLDN. The CDF's in Figure 5.10 further support geographic clustering of the unmatched events: there is a higher percentile of strikes with a close spatial proximity (Subplot c) than a close temporal proximity (Subplot b), indicating that many of the unmatched strikes are within the same storm cluster.

While the skew of unmatched SLDN events to low peak current may be explained by the low DE of the NLDN to cloud and weak peak current events, the close distribution of the positive and negative unmatched events may additionally result from the low SNR of these events. For lower peak current events, which will yield a lower SNR measurement at each receiver, the ability of each station to estimate the correct polarity diminishes, as seen by the difference between the correlation coefficient between the two polarities in Figure 5.3b. Without a fourth station, the polarity estimation is made only on the basis of the errors in the range estimates and the correlation coefficients, so a reduced accuracy in the polarity estimation with lower peak current is expected. The unmatched lower peak current events are therefore likely a combination of (mostly positive) cloud strokes and small negative CG strokes, with polarity mixing due to the low SNR for both types of events. This conclusion is supported by the slight bias towards positive unmatched events for strokes with peak current between 0 and 5 kA, and the bias towards negative peak current for strokes with peak current >5 kA, because cloud strokes tend toward a lower peak current

distribution compared to CG strokes [*Murphy and Cummins*, 1998].

The SLDN long-range network, using the VLF signature only, does not currently differentiate between the ground and cloud discharges. The NLDN uses a VLF/LF measurement to analyze the stroke rise and recovery times specifically to differentiate between cloud and ground strokes. VLF signatures from cloud discharges may be distinguishable from sferics generated by ground discharges, but an ability to make this distinction may also be complicated by the cloud discharge variability. A cloud discharge at a particular altitude z_0 will have a modified mode excitation spectrum in accordance with the height-gain function (2.61). The mode structure will additionally be affected by the dipole's angle, especially under an anisotropic ionosphere [*Pappert and Bickel*, 1970]. The conclusion about the degree to which one can determine the discharge type using long-range VLF measurements is left to future work.

Chapter 6

Summary and Suggestions for Future Work

6.1 Summary

This thesis puts forth the methodology and the algorithms for a new technology for long-range lightning geo-location. Using the VLF portion of the radiated fields, existing techniques have been adapted to mitigate powerline noise and a new nonlinear method to mitigate interference from man-made narrowband transmitters was developed. Since the new method also relies heavily on measuring the arrival azimuth of each sferic using orthogonal magnetic loop antennas, current arrival azimuth techniques were reviewed and a modification was introduced which mitigated polarization errors.

Chapter 4 developed a new paradigm for characterizing sferic waveforms on a global basis. Through an extensive empirical cataloging of sferic waveforms from a variety of source locations and propagation path profiles, we found that the average received waveform variation may be captured with a relatively small number of parameters. Specifically, for a fixed day/night path profile, one can construct a distance-indexed, logarithmically-spaced canonical waveform bank. At geographically remote stations capable of measuring the arrival azimuth of individual sferics, each received waveform may be compared to an azimuth- and local time-dependent stored

146

waveform bank. The result of this correlation gives an estimate of the sferic range, and, crucially, allows the identification of a low time-variance feature of the waveform for use in a triangulation algorithm at the central processor.

Through the use of a simple ray-hop model and an FDTD simulation, the dominant features of nearby lightning strokes were reproduced. For more distant strikes, the sferic waveform is more appropriately modeled as a series of modes. For these long-range strikes, a rich theoretical background has been previously developed by *Budden* [1961a] and *Wait* [1970], among others, with several published propagation parameters. Several more recent robust numerical models exist to model this type of long-range VLF propagation [*Pappert and Ferguson*, 1986; *Lehtinen and Inan*, 2008], which match well with the empirical propagation parameters derived from our empirical measurements.

These published and modeled results, together with the empirical spread of the received waveform, give us the insight on how best to extract range and timing information from each sferic. After a cross-correlation with each entry of the waveform bank is used to identify the range, either a threshold crossing time is identified as the ground wave or an early zero-crossing time is isolated to yield a reliable timing measurement. In this way, the spread in arrival times is minimized for a wide range of distances, giving equally good results for sferics with high and low modal content. The timing spread is consistent with variations in the initial rise time of the measured ground wave in *Lin et al.* [1979].

Using this new technique for extracting arrival time and correcting for second-order propagation effects empirically, three stations were used to triangulate events in the continental United States and surrounding areas for several synoptic periods in a 24-hour time span. By using NLDN as a reference dataset, strike location errors for daytime paths had a median value of ~ 1 km, whereas nighttime paths had a median location error of ~ 2 km. An overall stroke detection efficiency of 40–60% was, and a significant amount of tightly clustered lightning activity detected that was not registered by NLDN and that may be from many weak CG and cloud discharges. This location accuracy and detection efficiency either compare favorably or are significantly better than the performance metrics reported by the existing long-range networks,

which were summarized in Section 1.3.

6.2 Suggestions for Future Work

In this work, only two waveform banks were used in the triangulation process: one corresponding to all daytime and one corresponding to all nighttime propagation. Figure 5.7 suggests that the polynomial regression curves for the intermediate day/night paths may be linearly interpolated between the all-day and all-night curves. Nevertheless, the day-path and night-path waveforms are distinctly different, as is evident in Figure 4.5. For a robust geo-location network, intermediate waveform banks should be developed, through either additional measurements or modeling.

To flush out these intermediate waveform banks and to make better a priori predictions of the secondary correction grids for the propagation delay and the received amplitude, a more robust subionospheric propagation model should be developed. The simple ray-hop model outlined in Chapter 4 provides a starting basis for a more complete model. In other work the modal model developed in *Pappert and Ferguson* [1986] has been successfully employed to use average sferic measurements to estimate the D region ionosphere [*Cummer et al.*, 1998]. While in this work an averaged ionosphere was assumed between the source and the receiver, the underlying LWPC software is capable of a segmented path and therefore may be well-suited to predict sferic waveforms for real-world path profiles. Modeled results based on FDTD simulations may also be used to account for a variable path profile [*Chevalier and Inan*, 2006].

The correction grids used in this work were derived from summer propagation paths. It is known that the ionosphere has seasonal variations [e.g., *Watt*, 1967, p. 361], and corrections to account for this will no doubt need to be made. Also, snow and ice has a lower conductivity, which may further impact the seasonal variation of the received waveforms.

The waveform banks were also trained primarily on negative ground strokes from lightning from the mid-west region of the United States. The particular day analyzed in this work did not have many positive ground strokes reported by the NLDN,

resulting in insufficient statistics available to fully evaluate the ability of a waveform bank trained on negative ground strokes to successfully cross-correlate with positive ground strokes. It is possible that the DE and LA for positive strokes may benefit from a separate waveform bank (applied at each receiver in parallel to the 'negative stroke' waveform bank) and/or correction grid. Since cloud discharges radiate a spectrum consistent with a similar discharge process but a shorter channel length, including a weaker field at low frequencies [*Vine*, 1987], the source term B_0 for cloud discharges may also be sufficiently different to warrant a separate waveform bank and/or correction grids.

An expanded and refined waveform bank may not only improve the detection efficiency for different discharge types, it may also improve the single-station range estimation by better predicting the expected range of sferic waveshapes at each sensor. This improvement would lead to both an increase in overall DE and a reduction in the number of spurious events. As discussed in Section 5.2.1, the challenge of correctly matching sferics from different sensors to a particular stroke was aided through the use of an arrival azimuth measurement and a range estimation reported by each receiver. Improvements to either of these two measurements further decreases the complexity of the sferic matching problem, which in turn helps to reduce the chance of miss-correlating sferics and improve the overall detection efficiency. By improving the range estimation from $0.2R$ to $0.1R$, for example, the number of possible combinations to consider will be reduced by a factor of $\sim$4. In Figure 5.17, the overall detection efficiency is seen to fall off for higher peak current strokes. These strokes, especially for positive ground strokes, are precisely the events needed for studying high-altitude transient luminous events [*Boccippio et al.*, 1995]. A further refinement in the waveform bank may therefore also be necessary to accommodate for saturation at close range.

Finally, the network algorithm presented in this thesis assumes that the propagation path profiles are steady over large timescales, changing only with the angle of the sun. More sudden ionospheric perturbations caused by energetic particles precipitating out of the magnetosphere and temporarily enhancing the ionization of the lower ionosphere have been well-documented [e.g., *Inan et al.*, 1985]. This sudden

change perturbs the ionospheric path profile and may temporarily skew the arrival time and/or peak current estimate at one or more receiver sites. With sufficient network redundancy, a temporal spike in the cost function may indicate a sudden change in a path profile, which perhaps could in turn be used as a diagnostic tool to monitor the ionospheric stability.

Appendix A

Best-fit line

In section 3.2 the arrival azimuth was calculated by using the slope of a straight line passing through the origin fitted to a collection of points, which were either parametric points from a time-domain waveform or polar points derived from a Fourier Transform. This appendix derives the slope a for this line.

A least-squares linear regression fit minimizes the sum of squared vertical errors. If the slope is close to infinity, small deviations in slope will lead to substantial swings in the residuals. Instead, a method is needed that is symmetric with respect to the slope. A solution is to minimize the sum of squared perpendicular errors to the best-fit line. For a more general derivation of the best-fit line through a collection of points which minimizes the sum of the squared normal error to the line, see *Pearson* [1901].

The shortest distance between a point $P_i = (x_i, y_i)$ and a line L with slope a that goes through the origin is

$$d(P_i, L) = d_i = \frac{-ax_i + y_i}{\sqrt{1 + a^2}}. \tag{A.1}$$

Given a collection of points P_i, our goal is to find the slope a which minimizes the sum of squared distance errors:

$$\min_a \sum_i d_i^2 = \min_a \frac{\|\mathbf{y} - a\mathbf{x}\|^2}{1 + a^2} \tag{A.2}$$

151

where $\mathbf{x} = [x_1, x_2, \ldots, x_N]^T$ and $\mathbf{y} = [y_1, y_2, \ldots, y_N]^T$. Setting

$$\frac{d}{da}\left(\frac{\|\mathbf{y} - a\mathbf{x}\|^2}{1 + a^2}\right) = 0 \tag{A.3}$$

yields

$$a = \alpha \pm \sqrt{1 + \alpha^2} \tag{A.4}$$

where

$$\alpha = \frac{\mathbf{y}^T\mathbf{y} - \mathbf{x}^T\mathbf{x}}{2\mathbf{y}^T\mathbf{x}}. \tag{A.5}$$

The two solutions for a give the slopes corresponding to the maximum and minimum sum of square distances. These will be orthogonal; the error term

$$\frac{\|\mathbf{y} - a\mathbf{x}\|^2}{1 + a^2} \tag{A.6}$$

is calculated for each value of a given by (A.4) and the slope corresponding to the minimum error is chosen as the best fit.

Appendix B

Calibration

Direct comparison of received waveforms form different receivers is only meaningful if waveform differences due to hardware variation are first compensated for. In this appendix, the calibration procedure applied to all data used in this thesis is reviewed. A more thorough treatment can be found in *Paschal* [1988], which covers all of the design elements of a VLF receiver. A description of a more recent hardware design, the so-called Atmospheric Weather Electromagnetic System for Observation, Modeling, and Education (AWESOME) receiver, which was also used to collect data for this work, is found in *Cohen et al.* [2009].

B.1 Amplitude

We wish to relate the numerical values returned by the analog to digital converter to the magnetic flux that passes through the (electrically small) magnetic field antenna. By Faraday's law of induction, the voltage v_{ind} that will be induced on the terminals of a conductive loop which encloses a surface S is

$$V_{\text{ind}} = -\frac{d}{dt}\Phi \tag{B.1}$$

where the total magnetic flux Φ through the loop is

$$\Phi = \int_S \mathbf{B} \cdot d\mathbf{s}. \tag{B.2}$$

The definition of self-inductance may be used to relate the current to the magnetic field. The definition of self inductance is

$$L = \frac{\Phi}{I}. \tag{B.3}$$

For a loop antenna with N windings and area A, $\Phi = BNA$. So (B.3) becomes

$$L = \frac{BNA}{I}. \tag{B.4}$$

If the current I through the antenna is known, and given antenna values L_a, N, and A, one can calculate the incident magnetic field strength B in Teslas. If a known current I_{cal} is injected in series with the antenna, then the effective incident magnetic field B_{cal} can be calculated. From (B.4), the resulting magnetic field value is given by

$$B_{\text{cal}} = \frac{L_a I_{\text{cal}}}{NA}. \tag{B.5}$$

If the data is digitally recorded when this I_{cal} is injected, one can then relate the measured digital amplitude a_{cal} to the magnetic field B_{cal}. This is typically done at several frequencies. In the AWESOME receiver, a Pseudo-Random Number Generator (PRNG) in the calibration circuit generates a comb signal with frequency components 250.244 Hz apart, allowing for the magnitude calibration of the full bandwidth to be done at once. The calibration number, the factor that converts measured amplitudes to the incident magnetic field strength, is simply $B_{\text{cal}}/a_{\text{cal}}$.

B.2 Phase

There are several sources of distortion introduced between the magnetic loop antenna and the recorded broadband data. In addition to a linear phase delay introduced by

the (site-dependent) length of cable connecting the preamp to the line receiver, a non-linear phase will be introduced from the preamp and filter cards. Also, the National Instruments A/D converter (NiDAQ) interleaves samples from each channel, thereby introducing an additional constant offset of all but the first channel.

In the AWESOME receiver, each channel uses its own calibration chip, preventing a simultaneous phase calibration with the amplitude calibration. We therefore seek an alternative method for calibrating the hardware-dependent phase distortion. Our goal is to refer all phase offsets back to the input of the NS channel (channel 0) on the A/D card. To do this, we measure the phase ratio twice. First, we measure the phase difference between the auxiliary input (channel 2) and the NS-input to the A/D cards by using a pass-through card on both channels. Call this result $h_{\text{ns}'}^{\text{aux}'}$, where the prime denotes a measurement that bypasses all of the electronics and directly feeds the source to the A/D card. Next we measure the phase difference between the input to the NS (or EW) preamp and the auxiliary input to the NiDAQ card. Call this result $h_{\text{aux}'}^{\text{CH}}$, where CH may be NS or EW. Combining $h_{\text{ns}'}^{\text{aux}'}$ with $h_{\text{aux}'}^{\text{CH}}$, which will be a multiplication in the frequency domain or a convolution in the time domain, we recover the overall system phase response of channel CH with respect to a directly sampled input on the NS channel. We can invert the result to form a distortion-compensation filter, which allows for a phase calibration on each channel. To calibrate the auxiliary channel, simply reverse EW for AUX in the above procedure.

If the phase difference is measured at several frequencies, we can interpolate on a DFT grid and invert to form an impulse response.

The overall system response is recovered by combining the magnitude response form the previous section with the measured phase response. Figure B.1a shows the system response measured using this technique.

To generate a distortion compensation filter, the system response is inverted. A simple inversion would yield a large gain factors near DC and Nyquist, so care must be taken to saturate the pre-filter at the edges of the usable bandwidth. Figure B.1b shows the prefilter resulting from inverting the system response shown in Figure B.1a.

When re-sampling the frequency response, one must be cognizant of the properties
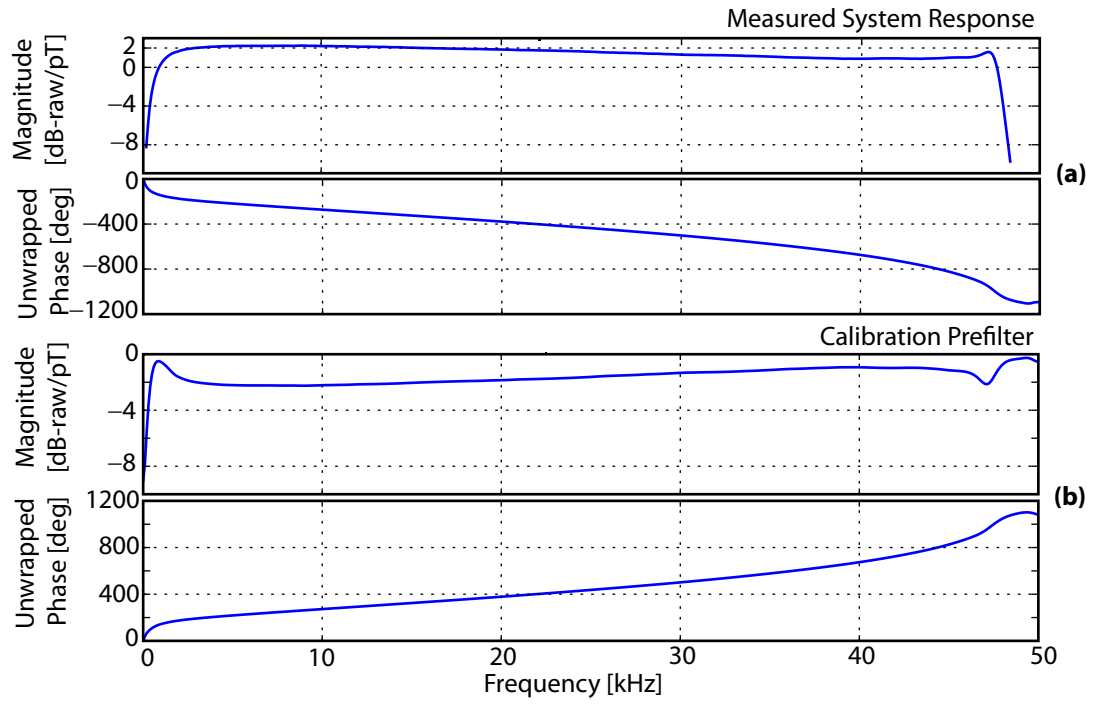
Figure B.1: (a) AWESOME receiver system response. (b) Prefilter used to calibrate the recorded broadband data.

of a DFT. At DC and Nyquist (0 Hz and $f_s/2$), the phase must be 0 or a multiple of π. For real signals, the magnitude is even and the phase is odd. After interpolating from $0 - f_s/2$, one must use these rules to map out the whole DFT range.

Note that the amplitude response of the receiver is relatively flat, but from (B.1) and (B.2), the induced voltage on the antenna terminal leads is proportional to frequency. The sampled voltages are proportional to the induced current, however. For frequencies above the input turnover frequency, defined as the frequency where the reactance of the loop is equal to the resistance of the input circuit, the input impedance Z_{in} is dominated by the reactance of the loop ωL_a and so the induced current $I_{\text{ind}} = V/Z_{\text{in}}$ is directly proportional to the incident magnetic field [*Paschal*, 1988, p. 15]. For a 1.0 Ω, 1.0 mH antenna, with the input turnover frequency set to twice the antenna turnover frequency [for noise reasons; see *Paschal*, 1988, p. 15], this frequency is $2R_a/(2\pi L_a) = 318$ Hz.

B.3 Receiver Sensitivity

The receiver sensitivity dictates a minimum input-referred noise level of a receiver, which places a lower bound on the minimum detectable field strength. This sensitivity is controlled by a gain setting on the preamp and by the antenna factor NA. To maximize the sensitivity potential of a receiver, the receiver sensitivity should be below the natural background noise (which is influenced by both man-made interference and natural emissions). This section discuses the calculation of the receiver sensitivity.

If the data signal $x[n]$ is calibrated and carries the units of, say, picoteslas (pT), then the periodogram P at frequency kf_s/N is given by

$$P_l\left(\frac{kf_s}{N}\right) = \frac{F}{f_s} \frac{|\text{DFT}\{x[n]w[n-l]\}[k]|^2}{\sum_n |w[n]|^2} \left[\frac{\text{pT}^2}{\text{Hz}}\right], \quad F = \begin{cases} 1, & k = 0, \frac{N}{2} \\ 2, & \text{otherwise} \end{cases} \quad \text{(B.6)}$$

and so $\sqrt{P}$ carries units of $\text{pT}/\sqrt{\text{Hz}}$. The spectral resolution is governed by the choice of window $w[\cdot]$. To recover the amplitude of a deterministic tone in pT, $\sqrt{P}$ must be

integrated over the relevant bandwidth[1].

An estimate of the Power Spectral Density (PSD) is obtained by taking the average of many successive periodograms, where each P_l is calculated with a different window offset l:

$$\text{PSD} = \frac{\sum_l P_l}{L} \tag{B.8}$$

where L is the total number of sample periodograms used. This technique for estimating the PSD is known as Welch's method [*Smith*, Accessed 2009].

Figure B.2 shows the measured, uncalibrated PSD from several hardware configurations using the AWESOME receiver. The AWESOME receiver records values with 16-bit sampling, covering the range -2^{15} to $2^{15} - 1$ in increments of 1. Without any noise sources, this gives a dynamic range of $20 \log_{10}(2^{16}) = 96$ dB. On a PSD plot, the lower level at the base of this dynamic range is given by the quantization noise floor (QNF), which is the per-root-Hz contribution of the uniform distribution of ± 0.5 from the least significant bit (LSB). With the LSB set to $\Delta = 1$ and a sampling frequency $f_s = 100,000$ samples/second, the one-sided quantization noise floor (QNF) is

$$\text{QNF} = 10 \log_{10} \left(\frac{\Delta^2}{12} \frac{1}{f_s/2} \right) = -57.8 \left[\text{dB} \frac{\text{raw}}{\sqrt{\text{Hz}}} \right], \tag{B.9}$$

and is shown by the dotted black line. The blue curve shows the measured noise floor, obtained using Welch's method, of the receiver with no filter card. This noise floor is generated from a noise source after the filter card. The green curve shows the noise floor with the filter card inserted into the preamp, but with no preamp connection. This noise floor is generated by a noise source introduced in the anti-aliasing filter

---

[1]If one wishes for the amplitude of each frequency bin to directly correspond to the amplitude of a single-sided sinusoid, then the following expression should be used instead:

$$A[k] = F \frac{\text{DFT}\{x[n]w[n-l]\}[k]}{\sum_n w[n]}, \quad F = \begin{cases} 1, & k = 0, \frac{N}{2} \\ 2, & \text{otherwise} \end{cases}. \tag{B.7}$$

The denominator scaling factor compensates for the reduction in signal strength from the windowing function $w[\cdot]$ and scales the spectral magnitudes such that the amplitude is directly comparable to the time-domain magnitude of a tone. This expression is used to normalize all spectrograms in this thesis.

card and brings the noise baseline to ~-40 dB-raw/$\sqrt{\text{Hz}}$. This raised noise floor is larger than the QNF by $\sim(-40)-(-57)=17$ dB (~3 bits), which reduces the overall dynamic to $\sim96-17\simeq80$ dB.

The next four plots (red, cyan, magenta, and yellow) show the uncalibrated (raw) noise floor with the preamp connected at the four possible gain settings: 0, 10, 20, and 30 dB, respectively, with an impedance-matched dummy loop at the antenna terminals. At the 0 dB gain setting, the noise floor is dominated by the filter card above 20 kHz and by the preamp card below 20 kHz. Increasing the gain in the preamp by 10 dB increases the noise floor by ~7 dB below 20 kHz and ~3 dB above 20 kHz. Since the input signal is amplified by 10 dB, there is an improvement of 3 dB (7 dB) below (above) 20 kHz in the signal to noise ratio (SNR). The dynamic range will also be reduced by 7 dB (3 dB) below (above) 20 kHz. Increasing the gain switch to 20 dB, the noise floor is uniformly increased by 10 dB, and so the SNR will remain the same while the dynamic range is reduced by another 10 dB. The same effect is seen in the 30 dB gain setting.

The optimal gain setting is therefore application dependent. If dynamic range is of primary importance, then the 0 dB gain setting should be used. If sensitivity is more important, than the 10 dB gain setting should be used. There is little added benefit by going to the 20 dB and 30 dB gain settings.

The PSD curves drawn in Figure B.2 correspond to raw uncalibrated data. The input-referred sensitivity is calculated by scaling each frequency component by the calibration number $B_{\text{cal}}/a_{\text{cal}}$, where a_{cal} is the amplitude of the recorded digital signal and where B_{cal} depends on the antenna A area and number of windings N, as in equation (B.5).

Increasing NA of the antenna can improve sensitivity and SNR. If the local measured noise floor is close to the receiver sensitivity, then that site will benefit from an antenna with a larger NA product. Note, too, that increasing NA increases the gain, thereby reducing the maximum peak amplitude before receiver saturation.

Figures B.3a,b show the measured PSD in blue and the measured receiver sensitivity in magenta for receivers in Tucson, AZ and Juneau, AK, respectively. Also shown is a RMS background noise level measured from month-long averages in the spring
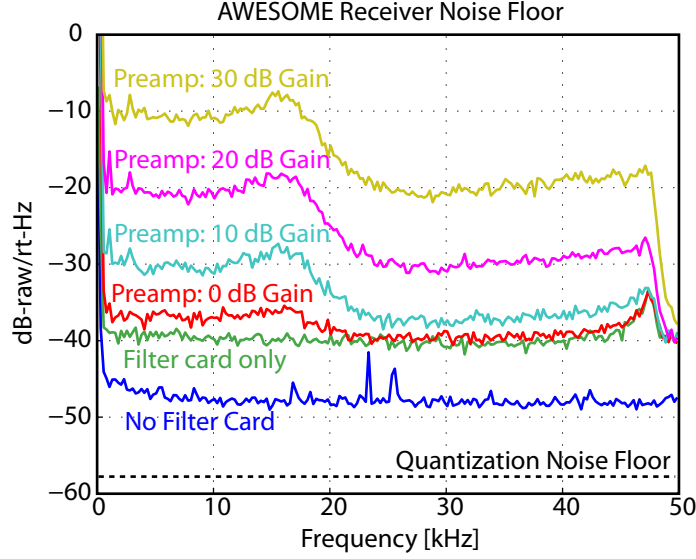
Figure B.2: AWESOME receiver noise floor for several hardware configurations.

at Stanford University [*Chrissan and Fraser-Smith*, 1996]. This curve is labeled as a 'Natural Noise Floor' in the figure, but it should be understood that this measurement was taken over a very long averaging period and therefore includes contributions from sferics. The receiver in Tucson used a 'T1' antenna (A=1.69 m^2, N=12); the receiver in Juneau used a 'T2' receiver (A=17.64 m^2, N=6). The sensitivity of T2 is $20\log_{10}[(17.64)(6)/(1.69)(12)]$=14.4 dB better than T1. In this figure the receiver sensitivity is referred to the input, and so degrades near DC and Nyquist as a_{cal} decreases. At Juneau, the larger T2 antenna used allows for measuring a natural noise floor which is ∼15 dB lower than the sensitivity of the receiver at Tucson. The rise near ∼8 kHz in the measured PSD in at Juneau appears to be caused by the receiver, however.
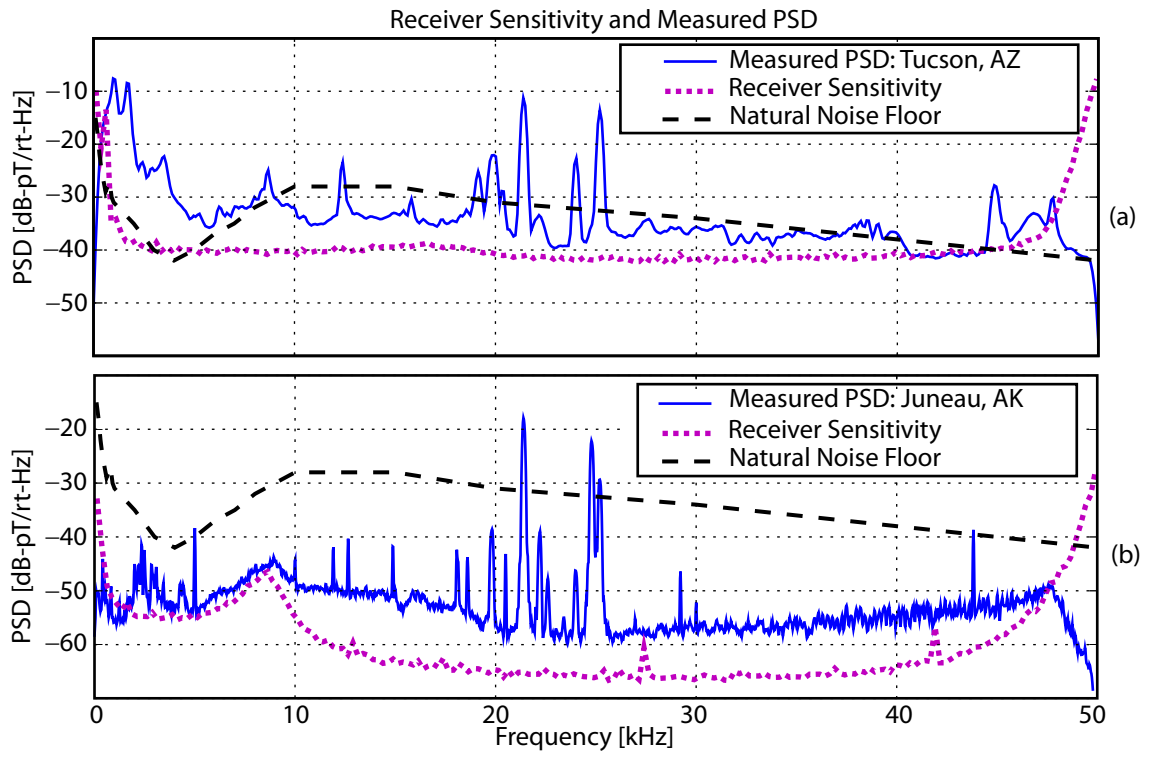
**Figure B.3:** Measured PSD and sensitivity at (a) Tucson, AZ and (b) Juneau, AK. The averaged natural noise floor is shown in both plots as a reference.

Appendix C

Network Design

The error distributions associated with the arrival time, azimuth, range, and amplitude measurements translate to an error in the location position and strike time. This error distribution in turn is distorted by the geometry of the network and the strike location. Figure C.1, which maps out an arrival time difference grid between two different pairs of stations, gives an intuitive understanding of this distortion. The white lines highlight a specific arrival time difference for each station pair. While this work does not explicitly use intersecting ATD curves, the geometric effects are easier to visualize using this construct. The strike location is identified at an intersection between the two ATD curves. If this strike location is somewhere in the region labeled 'Poor Spatial Resolution' in (b), then a small change in the arrival time difference yields a substantial error in the location estimate. Furthermore, if two arrival time difference curves intersect with a shallow angle, which is the case for the two highlighted lines in the curve (the intersection point is out of view of these maps), then again a small timing error yields a significant location error. Thus geometric effects are minimized if the strike position is in the interior of the network.

In Figure 5.14 the measured error ellipse was compared to a Monte Carlo simulation with an assumed Gaussian error distribution on the arrival time and azimuth. To evaluate the geometric effects for a sensor network, one would have to repeat this simulation at each location within the network coverage area. Since each point in the simulation requires an iterative location solution, this is a computationally intensive
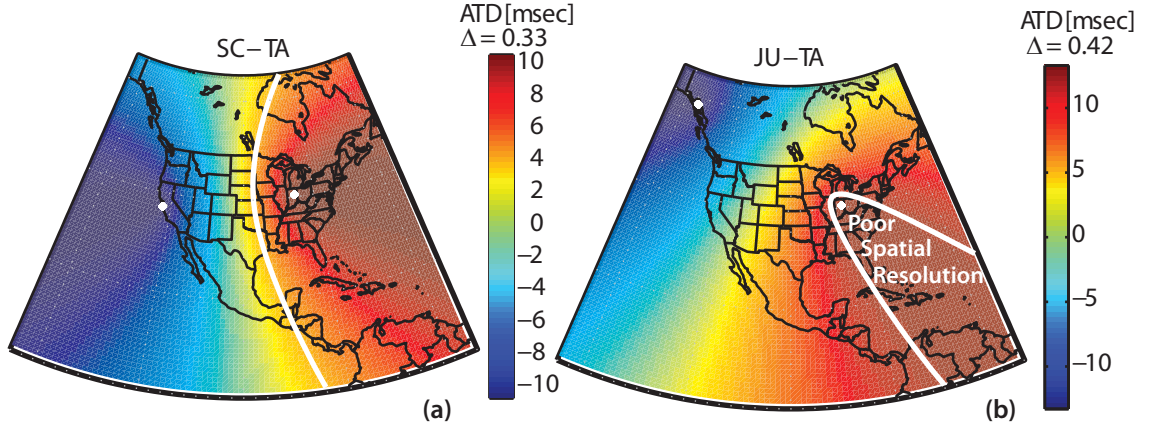
Figure C.1: Arrival time difference (ATD) maps between two different station pairs.

procedure.

A more elegant and computationally efficient approach is to evaluate the Hessian of the cost function at each location. Let λ_0, ϕ_0 be the nominal geo-located latitude and longitude solution given by an array of receivers at latitudes $\boldsymbol{\lambda}$ and longitudes $\boldsymbol{\phi}$ with arrival times and arrival azimuths $\mathbf{t}$ and $\boldsymbol{\theta}$, respectively. At this nominal point, assuming all systematic offsets have been removed, the cost χ^2 is zero. Assuming that the time and azimuth measurements have a normal distribution with associated RMS errors $\boldsymbol{\sigma}_t$ and $\boldsymbol{\sigma}_\theta$ leads to a joint Gaussian distribution of solutions around this nominal point. This distribution may then be characterized with an ellipse.

In the geo-location algorithm each solution has an associated value of χ^2 which indicates the level of consistency among all of the measurements. If we assume that χ^2 is due to the random MSE only, then we can estimate the expected distance error distribution by finding the ellipse formed by the distance vector perturbations from the nominal solution which yield this χ^2 value.

Let $\chi^2(\lambda_0, \phi_0, \boldsymbol{\lambda}, \boldsymbol{\phi}, \mathbf{t}, \boldsymbol{\sigma}_t, \boldsymbol{\theta}, \boldsymbol{\sigma}_\theta)$ be the cost surface centered at the nominal solution λ_0, ϕ_0. A perturbation of dx east and dy north will map to the coordinates $(dy/R_\mathrm{E}, \sec(\lambda)dx/R_\mathrm{E})$ yielding a cost value $f(x, y) = \chi^2(\lambda_0 + dy/R_\mathrm{E}, \phi_0 + \sec(\lambda)dx/R_\mathrm{E}, \boldsymbol{\lambda}, \boldsymbol{\phi}, \mathbf{t}, \boldsymbol{\sigma}_t, \boldsymbol{\theta}, \boldsymbol{\sigma}_\theta)$. The goal is to find the set of values $\{dx, dy\}$ (an ellipse) that yield a constant cost value $f_0 = \chi_0^2$.

This ellipse is the cross-section of a paraboloid fitted to $f(x, y)$ centered at $(x, y) =$

$(0, 0)$ where $f(0, 0) = 0$. The 2-D solution to this problem is a simple extension of a 1-D problem. Consider a function $f : \mathbb{R} \rightarrow \mathbb{R}$ which may be approximated by a quadratic function ax^2, where $a = f''(x)|_{x=0}/2$. The two points of x which yield $f(x) \approx f_0$ are $\pm\sqrt{f_0 a^{-1}} = \pm\sqrt{2f_0[f''(x)|_{x=0}]^{-1}}$.

For the 2-D solution, the second derivative becomes the Hessian H of f:

$$H = \begin{bmatrix} \dfrac{\partial^2 f}{\partial x^2} & \dfrac{\partial^2 f}{\partial x \partial y} \\ \dfrac{\partial^2 f}{\partial x \partial y} & \dfrac{\partial^2 f}{\partial y^2} \end{bmatrix}. \tag{C.1}$$

Let e_1, e_2 and $\mathbf{v}_1$, $\mathbf{v}_2$ denote the eigenvalues and eigenvectors of the matrix $2f_0 H^{-1}$, respectively, where $e_1 > e_2$. Then the major and minor axis lengths (which correspond to two times the x-intercept points in the 1-D case) are simply $2\sqrt{e_1}$ and $2\sqrt{e_2}$ and the major and minor axis directions are given by the eigenvectors $\mathbf{v}_1$ and $\mathbf{v}_2$.

Error ellipses shown in Figure C.2 were calculate in this way for the three-station configuration used in the trial network of Chapter 5, using $\chi_0^2 = 3$, $\boldsymbol{\sigma}_t = (10 \ \mu\text{s})\mathbf{1}$, and $\boldsymbol{\sigma}_\theta = (1°)\mathbf{1}$. As was reasoned above, the error ellipse has smaller axis lengths and is more symmetric in the interior of the network. Furthermore, The $\sim -20°$ slant in the south-west region seen in subplot (c) matches the angle of the error ellipse derived in the Monte Carlo simulation.
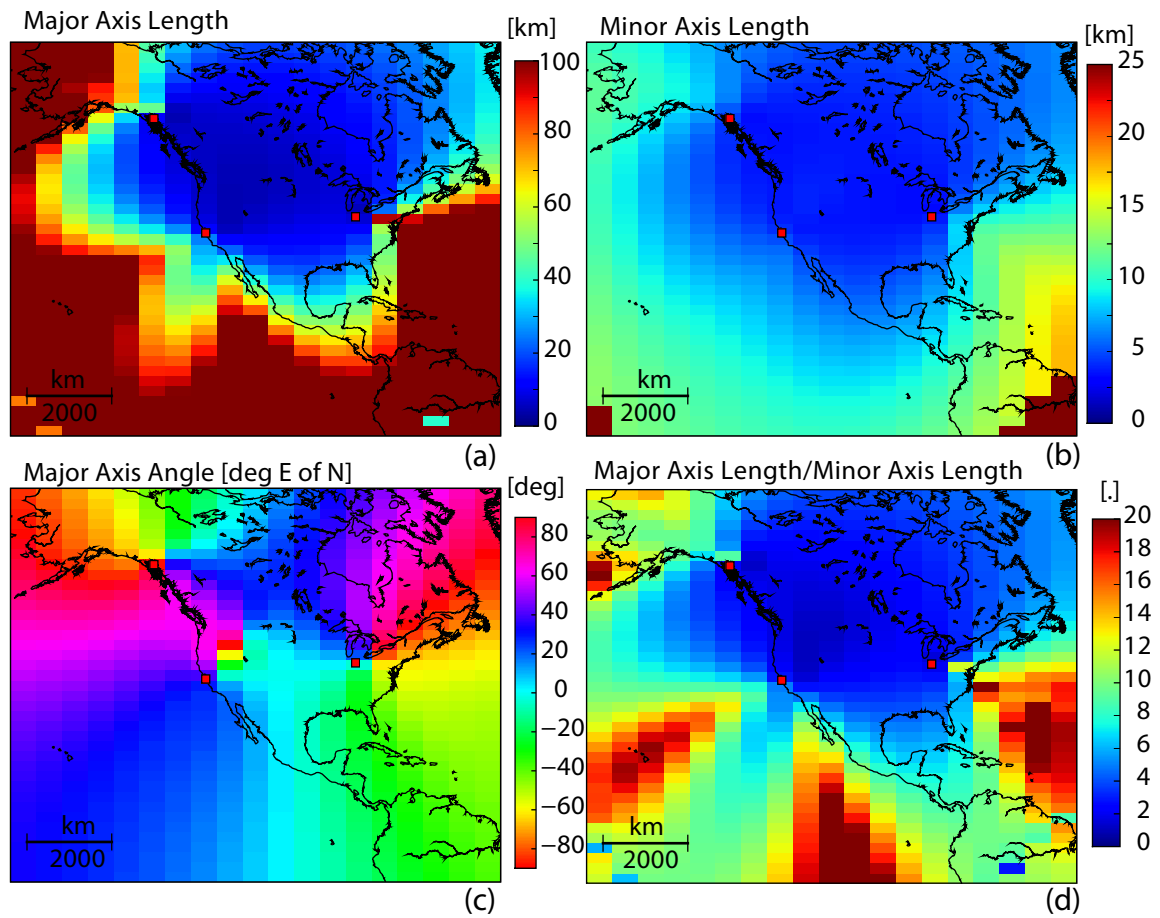
Figure C.2: Error ellipse parameters using the three stations marked by red squares. (a) Major axis length. (b) Minor axis length. (c) Major axis angle. (d) Ratio between the major and minor axis lengths.

Appendix D

Spherical Earth Waveguide

Section 2.4.4 highlighted Budden's approach for dealing with a curved earth whereby the refractive index in the free space below the ionosphere was changed to have an exponential dependence. Another approach widely used in the literature is to begin with a spherical waveguide and solve Maxwell's equations using this geometry. One such approach is derived in detail in *Wait* [1970]; we highlight some of the steps here which are relevant to the results used in this thesis.

Wait starts with a spherical coordinate system (r, θ, ϕ) and a spherical waveguide with inner radius a and outer radius $a + h$. The source is assumed to be at $r = r_0$ and $\theta = 0$ and be radially oriented. As with the parallel plate waveguide, the fields may be derived from a Hertz vector $\mathbf{U}$[1] which, with a radially oriented source, has only a radial component $U_r \equiv U$. The fields are obtained using

$$\mathbf{E} = \nabla \times (\nabla \times \mathbf{r}U), \tag{D.1}$$

$$\mathbf{B} = j\omega\mu_0\epsilon\nabla \times (\mathbf{r}U). \tag{D.2}$$

Away from the source, the function U satisfies $(\nabla^2 + k^2)U = 0$, which leads to solutions of the form $h_n^{(1,2)}(kr)P_n(\cos\theta)$, where $h_n^{(1)}(x)$ $(h_n^{(2)}(x))$ is a spherical Hankel function of the 1st (2nd) kind which describes the incoming (outgoing) wave, and

---

[1]In Chapter 2.1, U was defined in such a way that it carried the units Coulombs. In order to keep the notation consistent with *Wait* [1970], this Appendix will use the definition for $\mathbf{U}$ given by *Wait* [1970, p. 107, equations (1a) and (1b)], so that U now carries the units Volts.

$P_n(x)$ is a Legendre polynomial which captures the angular dependence θ as the wave travels along the waveguide. The solution has the form [*Wait*, 1970, p. 157]

$$U(r, \theta) = \sum_n D_\nu z_\nu(kr) P_\nu(\cos \theta), \ a < r < a + h \tag{D.3}$$

where $z_\nu(kr) = a_\nu h_\nu^{(1)}(kr) + b_\nu h_\nu^{(2)}(kr)$ is the height-gain function, $\nu(n)$ indexes the mode number, and D_ν is an excitation factor that is determined by evaluating U at the source.

By enforcing (2.26) near the source at $(r, \theta) = (r_0, 0)$ and assuming that the height-gain functions are orthogonal to solve for D_ν, *Wait* [1970, p. 160] finds[2]

$$U(r, \theta) \cong \frac{-Il}{2\omega h r_0 \epsilon} \sum_{n=0}^{\infty} \frac{z_\nu(kr_0)}{z_\nu(ka)} \frac{z_\nu(kr)}{z_\nu(ka)} \frac{P_\nu(-\cos \theta)}{\sin(\nu\pi)} \delta_n \tag{D.4}$$

where

$$\delta_n = \left(1 + \frac{\sin(2kh S_n)}{2kh S_n}\right)^{-1}. \tag{D.5}$$

In spherical coordinates, using (D.1) and the following definition for the Legendre Polynomial

$$\frac{1}{\sin \theta} \frac{\partial}{\partial \theta} \left(\sin \theta \frac{\partial P_\nu(-\cos \theta)}{\partial \theta}\right) = -\nu(\nu + 1) P_\nu(-\cos \theta), \tag{D.6}$$

the radial electric field may be written

$$E_r(r, \theta) \cong \frac{Il}{2h r_0 r \omega \epsilon} \sum_{n=0}^{\infty} \delta_n f_n(r_0) f_n(r) \frac{\nu(\nu + 1)}{\sin(\nu\pi)} P_\nu(-\cos \theta) \tag{D.7}$$

where $f_n(r) = z_\nu(kr)/z_\nu(ka)$. For $\nu \gg 1$ and θ not near 0 or π, *Wait* [1970, p. 161]

---

[2]Wait occasionally uses a scalar function ψ, which is related to U_r by $\psi = U_r/(j\eta)$. In this Appendix, any original formulas from *Wait* [1970] using ψ are converted back to an expression for U.

shows that

$$P_\nu(-\cos\theta) \cong \frac{1}{\sqrt{2\pi\nu\sin\theta}} \exp\left[j(\nu+1/2)(\pi-\theta)-j\pi/4\right] \tag{D.8}$$

The WKB approximation (see next subsection) shows that $\nu + 1/2 \approx kaC_n$; if in addition $\nu \gg 1$, then $\nu \simeq kaC_n$ and the radial (vertically oriented at the surface of the earth) electric field may be expressed as

$$E_r = c\frac{jIlk^{3/2}\mu_0}{\sqrt{8\pi}\sqrt{a\sin(d/a)}}e^{j\pi/4}\sum_{n=0}^{\infty}\left(\frac{2}{jhk}\delta_n\right)C_n^{3/2}f_n(r_0)f_n(r)e^{-jkC_nd}. \tag{D.9}$$

If one assumes $B_\phi = E_r/c$, then it is possible to recover the azimuthal magnetic field by removing a $\cos\theta_n$ factor (which accounts for the projection of an angled plane wave on a vertical electric field, which does not apply for the magnetic field). Under the WKB approximation, Wait also shows that the height-gain functions f_n recover the functional form for the parallel-plate waveguide. Finally, if the quantity in parenthesis is equated to the excitation factor r_n, then (2.62) is recovered, except for the substitution

$$\frac{1}{\sqrt{d}} \to \frac{1}{\sqrt{a\sin(d/a)}} \tag{D.10}$$

which corrects for cylindrical spreading in a parallel plate waveguide to spreading along a spherical waveguide.

Two approximations to the Hankel functions are given in *Wait* [1970] which simplify the expression for the solution (D.4). The first uses a WKB or second-order approximation and is suitable for lower frequencies ($\lesssim 15$ kHz). The second approach uses Airy integrals as a third-order approximation of the spherical Hankel functions. No attempt is made here to list all of the derivation steps; rather, an overview is given to assist the reader in parsing the two approaches given in *Wait* [1970].

D.1   2$^{\text{nd}}$ Order Approximation

A second order approximation is made by expanding $z_\nu(kr)$ into upgoing and down-going waves with factors of the form

$$\mathrm{e}^{\pm jk \int_0^z q\,dz} \tag{D.11}$$

where

$$q^2 = S_n^2 + \frac{2z}{a}C_n^2 \tag{D.12}$$

and $C_n = (\nu + 1/2)/(ka)$. S_n and C_n are analogous to $\sin\theta_n$ and $\cos\theta_n$ where θ is the angle measured from the horizontal (note that *Wait* [1970] referenced θ from the vertical and so swaps the definition of S and C). This is equivalent to assuming a parallel plate waveguide with a linearly varying refractive index, which was considered in Section 2.4.4, except here the refractive index is set to 1 at $r = a$ and an extra $\cos^2\theta$ term is included as an additional correction factor for the curved geometry of the upgoing and downgoing rays; this is usually small and is discussed in *Budden* [1961a, p. 280]. This approximation is referred to as the WKB approximation since it uses the phase-integral approximation.

D.2   3$^{\text{rd}}$ Order Approximation

A solution using third order approximation for the Hankel functions is also derived throughout *Wait* [1970]. In (D.3) the solution was summed over a set of waveguide modes in an analogous fashion as the parallel plate waveguide. Another approach is to start with the following expansion of a point source, summed over integer values n:

$$\frac{\mathrm{e}^{-jkr}}{-jkr} = \sum_{n=0}^{\infty}(2n+1)P_n(\cos\theta)\widetilde{f}(k,b,r,n) \tag{D.13}$$

where $\widetilde{f}$ is expressed in terms of $h_n^{(1),(2)}$. Writing similar expressions for the primary and homogeneous solutions to the inhomogeneous wave equation (which contains a point source at $(r, \theta) = (r_0, 0)$) and imposing an impedance match at the upper and lower boundaries, the solution takes the form

$$U = U_0 \sum_{n=0}^{\infty} (2n + 1) f(n) P_n(\cos \theta) \tag{D.14}$$

where $f(n)$ is a function of the impedance at the boundary, k, a, and b; see *Wait* [1970, pp. 176–177] for the exact expression. In this formulation, the series will be slow to converge. Wait transforms (D.14) to a contour integral to make the series converge more rapidly, where the terms will once again be associated with waveguide modes. A double-index change and use of the residue theorem shows that (D.14) is equivalent to

$$U = jU_0 \int_{\text{real+ axis}} \frac{n dn}{\cos(n\pi)} f(n - \frac{1}{2}) P_{n-1/2}(\cos(\pi - \theta)) \tag{D.15}$$

where the contour encircles the positive half of the real axis. Since $f(n - 1/2)$ is close to an even function, the part of the contour above the real axis may be replaced by a contour just below the real axis going form $-\infty$ to the origin. The new contour now runs just below the real axis, and is now extended to a semicircle encompassing the negative imaginary plane. As the radius goes to infinity, the contribution from this part vanishes, and the integral from the line parallel to the real axis may now be evaluated by finding the residues in the negative imaginary half-plane. Replacing $n - 1/2$ by ν yields

$$U = -jU_0 \int_L \frac{\nu + 1/2}{\sin(\nu\pi)} f(\nu) P_\nu(\cos(\pi - \theta)) d\nu. \tag{D.16}$$

This integral may be evaluated in the usual manner by taking the derivatives of $1/f(\nu)$. The resulting expression will contain terms involving Hankel functions (for

the expression from a point source placed on a spherical surface with no upper boundary, see *Wait* [1970, p. 111]). The third-order approximation is now made by expressing the spherical Hankel functions by Hankel functions of order $1/3$. With $x = ka$ and ν as a solution to the modal equation set up by finding the roots in the integrand of (D.16), when $|\nu - x| \ll x$ and $x \gg 1$, the spherical Hankel function may be approximated using

$$x h_{\nu-1/2}^{(2)}(x) \approx \mathrm{e}^{-j\pi/6} \left(\frac{-2\tau}{3} \right)^{1/2} x^{1/6} H_{1/3}^{(2)} \left[\frac{1}{3}(-2\tau)^{3/2} \right] \qquad (\mathrm{D.17})$$

where $\nu = x + x^{1/3}\tau$. This may be written in terms of an Airy integral:

$$w_1(t) = \frac{1}{\sqrt{\pi}} \int_{\Gamma_1} \mathrm{e}^{st - s^3/3} ds = \mathrm{e}^{-j\pi/3} \left(\frac{-\pi}{3t} \right)^{1/2} H_{1/3}^{(2)} \left[\frac{2}{3}(-t)^{3/2} \right] \qquad (\mathrm{D.18})$$

where Γ_1 goes from $\infty \mathrm{e}^{j2\pi/3}$ to the origin, then out along the real axis. With this hindsight, Wait re-casts the problem by starting with the Airy-integral approximation to the mode series generated by (D.16) for a point source placed on a spherical surface with no upper boundary; the resulting expression is (for a point source placed at $r = a$) [*Wait*, 1961]

$$V_0 = 2\sqrt{\pi x} \mathrm{e}^{-j\pi/4} \sum_{s=1,2,\dots}^{\infty} \frac{\mathrm{e}^{-jxt_s}}{(t_s - q^2)} \frac{w_1(t_s - y)}{w_1(t_s)} \qquad (\mathrm{D.19})$$

where the leading factors are such that

$$E_r \propto \frac{\mathrm{e}^{-jka\theta}}{a\sqrt{\theta \sin \theta}} V_0. \qquad (\mathrm{D.20})$$

Here y describes the radial dependence, x captures the angular dependence (which describes the propagation along the waveguide; for a source and observer on the surface, $\theta = ad$, where d is the great-circle distance), and q contains information

about the impedance at the boundary:

$$y = \left[\frac{2}{ka}\right]^{1/3} k(r-a); \qquad x = \left[\frac{2}{ka}\right]^{-1/3}\theta; \qquad q = -j\left[\frac{2}{ka}\right]^{-1/3}\Delta_g, \qquad \text{(D.21)}$$

with $\Delta_g = (1/n_g)\sqrt{1 - 1/n_g^2}$, where $n_g = \sqrt{\epsilon_g/\epsilon_0 + \sigma_g/(j\omega\epsilon_0)}$ is the propagation constant for the ground[3]. The roots t_s are solutions of $w_1'(t) - qw_1(t) = 0$. (D.19) may now be re-cast as the contour integral [*Wait*, 1961]

$$V = e^{j\pi/4}\sqrt{\frac{x}{\pi}} \oint \frac{e^{-jxt}w_1(t-y)}{w_1'(t) - qw_1(t)}dt. \qquad \text{(D.22)}$$

Wait now uses this equation as the starting point for finding the guided mode solutions in a curved waveguide. The solution proceeds in much the same way as Budden's derivation. The downgoing waves carry the functional form $w_2(t)$, where

$$w_2(t) = e^{j2\pi/3}\sqrt{\frac{-\pi t}{3}} H_{1/3}^{(1)}\left[\frac{2}{3}(-t)^{3/2}\right]. \qquad \text{(D.23)}$$

After one reflection, $\exp[-jxt]\,w_1(t-y)$ is converted to $A(t)e^{-jxt}w_2(t-y)$. A similar expression may be used for the lower boundary with reflection coefficient $B(t)$; the resulting expressions are

$$A(t) = -\left[\frac{w_1'(t-y_0) + q_iw_1(t-y_0)}{w_2'(t-y_0) + q_iw_2(t-y_0)}\right]; \qquad B(t) = -\left[\frac{w_2'(t) - qw_2(t)}{w_1'(t) - qw_1(t)}\right] \qquad \text{(D.24)}$$

where q_i is defined analogously to q but for the ionosphere:

$$q_i = -j\left[\frac{2}{ka}\right]^{-1/3}\Delta_i. \qquad \text{(D.25)}$$

This sets up an an infinite series, with powers of $B(t)$ and $A(t)$ indicating the number of ground and ionospheric reflections the wave has undergone. The summation of the resulting power series may be written

---

[3]This is the same expression for Δ as equation (2.32).

$$V = e^{j\pi/4}\sqrt{\frac{x}{\pi}} \oint \frac{e^{-jxt}[w_1(t-h) + A(t)w_2(t-y)]}{[w_1'(t) - qw_1(t)][1 - A(t)B(t)]} dt \tag{D.26}$$

where the contour is set to enclose poles at t_n, where t_n is a solution to $A(t_n)B(t_n){=}1$. Finally, using the residue theorem, the solution for V is [*Wait*, 1961]

$$V = -2\sqrt{\pi x}e^{-j\pi/4} \sum_{n=0}^{\infty} \frac{e^{-jxt_n}[w_1(t_n - h) + A(t_n)w_2(t_n - y)]}{[w_1'(t_n) - qw_1(t_n)]\left[\frac{\partial}{\partial t}A(t)B(t)\right]_{t=t_n}}. \tag{D.27}$$

The solution (D.27) is used as the starting point for many publications on numerical waveguide calculations; see, for example, *Wait and Spies* [1964].

Bibliography

Biagi, C. J., K. L. Cummins, K. E. Kehoe, and E. P. Krider (2007), National Lightning Detection Network (NLDN) performance in southern Arizona, Texas, and Oklahoma in 2003-2004, *Journal of Geophysical Research (Atmospheres)*, *112*, 5208–+, doi:10.1029/2006JD007341.

Bilitza, D. (2001), International reference ionosphere, *Radio Sci.*, *36*, 261–275.

Boas, M. L. (1983), *Mathematical Methods in the Physical Sciences, Second Edition*, John Wiley & Sons, Inc.

Boccippio, D. J., E. R. Williams, S. J. Heckman, W. A. Lyons, I. T. Baker, and R. Boldi (1995), Sprites, ELF Transients, and Positive Ground Strokes, *Science*, *269*, 1088–1091, doi:10.1126/science.269.5227.1088.

Boyd, S., and L. Vandenberghe (2004), *Convex Optimization*, Cambridge University Press.

Brook, M., M. Nakano, P. Krehbiel, and T. Takeuti (1982), The electrical structure of the Hokuriku winter thunderstorms, *J. of Geophys. Res.*, *87*, 1207–1215, doi:10.1029/JC087iC02p01207.

Bruce, C., and R. Golde (1941), The lightning discharge, *J. Inst. Electr. Eng.*, *88*, 487–520.

Budden, K. G. (1961a), *The wave-guide mode theory of wave propagation*, Logos Press.

Budden, K. G. (1961b), *Radio waves in the ionosphere; the mathematical theory of the reflection of radio waves from stratified ionised layers*, 542 pp., Cambridge.

Budden, K. G. (1962), The influence of the earth's magnetic field on radio propagation by wave-guide modes, *Royal Society of London Proceedings Series A*, *265*, 538–553.

Budden, K. G. (1985), *The Propagation of Radio Waves*, Cambridge University Press.

Burgess, W. C. (1993), Lightning-induced coupling of the radiation belts to geomagnetically conjugate ionospheric regions, Ph.D. thesis, Stanford Univ., CA.

Burke, C. P., and D. L. Jones (1992), An experimental investigation of ELF attenuation rates in the earth-ionosphere duct, *Journal of Atmospheric and Terrestrial Physics*, *54*, 243–250.

Butler, K. E., and R. D. Russell (2003), Cancellation of multiple harmonic noise series in geophysical records, *Geophysics*, *68*, 1083–+, doi:10.1190/1.1581080.

Byrne, G. J., A. A. Few, and M. E. Weber (1983), Altitude, thickness and charge concentration of charged regions of four thunderstorms during trip 1981 based upon in situ balloon electric field measurements, *Geophys. Res. Lett.*, *10*, 39–42, doi:10.1029/GL010i001p00039.

Chevalier, M. W., and U. S. Inan (2006), A technique for efficiently modeling long-path propagation for use in both FDFD and FDTD, *IEEE Antennas and Propagation, IEEE Transactions on*, *5*, 525–528.

Chrissan, D. A., and A. C. Fraser-Smith (1996), Seasonal variations of globally measured elf/vlf radio noise, *Radio Sci.*, *31*, 1141–1152.

Christian, H. J., et al. (1999), The lightning imaging sensor, in *ICAE 99 - International Conference on Atmospheric Electricity, 11th*, pp. 746–749.

Christian, H. J., et al. (2003), Global frequency and distribution of lightning as observed from space by the optical transient detector, *J. of Geophys. Res. (Atmospheres)*, *108*(D1), 4–1, doi:10.1029/2002JD002347.

Chronis, T., and E. Anagnostou (2003), Error analysis for a long-range lightning monitoring network of ground-based receivers in europe, *J. of Geophys. Res.*, *108*(D24), D24, doi:10.1029/2003JD003776.

Chronis, T. G., and E. Anagnostou (2006), Evaluation of a long-range lightning detection network with receivers in europe and africa, *IEEE Transactions on Geoscience and Remote Sensing*, *44*, 1504–1510.

Cohen, M. (2009), ELF/VLF phased array generation via frequency-matched steering of a continuous HF ionospheric heating beam, Ph.D. thesis, Stanford University.

Cohen, M., U. Inan, and E. Paschal (2009), Sensitive broadband ELF/VLF radio reception with the AWESOME instrument, *Transactions on Geoscience and Remote Sensing Letters, IEEE*.

Cummer, S. A., and U. S. Inan (2000), Modeling ELF radio atmospheric propagation and extracting lightning currents from ELF observations, *Radio Sci.*, *35*(2), 385–394, doi:10.1029/1999RS002184.

Cummer, S. A., U. S. Inan, and T. F. Bell (1998), Ionospheric D region remote sensing using VLF radio atmospherics, *Radio Sci.*, *33*(6), 1781–1792, doi:10.1029/98RS02381.

Cummins, K. L., and M. J. Murphy (2009), An overview of lightning location systems: History, techniques, and data uses, with and in-depth look at the U.S. NLDN, *IEEE Transactions on Electromagnetic Compatibility*, *51*.

Cummins, K. L., E. P. Krider, and M. D. Malone (1998a), The US national lightning detection network^{TM} and applications of cloud-to-ground lightning data by electric power utilities, *IEEE Transactions on Electromagnetic Compatibility*, *40*(4), 465–480, doi:10.1109/15.736207.

Cummins, K. L., M. J. Murphy, E. A. Bardo, W. L. Hiscox, R. B. Pyle, and A. E. Pifer (1998b), A combined TOA/MDF technology upgrade of the U.S. national

lightning detection network, *J. of Geophys. Res.*, *103*(D8), 9035–9044, doi:10.1029/98JD00153.

Dowden, R. L., J. B. Brundell, and C. J. Rodger (2002), VLF lightning location by time of group arrival (TOGA) at multiple sites, *Journal of Atmospheric and Terrestrial Physics*, *64*, 817–830.

Ferguson, J. A. (1998), Computer programs for assessment of long-wavelength radio communications, version 2.0, *Tech. rep.*, Space and Naval Warfare Systems Center.

Galejs, J. (1972), *Terrestrial propagation of long electromagnetic waves*, Pergamon Press New York.

Gopalakrishna, S., V. S. Varaprasad, and P. Sitaramaswamy (1984), Study of daytime VLF spectral parameters of atmospherics as a function of azimuth, *International Journal of Electronics*, *56*(3), 339–348.

Gradshteyn, I. S., and I. M. Ryzhik (1980), *Table of Integrals, Series, and Products*, Academic Press.

Helliwell, R. A. (1965), *Whistlers and Related Ionospheric Phenomena*, Dover Publications, New York.

Hiscox, W. L., E. P. Krider, A. E. Pifer, and M. A. Uman (1986), A systematic method for identifying and correcting "site errors" in a network of magnetic direction finders, *Tech. rep.*, LIghtning Location and Protection, Inc.

Horner, F. (1954), The accuracy of the location of sources of atmospherics by radio direction finding, d.S.I.R. Radio Research Station.

Inan, U. S., and A. S. Inan (2000), *Electromagnetic Waves*, Prentice-Hall.

Inan, U. S., H. C. Chang, and R. A. Helliwell (1984), Electron precipitation zones around major ground-based VLF signal sources, *J. of Geophys. Res.*, *89*, 2891–2906.

Inan, U. S., D. L. Carpenter, R. A. Helliwell, and J. P. Katsufrakis (1985), Subionospheric VLF/LF phase perturbations produced by lightning-whistler induced particle precipitation, *J. of Geophys. Res.*, *90*, 7457–7469.

Jacobson, A. R., R. Holzworth, J. Harlin, R. Dowden, and E. Lay (2006), Performance assessment of the world wide lightning location network (WWLLN), using the los alamos sferic array (LASA) as ground truth, *Journal of Atmospheric and Oceanic Technology*, *23*, 1082–1092.

Jayaratne, E. R., C. P. R. Saunders, and J. Hallett (1983), Laboratory studies of the charging of soft-hail during ice crystal interactions, *Quarterly Journal of the Royal Meteorological Society*, *109*, 609–630, doi:10.1256/smsqj.46110.

Kawamura, T., M. Ishii, and Y. Miyake (1988), Site errors of a magnetic direction finder for lightning flashes, in *NOAA, International Aerospace and Ground Conference on Lightning and Static Electricity p 487-494(SEE N89-10429 01-47)*, pp. 487–494.

Krehbiel, P. R. (1986), The electrical structure of thunderstorms, *Tech. rep.*, New Mexico Institude of Mining and Technology.

Krider, E. P., R. C. Noggle, and M. A. Uman (1976), A gated, wideband magnetic direction finder for lightning return strokes, *Journal of Applied Meteorology*, *15*, 301–306.

Leavitt, M. K., D. L. Carpenter, N. T. Seely, R. R. Padden, and J. H. Doolittle (1978), Initial results from a tracking receiver direction finder for whistler mode signals, *J. of Geophys. Res.*, *83*, 1601–1610.

Lee, A. C. (1990), Bias elimination and scatter in lightning location by the VLF arrival time difference technique, *Journal of Atmospheric and Oceanic Technology*, *7*, 719–733.

Lee, A. C. L. (1986), An experimental study of the remote location of lightning flashes using a VLF arrival time difference technique, *Quarterly Journal of the Royal Meteorological Society*, *112*, 203–229.

Lee, A. C. L. (1989), The limiting accuracy of long wavelength lightning flash location, *Journal of Atmospheric and Oceanic Technology*, *6*(1), 4349, doi:10.1175/1520-0426(1989)006⟨0043:TLAOLW⟩2.0.CO;2.

Lehtinen, N. G., and U. S. Inan (2007), Possible persistent ionoization caused by giant blue jets, *Geophys. Res. Lett.*, *34*, L08,804.

Lehtinen, N. G., and U. S. Inan (2008), Radiation of ELF/VLF waves by harmonically varying currents into a stratified ionosphere, with application to radiation by a modulated electrojet, *Geophys. Res. Lett.*, *113*, A06,301.

Lennon, C., and L. Maier (1991), Lightning mapping system, *The 1991 International Aerospace and Ground Conference on Lightning and Static Electricity*, *2*, 10.

Lewis, E. A., R. B. Harvey, and J. E. Rasmussen (1960), Hyperbolic direction finding with sferics of transatlantic origin, *J. of Geophys. Res.*, *65*, 1879–+.

Lin, Y. T., M. A. Uman, J. A. Tiller, R. D. Brantley, W. H. Beasley, E. P. Krider, and C. D. Weidman (1979), Characterization of lightning return stroke electric and magnetic fields from simultaneous two-station measurements, *J. of Geophys. Res.*, *84*, 6307–6314, doi:10.1029/JC084iC10p06307.

Morales, C. A., J. A. Weinman, E. N. Anagnostou, S. J. Goodman, and E. Williams (2002a), Continuous long-range thunderstorm monitoring by a VLF receiver network. part ii: Cloud-to-ground and intra-cloud detection efficiency, in preparation to be submitted to Journal of Atmospheric and Oceanic Technology.

Morales, C. A., J. A. Weinman, E. N. Anagnostou, and J. S. Kriz (2002b), Continuous long-range thunderstorm monitoring by a VLF receiver network part i: Instrumentation and location error analysis, submitted to Journal of Atmospheric and Oceanic Technology.

Morales, C. A., E. N. Anagnostou, E. Williams, and J. S. Kritz (2007), Evaluation of peak current polarity retrieved by the ZEUS long-range lightning monitoring system, *IEEE GEOSCIENCE AND REMOTE SENSING LETTERS*, *4*(1), 32–36.

Morgan, R. R. (1968), World-wide VLF effective conductivity map, *Tech. rep.*, Westinghouse Electric Corporation Report 8013F-1.

Murphy, M. J., and K. L. Cummins (1998), Two-dimensional and three-dimensional cloud discharge detection, in *Int. Lightning Detection Conf., Tucson, AZ.*

Nyman, D. C., and J. E. Gaiser (1983), Adaptive rejection of high-line contamination, *SEG Technical Program Expanded Abstracts*, *2*(1), 321–323, doi:10.1190/1.1893897.

Oppenheim, A. V., and R. W. Schafer (1989), *Discrete-Time Signal Processing*, Prentice-Hall Signal Processing Series.

Pappert, R. A., and J. E. Bickel (1970), Vertical and horizontal VLF fields excited by dipoles of arbitrary orientation and elevation, *Radio Sci.*, *5*, 1445–+.

Pappert, R. A., and J. A. Ferguson (1986), VLF/LF mode conversion model calculations for air to air transmissions in the earth-ionosphere waveguide, *Radio Sci.*, *21*, 551–558.

Paschal, E. W. (1988), The design of broad-band VLF receivers with air-core loop antennas, *Tech. rep.*, Stanford University.

Passi, R. M., and R. E. López (1989), A parametric estimation of systematic errors in networks of magnetic direction finders, *J. of Geophys. Res.*, *94*(D11), 13,319–13,328.

Pearson, K. (1901), On lines and planes of closest fit to systems of points in space, *Philosophical Magazine*, *2*, 559–572.

Pessi, A. T., S. Businger, K. L. Cummins, N. W. S. Demetriades, M. Murphy, and B. Pifer (2009), Development of a long-range lightning detection network for the Pacific: Construction, calibration, and performance, *Journal of Atmospheric and Oceanic Technology*, *26*, 145–166.

Press, W. H. (2002), *Numerical recipes in C++ : the art of scientific computing*, Cambridge University Press.

Rafalsky, V. A., A. P. Nickolaenko, A. V. Shvets, and M. Hayakawa (1995), Location of lightning discharges from a single station, *J. of Geophys. Res.*, *100*(D10), 20,829–20,838.

Ratcliffe, J. A. (1959), *The Magneto-Ionic Theory and its Applications to the Ionosphere*, Cambridge.

Reising, S. C., U. S. Inan, and T. F. Bell (1999), ELF sferic energy as a proxy indicator for sprite occurrence, *Geophys. Res. Lett.*, *26*(7), 987–990.

Richards, C. N., and G. A. Dawson (1971), The hydrodynamic instability of water drops falling at terminal veclocity in vertical electric fields, *J. of Geophys. Res.*, *76*, 3445–3455.

Rodger, C., J. Brundell, R. Dowden, and N. Thomson (2004), Location accuracy of long distance VLF lightning location network, *Annales Geophysicae*, *22*, 747–758.

Rodger, C. J., J. B. Brundell, and R. L. Dowden (2005), Location accuracy of VLF world-wide lightning location (WWLL) network: Post-algorithm upgrade, *Annales Geophysicae*, *23*, 277–290.

Rodger, C. J., J. B. Brundell, R. Holzworth, and E. Lay (2009), Growing detection efficiency of the world wide lighting location network.

Sao, K., and H. Jindoh (1974), Real time location of atmospherics by single station techniques and preliminary results, *Journal of Atmospheric and Terrestrial Physics*, *36*, 261–+.

Saucier, A., M. Marchant, and M. Chouteau (2006), A fast and accurate frequency estimation method for canceling harmonic noise in geophysical records, *Geophysics*, *71*, V7–V18.

Saxton, J. A. (Ed.) (1964), *Advances in Radio Research, Volume 1*, Academic Press.

Sazhin, S. S., and M. Hayakawa (1992), Magnetospheric chorus emissions - A review, *Planet. Space Sci.*, *40*, 681–697.

Schulz, W., and G. Diendorfer (1996), Detection efficiency and site errors of lightning location systems, in *Int. Lightning Detection Conference*, Austrian Lightning Detection and Information Systems (ALDIS), Tucson, AZ.

Schunk, R. W., and A. F. Nagy (1980), Ionospheres of the terrestrial planets, *Reviews of Geophysics and Space Physics*, *18*, 813–852.

Sechrist Jr., C. F. (1977), *The Ionospheric D Region, in The Upper Atmosphere and Magnetosphere*, National Academy of Sciences, Washington D.C.

Shafer, D. C. (1994), Spread-spectrum VLF remote sensing of the ionosphere, Ph.D. thesis, Stanford University.

Shvets, A. V., A. P. Nickolaenko, and M. Hayakawa (1997), Characteristics of nearby lightning discharges observed at Singapore, *Journal of Atmospheric and Solar-Terrestrial Physics*, *59*, 1717–1726.

Smith, J. O. (Accessed 2009), *Spectral Audio Signal Processing, October 2008 Draft*, http://ccrma.stanford.edu/∼jos/sasp/, online book.

Thomas, R. J., P. R. Krehbiel, W. Rison, T. Hamlin, D. J. Boccippio, S. J. Goodman, and H. J. Christian (2000), Comparison of ground-based 3-dimensional lightning mapping observations with satellite-based LIS observations in oklahoma, *Geophys. Res. Lett.*, *27*(12), 1703–+.

Thomson, E. M., P. J. Medelius, and S. Davis (1994), System for locating the sources of wideband dE/dt from lightning, *J. of Geophys. Res.*, *99*, 22,793–22,802, doi: 10.1029/94JD02150.

Thomson, N. R. (1981), Whistler mode signals - Spectrographic group delays, *J. of Geophys. Res.*, *86*, 4795–4802, doi:10.1029/JA086iA06p04795.

Thomson, N. R., M. A. Clilverd, and W. M. McRae (2007), Nighttime ionospheric D region parameters from VLF phase and amplitude, *J. of Geophys. Res. (Space Physics)*, *112*, 7304–+, doi:10.1029/2007JA012271.

Uman, M. A. (2001), *The Lightning Discharge*, Dover Publications, Inc.

Vine, D. M. (1987), Review of measurements of the RF spectrum of radiation from lightning, *Meteorology and Atmospheric Physics*, *37*, 195–204, doi:10.1007/BF01042441.

Wait, J. R. (1956), Radiation from a vertical antenna over a curved stratified ground, *Jouranl of Research of the National Bureau of Standards*, *56*(4), 237–244.

Wait, J. R. (1957), The geometrical optics of VLF sky wave propagation, in *Proceedings of the IRE*.

Wait, J. R. (1961), A new approach to the mode theory of VLF propagation, *Journal of Research of the National Bureau of Standards - D. Radio Propagation*, *65D*, 37–46.

Wait, J. R. (1970), On the reflection of electromagnetic pulses from the ionosphere, *Radio Sci.*, *5*(12), 1461–+.

Wait, J. R., and K. P. Spies (1964), Characteristics of the earth-ionosphere waveguide for VLF radio waves, *Tech. Rep. Technical Note 300*, National Bureau of Standards.

Wait, J. R., and L. C. Walters (1963), Reflection of VLF radio waves from an inhomogeneous ionosphere. part I. exponentially varying isotropic model, *J. Res. Nat. Bureau Stand.*, *67D*, 361.

Watt, A. D. (1967), *V.L.F. Radio Engineering*, Pergamon Press.

Weidman, C. D., and E. P. Krider (1986), The amplitude spectra of lightning radiation fields in the interval from 1 to 20 MHz, *Radio Sci.*, *21*(6), 964–970.

Weidman, C. D., E. P. Krider, and M. A. Uman (1981), Lightning amplitude spectra in the interval from 100 kHz to 20 MHz, *Geophys. Res. Lett.*, *8*, 931–934, doi:10.1029/GL008i008p00931.

Wood, T. G. (2004), Geo-location of individual lightning discharges using impulsive VLF electromagnetic waveforms, Ph.D. thesis, Stanford University.

Yamashita, M. (1978), Propagation of tweek atmospherics, *Journal of Atmospheric and Terrestrial Physics*, *40*, 151–153.

Yamashita, M., and K. Sao (1974), Some considerations of the polarization error in direction finding of atmospherics. i - effect of the earth's magnetic field. ii - effect of the inclined electric dipole, *Journal of Atmospheric and Terrestrial Physics*, *36*, 1623–1641.