Long-range lightning geolocation using a VLF radio atmospheric waveform bank

R. K. Said,[1] U. S. Inan,[1,2] and K. L. Cummins[3]

Received 3 February 2010; revised 7 September 2010; accepted 11 September 2010; published 3 December 2010.

[1] Lightning discharges generate broadband electromagnetic pulses with a peak component in the very low frequency (VLF; 3–30 kHz) range. VLF waves propagate through the Earth-ionosphere waveguide with relatively low attenuation, enabling the detection of these radio atmospherics at great distances from the lightning discharge. A new technique of long-range (≤6000 km) global lightning geolocation via sferic detection is presented. This new technique catalogs the dominant variation in expected received waveforms in a set of waveform banks, which are then used to estimate the propagation distance and accurately determine the arrival time. Using three sensors in a trial network, this new technique is used to demonstrate a median accuracy of 1–4 km, depending on the time of day. An overall cloud-to-ground (CG) stroke detection efficiency between ~40 and 60% is estimated by correlating individual lightning stroke events to data from the National Lightning Detection Network (NLDN). Additional events reported by the trial network are shown to have a tight spatial clustering to storm clusters identified by NLDN, suggesting that many of the unmatched events correspond to weak cloud-to-ground strokes, M components, or cloud pulses. Exploiting an empirical correlation between peak VLF field strength and peak current values reported by NLDN, we also provide unvalidated estimates of the peak current and lightning channel polarity. The trial network does not distinguish between cloud and ground discharges, so these peak current estimates only relate to an Earth-referenced channel current for the subset of reported events that are return strokes.

Citation: Said, R. K., U. S. Inan, and K. L. Cummins (2010), Long-range lightning geolocation using a VLF radio atmospheric waveform bank, *J. Geophys. Res.*, *115*, D23108, doi:10.1029/2010JD013863.

1. Introduction

[2] Recent satellite-based measurements estimate that there are an average of ~45 lightning flashes per second around the world [*Christian et al.*, 2003]. This work concerns the development of a new method to detect and geolocate as many of these lightning flashes as possible using a sparse network of ground-based VLF receivers.

[3] Lightning data finds many applications in both the scientific and commercial domains. Lightning flash rate has been coupled to mesocyclone evolution [*Macgorman et al.*, 1989], storm size [*Cherna and Stansbury*, 1986], and convective rainfall rates [*Tapia et al.*, 1998; *Petersen and Rutledge*, 1998]. Lightning itself triggers many secondary processes in the upper atmosphere and magnetosphere whose study benefits from consistent high-resolution cloud-to-ground (CG) lightning data, including Lightning-induced Particle Precipitation [*Inan et al.*, 1985; *Peter and Inan*, 2007], sprites [*Inan et al.*, 1995; *Lyons*, 1996], and terrestrial gamma ray flashes [*Fishman et al.*, 1994; *Inan et al.*, 2006; *Cohen et al.*, 2010b]. With sufficient location accuracy and low reporting latency, lightning detection networks may also be used as a tool to help mitigate the effects of lightning on electric power systems [*Cummins et al.*, 1998a].

[4] Lightning is an electrical discharge that partially neutralizes charge in a cloud. The current in a lightning channel also produces electromagnetic radiation at all frequencies from a few hertz [*Burke and Jones*, 1992] through to the optical band [*Weidman and Krider*, 1986]. Lightning detection networks measure specific portions of the electromagnetic spectrum, each with an associated set of benefits and tradeoffs.

[5] Using the optical band, the Optical Transient Detector (OTD) was launched in 1995 on a low Earth orbit (740 km) satellite on a 70° inclination. Through the year 2000, the optical sensor provided ~10 km spatial resolution and ~50% detection efficiency [*Christian et al.*, 2003]. A similar sensor, named the Lightning Imaging Sensor (LIS), was launched in 1997 on a 35° inclination and is still in operation on board the Tropical Rainfall Measuring Mission (TRMM)

[1]Department of Electrical Engineering, Stanford University, Stanford, California, USA.
[2]Electrical Engineering Department, Koc University, Istanbul, Turkey.
[3]Department of Atmospheric Sciences, University of Arizona, Tucson, Arizona, USA.

Copyright 2010 by the American Geophysical Union.
0148-0227/10/2010JD013863

[*Christian et al.*, 1999]. Orbiting at a lower altitude (350 km), LIS resolves lightning discharges at a 5 km spatial resolution. While these space-based sensors cover the entire planet within the orbital path, a given location is only covered periodically for a few minutes.

[6] A ground-based 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] and provides high-resolution data, resolving simultaneous leader branches from the main charge region to the upper charge region (cloud flashes) or to the ground (CG flashes), but has a coverage area limited by line-of-sight propagation.

[7] Operating in a lower-frequency band allows detection beyond strict line of sight paths. The Earth-ionosphere waveguide provides paths for both ground-wave and reflected-wave propagation in the low-frequency (LF; 30–300 kHz) and very low frequency (VLF; 3–30 kHz) bands. The reflected impulses are known as radio atmospherics or colloquially, sferics. In the United States, the National Lightning Detection Network (NLDN) consists of over 100 sensors that measure the arrival time, arrival azimuth, or both using the VLF/LF portion of the ground wave from individual lightning strokes [*Cummins et al.*, 1998a, 1998b]. The arrival azimuth is measured using a broadband magnetic direction finding (MDF) technique developed by *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 CG 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 accuracy of $\lesssim$500 m [*Cummins et al.*, 1998a], with a CG flash detection efficiency of ~90%.

[8] 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]. By comparing geolocation results with lightning data from LIS, the daytime and nighttime flash detection efficiencies near Hawaii were estimated at 22% and 61%, respectively, and 19% and 40%, respectively, over a region just north of Hawaii. As with the NLDN, arrival time estimates were determined using an amplitude threshold mechanism. The authors provided 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.

[9] Finally, ground-based long-range lightning detection systems use the VLF portion of the electromagnetic spectrum, utilizing the particularly low attenuation through the Earth-ionosphere waveguide at these frequencies [*Davies*, 1990, p. 389]. At long ranges, even small angle errors, due to polarization-error, site-error, or signal-to-noise (SNR) limitations, can lead to large location errors. For this reason, modern long-range systems rely on using timing information to geolocate the lightning discharge, requiring at least four sensors. In this paper, we use both arrival azimuth and time measurements to geolocate discharges using only three sensors.

[10] There are two existing long-range technologies in use that geolocate lightning discharges using only arrival times of VLF sferics. The first approach, which is referred to as the Arrival Time Difference (ATD) technique [*Lee*, 1986], determines the arrival time difference between sensors by cross correlating entire sferic waveforms from sensor pairs. The discharge location is found by minimizing the sum of squared error terms between the theoretical and measured ATD values. A differential propagation phase velocity is assumed in the ATD calculation, but the velocity does not take into account propagation effects from ground conductivity or the ionospheric conductivity profile. To mitigate this limitation, *Lee* [1989] formed a composite squared error function by grouping many discharges and used the extra degrees of freedom to calculate a series of offsets to each station. A location accuracy for clusters of lightning discharges on the order of ~1–2 km was achieved. Lee's ATD technique was first put to operational use by the UK Met Office in June 1988, replacing the labor-intensive cathode ray direction finding (CRDF) system [*Lee*, 1990]. Each detected sferic above a certain threshold was recorded for a 13.1 ms time window and sent back to a central processor.

[11] A new ATD network that is currently in operation, named Zeus, consists of 10 receivers over Europe and Africa, operated by University of Connecticut and National Observatory of Athens [*Chronis and Anagnostou*, 2003, 2006]. *Chronis and Anagnostou* [2006] evaluated the location accuracy using LIS data and found a median error of ~20 km. A comparison study using the European-based medium range network LINET as ground truth found the intranetwork daytime (nighttime) detection efficiency to be ~25% (9%), with a mean location error of 6.8 km [*Lagouvardos et al.*, 2009].

[12] Since the ATD method of geolocating lightning discharges directly cross correlates sferics from different sensors, the waveforms must be "similar." If the propagation paths between the source and two different receivers are very different — for example, due to vastly different path lengths or ionospheric reflection heights — the shapes may not be similar, compromising this method of geolocation. To compensate for different propagation paths, *Lee* [1990] proposes that a propagation correction filter 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.

[13] The second long-range geolocation technology captures the arrival time on each sferic independently so that, instead of the entire waveform, only one number need be sent back to the central processor for each sferic. For the arrival time, *Dowden et al.* [2002] measure an averaged time of group arrival (TOGA) by estimating the derivative of the sferic's phase as a function of frequency using a linear fit in the frequency range 6–22 kHz. The arrival time difference is then calculated using this averaged time of group delay with a velocity corresponding to a representative group velocity in the middle of the detection frequency range.

[14] This technique for determining the arrival time forms the basis of the World Wide Lightning Location Network (WWLLN), which uses vertical electric field antennas to preferentially measure the transverse magnetic polarization. WWLLN has been periodically upgraded since its first inception in 2002 [*Rodger et al.*, 2004, 2005, 2009] and

included 40 receivers as of July 2010 (http://webflash.ess.washington.edu/). By comparing with events reported by the Los Alamos Sferic Array (LASA), [*Jacobson et al.*, 2006] report a spatial accuracy of ~15 km. [*Rodger et al.*, 2009] report an algorithm upgrade which boosts the detection efficiency to ~35% for discharges stronger than 100 kA, dropping to ~20% for 50 kA discharges, and then dropping below a 10% detection efficiency for discharges weaker than 25 kA.

[15] The methodology for long-range lightning geolocation developed in this paper utilizes an awareness of sferic waveforms at various distances. This information is used to accurately determine the arrival time and estimate the propagation distance to each sensor. *Pessi et al.* [2009] show a histogram of sensor threshold delays to examine the early components of long-range sferics. The clustering of arrival times around the ground wave and subsequent ionospheric hops is consistent with results presented here, though the dependence on distance of these features was not exploited.

[16] A generalized range estimation technique is presented in section 5 that is applicable to VLF sferics at all propagation distances. Previous range estimation methods have shown success, though usually for a subset of propagation conditions or source functions. One such method uses the time progression of sky waves to estimate the distance d and/or the effective ionospheric reflection height h using the differential time between the ground wave and subsequent sky hops [*Schonland et al.*, 1940]. If h is assumed to be constant for all ionospheric reflections, three features (ground wave and first two sky hops or three sky hops) are sufficient to determine both d and h if the hop order number n of each feature is known; four peaks are necessary if n cannot be identified [*Horner*, 1954]. *Caton and Pierce* [1952] applied a graphical approach to this technique that attempted to align the features in a measured sferic waveform to a family of ray-hop arrival time curves indexed by h. The accuracy of this ray-hop approach depends on a constant reflection height and the ability to consistently identify the onset of the sky waves, conditions that are not met on all path distances and profiles. *McDonald et al.* [1979] analyzed waveforms from a storm ~300 km from a sensor and found the inferred height to vary by 1 km or more depending on which ray-hop feature was marked as the arrival time. *Horner* [1954] recognized a transition to an "oscillatory-type" waveform, where individual ray-hop features could no longer be consistently identified, at ~2000 km at night and ~500 km during daytime conditions.

[17] Other methods for range estimation that use the frequency domain have also been proposed. *Sao and Jindoh* [1974] estimate range by applying a formula, introduced by *Wait* [1970, p. 314], that relates the delay between the impulsive VLF sferic and the lower-frequency (<1 kHz) "slow tail" to the propagation distance. *Gopalakrishna et al.* [1984] introduce a method that measures the relative group delay and spectral content between multiple frequencies, though this technique assumes single-mode propagation and so is ill-suited for short ($\lesssim$500 km daytime, $\lesssim$2000 km nighttime) propagation paths. *Rafalsky et al.* [1995] use a technique that attempts to isolate the dispersion profile of the first transverse electric (TE_1) mode. This method, which requires both an electric and magnetic field measurement, yielded errors in range estimates of $\lesssim$7% when the TE_1 mode has sufficient amplitude. Another frequency-based approach utilizes the distance-dependent oscillating pattern of the wave impedance at frequencies <50 Hz [*Llanwyn Jones and Kemp*, 1970; *Kemp and Llanwyn Jones*, 1971; *Burke and Jones*, 1995]. This technique leverages a theoretical model for the electric (E) and magnetic (H) fields excited by an impulse in a spherical cavity to iteratively fit the theoretical impedance profile E/H (parameterized by the great-circle distance to the source) to measured data. Albeit limited to strong discharges, this method is able to estimate the source-receiver range at global distances.

[18] This paper presents a long-range lightning geolocation network that improves on the location accuracy and detection efficiency of existing long-range systems, and provides peak current and polarity estimates which have not yet been incorporated (for every discharge) into any other long-range network. Section 2 reviews the geolocation methodology and establishes a framework for the single-sensor measurements extracted in subsequent sections. Section 3 reviews the data acquisition hardware deployed at each sensor and the arrival azimuth calculation applied to each sferic. Section 4 presents a survey of VLF sferic waveforms under a variety of propagation paths. This survey is leveraged in section 5 to extract consistent arrival time information under a multiplicity of path profiles. An overview of the individual sensor measurements needed to geolocate individual lightning discharges is also presented in section 5. In addition, the ability to use VLF amplitudes as a proxy for peak current is established, and a method to estimate the range to each sensor is also presented. Section 6 presents a case study of the geolocation algorithm using three sensors located in the United States, and sections 7 and 8 provide some concluding remarks.

2. Geolocation Methodology

[19] A combination of arrival time, arrival azimuth, range estimation, and amplitude is used to measure the discharge time, peak current, and location. The discharge time and location is found by minimizing a cost surface χ^2, defined as the sum of normalized squared errors between sensor measurements and predicted values based on the source parameters. With one (four-variable) term in the summation for each sensor and N total sensors,

$$\chi^2(\mathbf{r}, t, I_{\max}) = \frac{1}{4N-4}\sum_{i=1}^{N}\left[(\Delta t)_i^2+(\Delta\theta)_i^2+(\Delta A)_i^2+(\Delta R)_i^2\right], \quad (1)$$

where $\mathbf{r}$ is the discharge's coordinates on the geodesic, usually parameterized by latitude and longitude, t is the discharge time, and $I_{\max}$ is the discharge's peak current. The four terms in square brackets represent the timing, azimuth, amplitude, and range error. For example, the arrival time error term is

$$(\Delta t)_i = \frac{\frac{1}{v}\widehat{d}(\mathbf{r}_i, \mathbf{r}) - (t_i - t)}{\sigma_{t,i}}, \quad (2)$$

where $\widehat{d}$ is the distance between sensor i located at $\mathbf{r}_i$ and the discharge location $\mathbf{r}$, v is the propagation speed of the sferic, $(t_i - t)$ is the difference between the measured arrival time t_i and the discharge time t, and $\sigma_{t,i}$ is the assumed standard deviation of the arrival time measurement. Expressions for

the azimuth error $(\Delta\theta)_i$, amplitude error $(\Delta A)_i$, and range error $(\Delta R)_i$ terms take similar forms, each with its own normalization term $\sigma_{\theta,i}$, $\sigma_{\mathrm{A,i}}$, and $\sigma_{\mathrm{R,i}}$, respectively. The four terms in equation (1) are not independent; the range determination, for example, additionally depends on the arrival azimuth to determine the propagation direction. Given typical values of σ_θ and σ_{R} demonstrated in section 5, as discussed below only the first two terms are used to estimate the discharge location, and the second two terms are independently used as consistency checks on the solution. Thus in practice we assume that each error term is independent of the other three.

[20] Correctly combining sferics from different sensors in a long-range lightning geolocation network is a problem that can easily grow cumbersome due to the long baseline distances between sensors, leading to hundreds of possible combinations of time-correlated sensor reports. Using timing only, the arrival time difference $t_i - t_j$ between sensors i and j may be used to reduce the possible combinations, since the time difference should not exceed the baseline travel time. In addition to providing valuable redundancy to evaluate the merit of one combination over the rest using equation (1), the range, azimuth, and amplitude estimates can be used to further reduce computational overhead associated with evaluating different sferic combinations. Assuming each range estimate is accurate to within $\alpha \times 100\%$, the single-sensor projected discharge time window is

$$\text{discharge time window} = t_i - (1 \pm \alpha)R_i/v, \tag{3}$$

where $v \lesssim c$ is the approximate propagation speed and t_i and R_i represent the arrival time and range estimate, respectively, at sensor i. Given a reference discharge window using equation (3), a reduced set from all other sensors $\{j \neq i\}$ is chosen by applying equation (3) to each sensor j and keeping events with overlapping discharge time windows. From the resultant list of candidate events from each receiver, initial geolocation estimates are made using three arrival time measurements at a time (an O(1) operation for each estimate). The arrival azimuth error and amplitude terms in equation (1) are then evaluated to choose the best combination before the final geolocation result is computed numerically by minimizing equation (1).

[21] At long range and for a sufficiently small σ_t, the minimization of equation (1) is weakly dependent on the azimuth, range error, and amplitude terms compared to the arrival time term. As an example, consider a source-receiver distance d = 2000 km, and assume the standard deviation normalization terms σ_t, σ_θ, σ_{R}, and σ_{A} are 5 μs, 1°, 0.2d, and 3 dB, respectively; the corresponding location uncertainties from each term are then $(5\ \mu\mathrm{s})(c) = 1.5$ km, $2\pi(d)(1/360) = 34.9$ km, $0.2d$ = 400 km, and 1000 km, respectively, assuming a differential sferic attenuation rate of 3 dB/Mm at 2000 km. Since the location r and event time t are weakly dependent on amplitude and range estimates, the corresponding terms in equation (1) are not included when the geolocation is iteratively calculated by minimizing χ^2. Once this location estimate is found, the final χ^2 is computed using all terms in equation (1) to evaluate the overall consistency from all measurements.

[22] Since the location accuracy of the discharge position is heavily dependent on the arrival time accuracy and knowledge of the propagation speed v, this paper thus focuses on determining v and minimizing the timing uncertainty σ_t. The other parameters are still important, however, since they are critical in preventing sferic miscorrelation (leading to incorrect locations) and evaluating the internal consistency of a geolocation result.

3. Data Acquisition and Sferic Preparation

[23] Each sensor is equipped with two orthogonal air-core magnetic loop antennas. Using GPS-synchronized timing, data for this work were collected using desktop PCs equipped with a commercial National Instruments A/D card with 16-bit resolution, sampling each channel at 100 kSamples/s. The receiver hardware specifications are provided by *Cohen et al.* [2010a] and achieve a relatively flat gain between 800 Hz and 47 kHz. The magnitude and phase response of each channel was measured and calibrated in postprocessing.

[24] Data from four sensors were used in the sferic waveform studies in sections 4 and 5: Taylor, Indiana (TA; 40.5°N, 85.5°W); Santa Cruz, California (SC; 37.1°N, 122.2°W); Juneau, Alaska (JU; 58.6°N, 134.9°W); and Chistochina, Alaska (CH; 62.6°N, 144.6°W). The geolocation results in section 6 were obtained using only the first three sensors.

[25] The SNR is limited by the system noise and naturally and artificially generated signals like VLF transmitters [*Watt*, 1967, chapter 2] and harmonics from nearby powerlines (often visible up to 5 kHz or higher). Natural noise sources, which are incoherent and broadband, include sferics from very distant (>10,000 km) lightning discharges and, depending on the sensor location, chorus, which ranges from hundreds of hertz to 5 kHz [*Sazhin and Hayakawa*, 1992]. A least-squares method for powerline noise mitigation in geophysical records is described by *Nyman and Gaiser* [1983], *Butler and Russell* [2003], and *Saucier et al.* [2006], and a technique to isolate narrowband transmitter signals in VLF data is considered by *Shafer* [1994, chapter 3]. These techniques for reducing coherent interference sources have been applied to the data used in this paper in order to improve the SNR of sferic measurements.

[26] Individual sferic waveforms must be identified and extracted from the two high-sensitivity magnetic field channels. Each channel is first filtered between 5 and 15 kHz in order to isolate the frequency range with the greatest contribution from distant sferics. The occurrence of a received sferic is identified when the envelope of the resulting horizontal magnetic field value rises above a predetermined threshold. The specific threshold value is governed by the local noise floor and the desired sensitivity of the receiver. This 5–15 kHz filter is only applied for the purpose of sferic identification; the sferic waveforms used in this paper are taken from the full broadband data record.

[27] Most sferic waveforms from daytime paths and long-range nighttime paths have a practical waveform length of <1 ms, and, even for shorter nighttime path sferics, for the present work 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 time when the VLF amplitude crosses 50% of the peak amplitude.

[28] The azimuth, used later in section 5, is first used to digitally rotate the magnetic field data and isolate the signal along its arrival azimuth. This rotation forms the time domain signal of the received sferic that is used for the rest

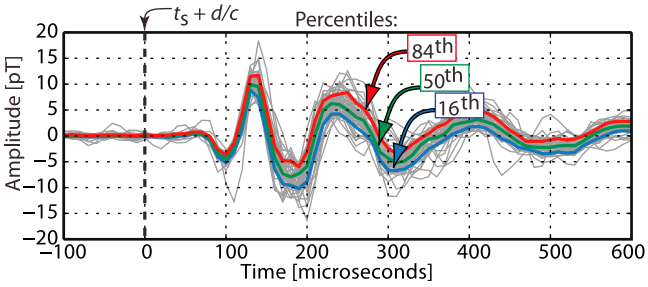


Figure 1. Process for producing an averaged waveform, illustrated using a 100% nighttime path. Each sferic is scaled in amplitude by the NLDN-reported peak current value I_{max}. Fifty recorded sferics from a storm cluster ~4500 km from Chistochina are plotted in gray, with a time axis adjusted such that zero is referenced to the NLDN-reported stroke time t_s plus the delay introduced by propagation over a distance d at the speed of light c: $t_0 = t_s + d/c$. The locus of received waveforms is used to define a canonical waveform. The three curves represent the 16th (blue), 50th (green), and 84th (red) percentile values of this collection at each time instant.

of the geolocation algorithm. The azimuth is calculated by taking the arctangent of the slope from a linear fit to the points in a time domain Lissajous parametric plot, with the north-south and east-west channel plotted on the x and y axes, respectively. Since there is no dependent variable, the best fit line minimizes the sum of the squared perpendicular distances between each point and the fitted line [*Pearson*, 1901]. To mitigate against polarization errors [*Horner*, 1954], the arrival azimuths in this work were calculated using only the leading 200 μs of the sferic.

[29] The field polarity on each magnetic loop antenna depends on the loop's winding direction and the incident quadrant of the signal. Along with the 180° arrival azimuth ambiguity intrinsic to magnetic direction finding systems, this angle-dependent polarity can be resolved with an electric field antenna [e.g., *Kemp*, 1971]. Though introduction of an electric field antenna is perfectly consistent with the methods put forth in this paper, the present results were derived using only magnetic field antennas. The quadrant-specific polarity is resolved after an initial geolocation result sets the absolute arrival azimuth and therefore incident quadrant of the sferic at each sensor.

4. Waveform Bank

[30] Short-range lightning location systems rely on the arrival time of the initial rising edge of the ground wave, which depends on the ground conductivity and the source current profile. After propagating 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 propagation path between the discharge and the receiver.

[31] It is shown here that the received waveforms from subsequent negative CG discharges conform to a tightly clustered canonical shape that depends primarily on distance and ionospheric profile. To investigate the sources of the variability in the received waveform, known CG stroke 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. Reference NLDN stroke locations in sections 5 and 6 are taken from negative CG discharges that are classified as a subsequent stroke, as defined by *Cummins et al.* [1998b]. The performance of the network with respect to other classes of discharges, including first negative and positive CG strokes and cloud discharges, is addressed in Table 1 and section 6. Data from each sensor was recorded on a synoptic schedule from 26 August through 6 September 2007. Within these recorded time periods, all NLDN-reported events, covering the continental United States and immediate surroundings [*Cummins et al.*, 1998b], were used as reference discharges.

[32] An example of this procedure is shown in Figure 1, which plots a collection of 50 sferics from the same storm that have propagated over a nighttime path. 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 50th 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 (the 'd/c' line), as the 'canonical waveform' for this specific propagation path. Figure 1 highlights a central observation of this paper that is further analyzed in section 5: for this specific source-receiver configuration and time, the 50 polarity-adjusted recorded waveforms conform to a well-defined shape.

[33] The number of sferics needed to arrive at a robust average may be determined as follows. At each time instant, the 90% confidence interval for the mean is first estimated using the sample mean and variance [*Leon-Garcia*, 1994, p. 291]. The range of values given by this confidence interval is then normalized by the peak amplitude of the canonical waveform, giving a normalized range at each sample. Results presented here were obtained by including additional sferics until the maximum of these normalized confidence intervals fell below 0.1. Empirically, we found that this threshold is usually crossed at $n \approx 50$ sferics.

[34] The geolocation algorithm in section 5 depends on a locally stored "waveform bank" that catalogs expected sferic waveforms to be seen by the sensor. The algorithm uses a normalized cross correlation between this stored waveform bank and each measured sferic. Thus the overall "shape" of the sferic waveforms is of particular interest. In the following analysis, the relative slope and delay of the zero-crossing points in the time domain waveform are used as convenient indicators of a sferic's phase and normalized amplitude profile.

[35] To minimize the number of entries needed in these waveform banks, it is useful to minimize the number of propagation parameters (beyond distance) that need to be individually considered at the sensor. To this end, Figure 2 illustrates the dependence of the received waveform on the ionospheric day/night profile and the ground conductivity. The same collection of sferics from Figure 1, taken from an

all-night path, is shown in Figure 2a (top) (median and $\pm\sim 1\sigma$ curves). Also shown, in Figure 2a (middle), are the statistical outlines for a mixed day/night path (with the same propagation distance) with the terminator bisecting the great circle path between the sensor and the storm cluster; finally, Figure 2a (bottom) plots the outlines for an all-day path. The first negative deflection is near 100 μs after the d/c line in the nighttime path. This features moves in to ~70 μs for the daytime path and is roughly halfway between these two values for the day/night mixed path. The time variation between the daytime and nighttime path is therefore much larger than the waveform spread from an isolated storm region (in this case, 4500 ± 200 km and 105 ± 5° east of north from the sensor) with a constant day/night profile. Thus the day/night profile is an important parameter for characterizing the received waveform.

[36] Figure 2b demonstrates the relative sensitivity of the phase profile, as seen in the zero-crossing progression, on the ground conductivity and, to some extent, the Earth's magnetic field configuration in comparison to the day/night profile. While the peak amplitude of the daytime sferic can reach the same magnitude as the nighttime sferic over a lower-conductivity path, the zero-crossings for the daytime path are clearly separated from those for the nighttime paths. The daytime green curve has nearly the same amplitude as the nighttime blue and red curves, especially near the onset of the sferic. However, the early zero-crossing delays after the d/c line for the three daytime curves, clustered near 90 μs, deviate by ~10 μs. The same feature in the three nighttime waveforms cluster with a similar spread near 120 μs. The later zero-crossing delays form a similar grouping between the daytime and nighttime profiles. Given the clustering of the zero-crossings for an order-of-magnitude variation of ground conductivity paths for both the all-day and all-night profiles, the day/night path is thus a more important parameter to capture the phase structure of the waveform.

[37] The influence of Earth's magnetic field is also partially lumped into these results, though an eastward-propagation path with a similar propagation distance was not available with the constellation of receivers used in this study. We expect, however, 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 and is less than the difference introduced by the ionospheric height change between a daytime and nighttime ionosphere. We base this conclusion on work by *Wait and Spies* [1964], who used numerical methods to evaluate the phase of the reflection coefficient off of an exponential anisotropic ionosphere. With an 84° vertical incidence angle (grazing incidence; $\cos\theta = .1$) at $f = 10$ kHz, the calculated phase difference between east-to-west and west-to-east propagation of the ionospheric reflection coefficient was $\lesssim 5°$. This dependence affects the phase measured on the ground much less than the reflection height. For the same propagation distance, a differential increase of height δx adds $\delta d \approx 2\cos\theta\delta x$ to the first ionospheric hop path length. At $\cos\theta = .1$, for an increase in ionospheric height from 70 km for daytime [*McRae and Thomson*, 2000] to 85 km for nighttime [*Thomson et al.*, 2007], the added phase of the first ionospheric reflection at 10 kHz ($\lambda \simeq 30$ km) is 36°.

[38] Given these findings, we conclude that one may capture the dominant variation in waveform shape with canonical waveform banks indexed only by the day/night profile and distance. Given these two parameters, a sferic from a 'typical' lightning discharge, defined for this paper to be a subsequent negative CG stroke, will have a good correlation with 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 Earth's magnetic field may delay and attenuate the waveform with respect to a canonical shape but have a smaller effect on the phase structure and therefore zero-crossing progression of the waveform.

[39] An empirically derived waveform bank is shown in Figure 3, where each waveform is taken from the median line as defined in Figure 1. Waveforms were derived from sferics measured at Taylor, Juneau, and Chistochina. At each distance, preference is given to high-conductivity paths to minimize distinct features from one distance to the next. A number of features may be readily identified in the waveform bank: the ground wave dies out beyond ~1500 km (peak closest to d/c line), and the sky waves move toward the d/c line as the ionospheric grazing angles become more shallow (at greater distances). 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 due to the lower attenuation from each reflection at night.

5. Sferic Parameters

[40] The complications in determining a consistent arrival time imposed by a propagation path-dependent waveform structure are managed by leveraging the waveform banks from Figure 3. These waveform banks are used to accurately measure arrival time by enabling a reliable identification of high-slope regions of the sferic, such as the onset of the ground wave at short distances and early zero-crossing delays associated with a minimum number of ionospheric reflections. The waveform banks are also used to derive a range estimate at each sensor, which is used at the central processor to help resolve the polarity estimate. After the methodology for determining the arrival time and range estimate is detailed, this section establishes the performance of other single-sensor measurements, including arrival azimuth, polarity estimation, and amplitude.

[41] The nonuniform spread as a function of time in Figure 1 suggests that some sferic features provide a more reliable timing measurement than others. Figure 4 further illustrates this point. 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 low-time-variance feature to measure a consistent arrival time. These plots suggest that one should use an early portion of the sferic, where the slope (and therefore SNR) is high, to target a low time variance feature of a sferic. For sferic waveforms that have propagated a short distance, the arrival time is taken from the rising edge of the ground wave. For longer propagation paths, where the ground wave signal is

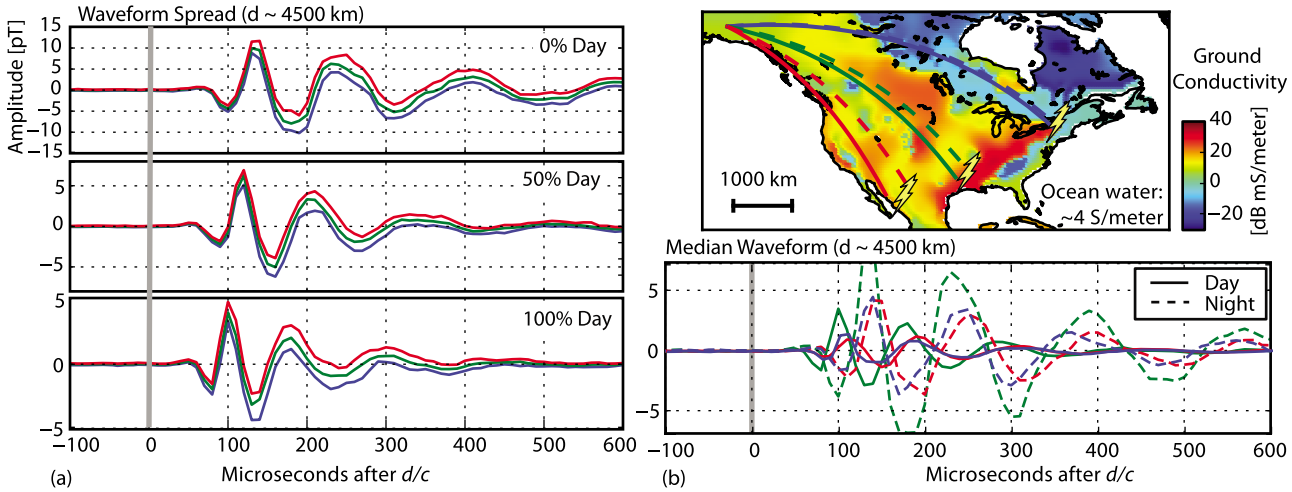


Figure 2. Waveform dependence on the ionospheric and ground profile. (a) Waveform averages (median and $\pm 1\sigma$ approximation) for three different day/night path profile parameters, using 50 sferics in each average. (b) (bottom) Six canonical waveforms from different path profiles. Daytime (nighttime) propagation paths and corresponding waveforms are shown in solid (dotted) lines. (top) Conductivity map corresponds to the upper VLF range (10–30 kHz, from *Morgan* [1968]).

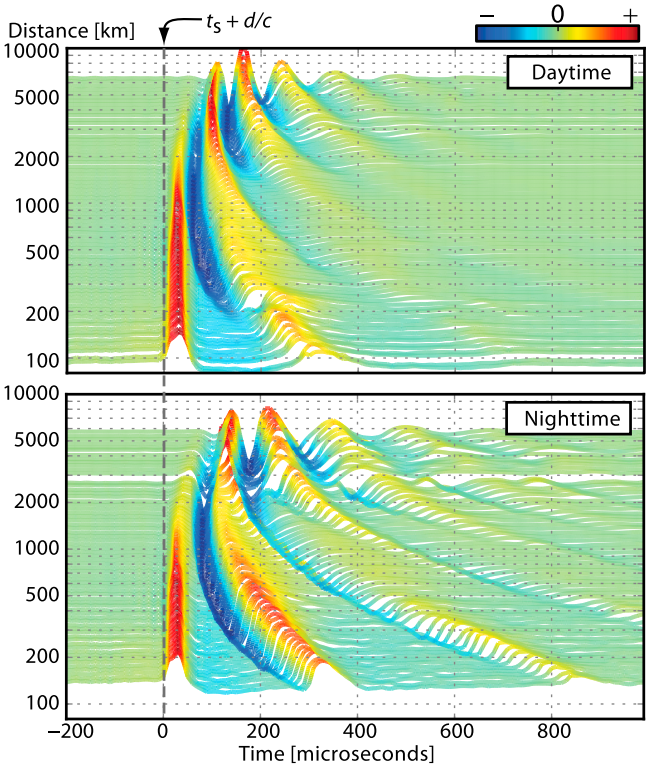


Figure 3. Complete day and night waveform banks. Each row indicates a specific waveform bank entry. Each waveform has been normalized so that the maximum amplitude is unity. The time axis (abscissa) is referenced with 0 corresponding to a speed-of-light propagation; the distance index, on the ordinate, is plotted on a log scale to best capture the rate of feature changes.

heavily attenuated with respect to the ionospheric hops, the arrival time is measured from an early zero-crossing point.

[42] While the above procedure may produce consistent arrival time measurements for a particular propagation path profile, knowledge of the relative propagation speed associated with the chosen feature of the sferic is still necessary in equation (2). If the received sferic is properly aligned to an entry of a locally stored waveform bank, the arrival time can instead be referenced to the d/c line, negating any need to evaluate the (frequency-dependent) propagation speed. The central processor may then set $v = c$ for all propagation paths. Thus if the sensor has a locally stored waveform bank entry that matches the propagation path of the sferic, then an initial estimate of the d/c point may be approximated by calculating the offset of the peak cross correlation between the measured sferic and the waveform bank entry.

[43] As seen in section 4, the zero-crossing structure is heavily dependent on the day/night profile and propagation

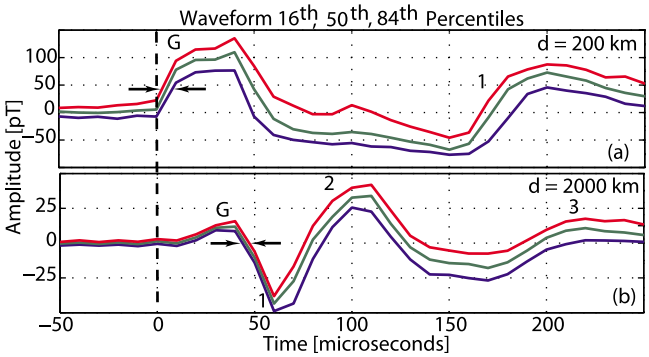


Figure 4. Expanded view of the initial portion of the measured daytime canonical waveform distribution at (a) 200 km and (b) 2000 km.

distance. The sensor uses the time of day and measured arrival azimuth of the sferic to determine the day/night profile along the arrival direction for multiple distances. If there is a 180° azimuth ambiguity, the day/night profile is projected in both possible directions. The sensor stores a daytime, nighttime, and ideally a series of mixed day/night banks that can be loaded into memory as needed, and a custom waveform bank is selected on the fly for each incoming sferic. In this work, only an all-day and all-night waveform bank was used.

[44] The waveform bank entries are all normalized to a specific polarity of the causative CG stroke, so the polarity of the peak magnitude of the cross correlation depends on the polarity of the source discharge that generated the sferic. The sferic in question is cross correlated with each entry in the custom waveform bank, and the peak cross-correlation value and delay is noted for each entry. Ideally, the 'correct' polarity yields the largest cross-correlation peak. For example, a sferic from a negative CG stroke cross correlated with a waveform bank derived from negative CG strokes would ideally yield a maximum positive cross correlation that is larger than the minimum negative correlation. Dispersion and attenuation suffered by the sferic in the Earth-ionosphere waveguide results in a relatively narrowband signal at large distances (the so-called oscillatory shape), which in turn gives rise to an oscillatory cross-correlation signal that has a maximum and minimum of approximately the same magnitude. The global maximum and minimum across all waveform bank entries, corresponding to the 'best' correlation for each assumed polarity, may therefore have approximately the same amplitude, leading to a polarity ambiguity at the sensor. Each sensor therefore records the zero-crossing delay and range estimate, described below, for both the maximum and minimum waveform bank cross correlation. The correct polarity is resolved at the central processor.

[45] The cross-correlation offset corresponding to the peak correlation (for each polarity) estimates the d/c line using the entire waveform, and therefore does not take advantage of the lower time variance features toward the beginning of the sferic. For discharges closer than a predefined distance d_0, we presently identify the arrival time at the point the leading edge of the ground wave crosses 50% of the peak sferic magnitude. The distance d_0 is chosen so that most of the sferics with propagation distances $<d_0$ 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. From empirical studies over the United States, d_0 may be set between 800 and 1000 km.

[46] For discharges farther than d_0, an early zero-crossing time of a measured sferic is used as the arrival time and is measured using the waveform bank cross correlation as follows. For each waveform bank entry, the delay after the d/c line of the first zero-crossing after the amplitude rises higher than 25% of the maximum value is identified. After the cross-correlation step described above, the measured sferic is time-aligned to the waveform bank entry that gave either the maximum or minimum cross correlation with an offset determined by the peak magnitude of the cross correlation to that entry. The zero-crossing is then identified as the zero-crossing in the sferic that is nearest to the corresponding precomputed '25%' zero-crossing in the chosen waveform bank entry.

[47] 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 early portion of the waveform. Extracting timing information from the early portion of the sferic may also help guard against small variability in the path profile since the early features have reflected fewer times off of the ionosphere.

[48] 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 uses 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 the method described here extends this ability to arbitrary differential sferic paths and reduces the required communications bandwidth.

[49] Along with the 50% threshold delay value, the absolute, GPS-referenced time of the zero-crossing for each polarity is sent back to the central processor. These timing estimates need to be referenced back to the d/c line to recover the speed-of-light assumption in the geolocation algorithm. Ideally, one would have an empirically or theoretically derived database for the delay after the d/c line of the threshold- and zero-crossing delays for each region, season, ionospheric profile, and source type. With an existing medium range network such as NLDN, reference source locations may be used to build up such an empirical database over the covered region (results using such a database are shown in section 6). Without such a reference network, it is nevertheless advantageous to use an approximate, distance-dependent correction factor. While timing deviations due to the nuances of each propagation path would remain, the mean offset of the relevant feature from the d/c line is removed.

[50] To derive good approximate correction factors, the same empirical NLDN-referenced database that was used to derive the waveform bank is again used, but now with an eye toward the low-variance timing features. Figure 5 characterizes the early zero-crossing delays under a daytime ionosphere. Using NLDN-referenced waveforms, the first zero-crossing value after the initial rise above 25% of the peak amplitude is clustered at discrete distances. By comparing with Figure 3, the first level zero-crossing, "L1", is the first zero-crossing after the ground wave, and has a negative slope for a negative CG discharge. The L2 zero-crossing is triggered by the first ionospheric reflection, after the ground wave attenuates below the 25% level from the peak amplitude. Similarly, the L3 zero-crossing is triggered after the first ionospheric hop attenuates below the 25% level.

[51] The zero-crossing averages represented in Figure 5 are taken from a multitude of paths. The relative consistency of the median zero-crossing levels between overlapping distances covered by two sensors, coupled with the low estimated standard deviation, further highlights the insensitivity of the early zero-crossing delay with the path profile. A reasonable correction factor for a wide variety of paths may therefore be calculated by using these zero-level

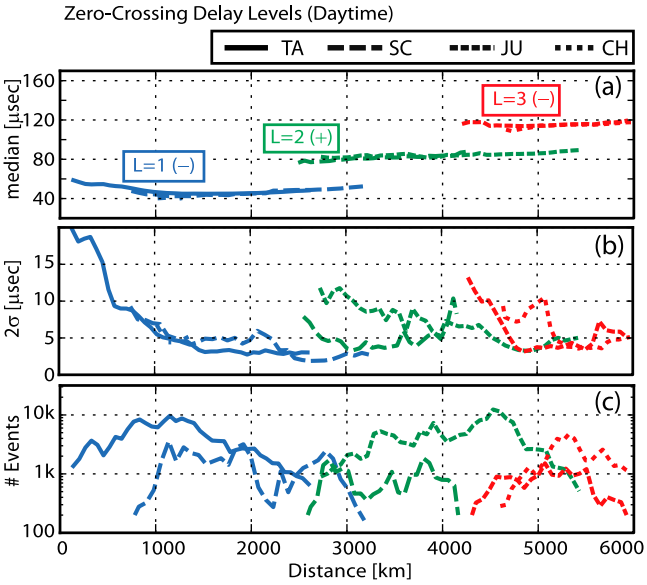


Figure 5. Statistical characterization of daytime zero-crossing delays versus distance from four sensors. (a) Median zero-crossing delays of the first three levels, (b) difference between the 84th and 16th zero-crossing delay in each distance bin, and (c) number of events contributing to each measurement. The polarity of the slope for a negative CG-referenced waveform is paired with each zero-crossing level label.

delays, with the same day/night profile as the sferic in question. Using a second-order regression curve fitted to the zero-crossing delays for each level, a reasonable correction factor may thus be computed by indexing the correct regression curve for the reported level at the approximate propagation distance. The large timing deviations for distances below ~800 km in Figure 5b illustrates why the ground wave threshold crossing is used for short propagation distances.

[52] For both polarities, the zero-crossing level of the chosen waveform bank entry is also transmitted back to the central processor. After the approximate propagation distance is determined from an initial geolocation estimate, the central processor then subtracts the appropriate delay based either on the threshold-crossing point or the appropriate zero-crossing level and then recomputes the geolocation result.

[53] As an alternative, another, more precise timing adjustment involves using an empirical grid-based correction factor, as mentioned above. Figure 8c shows the empirical delay factors at Juneau (JU) for the second zero-crossing level using sferics from NLDN-detected lightning discharges.

[54] Figure 6a plots the delay profile and uncertainty versus distance of the 50% threshold delay, again referenced to the d/c line. For distances ~100–1000 km, 68% of the threshold delays measured at Taylor (TA) were within 5 μs of the median value at each distance, where the median itself varied from 5 μs at 100 km to 20 μs at 1000 km. The threshold crossing value therefore provides a consistent arrival time measurement at short range, where the L1 zero-crossing delay is less consistent (Figure 5b). At slightly larger distances where the ground wave no longer reliably triggers the 50% threshold-crossing delay, the more consistent L1 zero-crossing delay is used.

[55] Figures 6b and 6c characterize the remaining sensor measurements used by the central processor. At distances

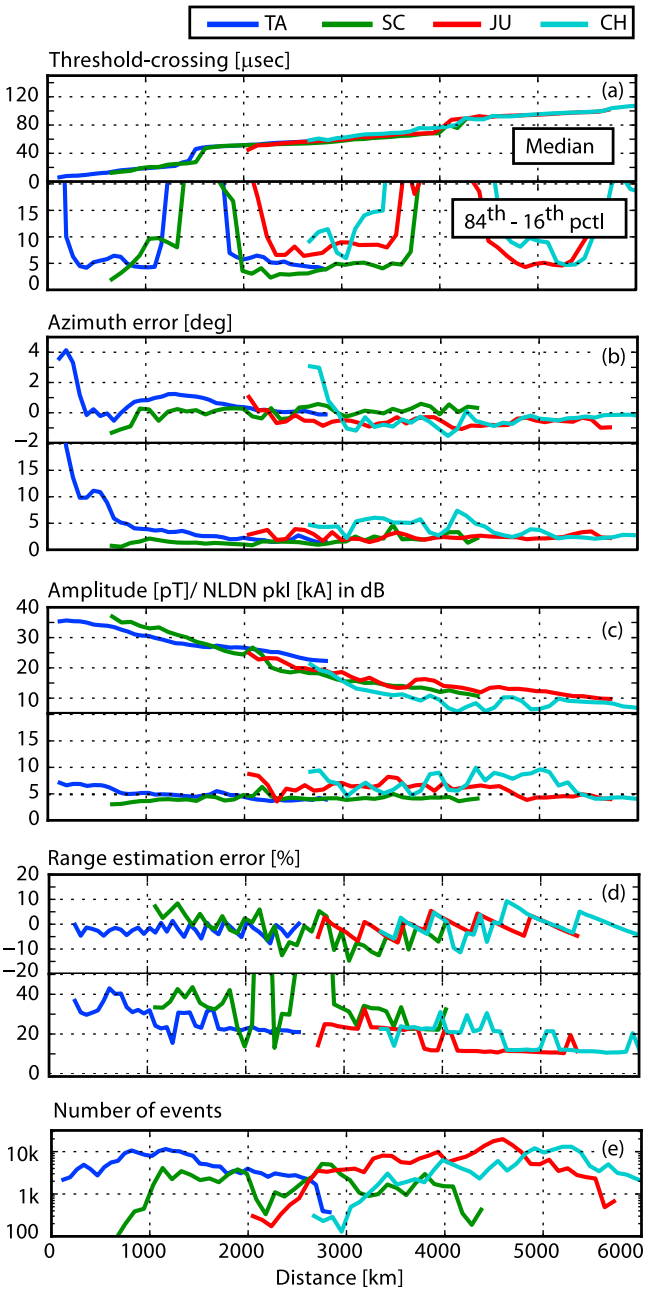


Figure 6. Median and 16th to 84th percentile range versus distance for four single-sensor parameter measurements using NLDN-referenced waveform data. All sferic data taken from subsequent strokes in a negative multistroke CG flash. (a) Threshold-crossing delay after the d/c line. (b) Arrival azimuth error, after site-error correction. (c) VLF peak amplitude normalized to a 1 kA peak current stroke, as reported by NLDN. (d) Range estimation error using a 40-element daytime waveform bank with entries logarithmically spaced from 100 to 6000 km. (e) Number of events contributing to each measurement.

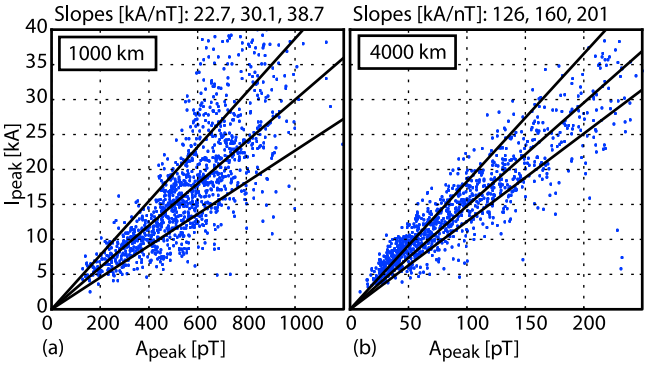


Figure 7. Scatterplots showing the relationship between the NLDN-reported peak current I_{peak} and the peak VLF amplitude A_{peak} for a daytime ionosphere at two distance intervals: (a) 1000–1100 km and (b) 4000–4300 km, showing 1 out of 5 samples (evenly sampled by occurrence time) for plot clarity. The slopes for the average and '$\pm\sigma$' lines (see text) are given above each plot. Receiver saturation is evident for large peak current (>30 kA) strokes in Figure 7a, indicating that the error estimate in Figure 6c for nearby discharges, which incorporates these more intense strokes, overestimates the peak current error range for smaller discharges. The NLDN-referenced sferics used to generate Figure 7b were taken from a 10° azimuth range from Juneau, resulting in a tighter clustering than indicated by Figure 6c at the same distance.

>1000 km, assuming a Gaussian error distribution, the standard deviation of the arrival azimuth is $\lesssim$2° (Figure 6b). The larger azimuth range at close range is due to signal saturation at the receiver (>0.7 nT), as seen in Figure 8b. The jump in the median arrival azimuth error at Chistochina (CH) at 2500 km (Figure 6b) is confined to arrival azimuths near 130°, possibly a result of higher-order harmonics in the site error or polarization errors due to propagation over many land/sea boundaries along the Pacific Coast.

[56] Figure 6c demonstrates the close correspondence between the NLDN-reported peak current and the peak VLF amplitude. The NLDN peak current is itself an empirically calibrated conversion between measured field waveforms from VLF/LF sensors and the peak current measured at the base of a rocket-triggered lightning channel [*Cummins et al.*, 1998b; *Jerauld et al.*, 2005]. Because of the similar radiation fields produced by triggered and subsequent return strokes [*Le Vine et al.*, 1989], this calibration is thought only to apply to subsequent strokes in a negative CG flash. In particular, the peak current estimate is not validated for first return strokes (those preceded by a stepped leader), and the utility of this parameter further breaks down in a cloud discharge, for a variety of geometrical reasons. The estimated peak current values reported by the trial network in section 6 are therefore understood to be validated only in subsequent negative CG strokes, an important caveat since the trial network does not identify different discharge types. These values are only informative in other classes of discharges insofar as the radiation field amplitude, estimated using the propagation-adjusted VLF amplitude, serves as a proxy for the intensity of the discharge in question. Nonetheless, the consistency of the VLF amplitude at a given distance with the NLDN estimated peak current also allows for another consistency check at the central processor through equation (1) and permits any application that relies on NLDN-reported peak current to gain a similar utility using the network described in this paper.

[57] Figure 6c (bottom) shows the estimated difference between the $\pm\sigma$ values for the ratio between the VLF amplitude A_{peak} and the NLDN-reported peak current I_{peak} at each distance interval, measured in dB. Figure 7 illustrates this bound at two distances by showing a scatterplot of I_{peak} versus A_{peak}. In each case, the average and $\pm\sigma$ values for the set $\{A_{peak}/I_{peak}\}$ are estimated, as in Figure 6c. The inverse of these measurements gives the expected linear slope relating A_{peak} and I_{peak} at the specified distance; the corresponding lines for the median and $\pm\sigma$ estimates are drawn on each plot. The ratio between the two bounding slopes at 1000 km is 38.7/22.7 = 1.69 = 4.6 dB. Thus a 4.6 dB '2σ' estimate in Figure 6c indicates that ~2/3 of all the discharges in that distance bin follow an average linear relationship between A_{peak} and I_{peak} within a factor of 1.69 of each other.

[58] When each measured sferic is cross correlated with a locally stored waveform bank in order to roughly determine the *d*/*c* point and allow for the correct zero-crossing level to be determined, the highest-peak cross correlation across all entries in the waveform bank also yields a range estimate, taken here to be the distance index of the waveform that yielded the highest cross correlation. Figure 6d shows the median and range uncertainty of the range estimate using this method. The sawtooth pattern in the median range error is due to the discretized waveform bank; a more sparsely populated waveform bank would yield larger swings and increase the range estimation error. These results suggest an overall RMS range estimation accuracy of ~20%; in the results that follow, the range uncertainty is therefore set at $\sigma_R = 0.2R$. The higher range error at Santa Cruz (SC) between 2000–3000 km (>50% 2σ range) may be due to higher attenuation compared to similar propagation distances measured at Taylor (TA) for ionospheric hops greater than 1, resulting in an overestimation of the propagation distance. The potential "full-cycle" errors that may result from this overestimation are addressed in section 5.2.

[59] Many of the deviations from the minimum value across the distance range in Figure 6 are resolved by plotting a given parameter against a different dependent variable, such as the received peak VLF amplitude or the originating geographic region. Figures 8a and 8b plot the threshold delay and azimuth range at Taylor against distance and the peak VLF amplitude. The range in both the threshold delay and, to a larger extent, the arrival azimuth error rises sharply with increasing VLF amplitude due to saturation at the high-gain antenna. In these cases, the central processor should appropriately modify the weighting factor for the relevant parameter in equation (1). At short distances, a larger σ_θ has less of an effect on the overall performance since the sensitivity of the geolocation result on azimuth is linearly proportional to distance and there is less absolute distance uncertainty in the range estimate.

5.1. Polarity Estimation

[60] Each assumed polarity yields its own best correlation with the waveform bank, which also sets the estimated propagation distance. The central processor must use this information from multiple sensors to estimate the source

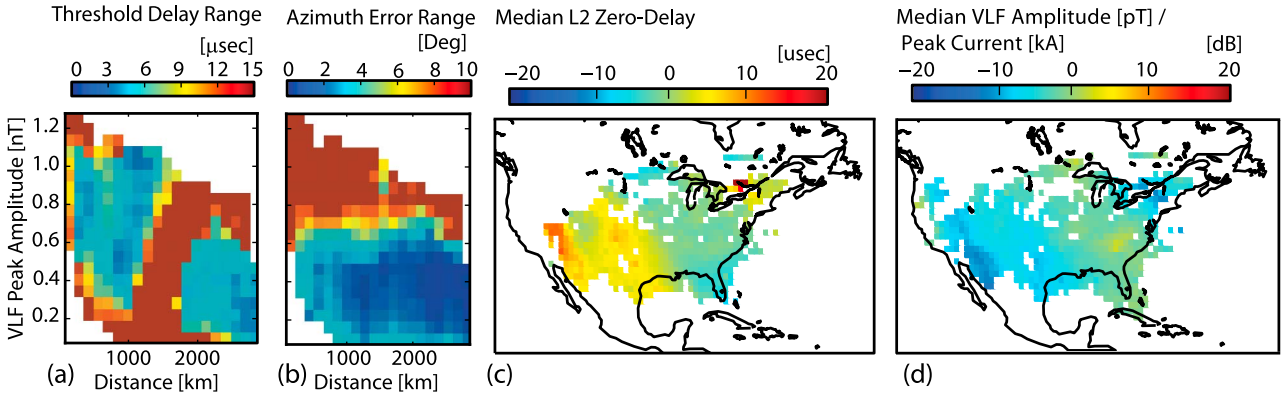


Figure 8. The 16th to 84th percentile range for the (a) threshold delay and (b) arrival azimuth error at Taylor versus distance and peak VLF amplitude, demonstrating performance dependence on both distance and signal amplitude. (c) Deviation of the daytime median L2 delay from fitted regression curves, plotted in 1° × 1° bins over the United States. The L2 delay is only plotted over distance ranges where the ground wave is less than 25% of the maximum peak. (d) Deviation of the median daytime conversion factor between peak current and VLF amplitude from the average value obtained using equation (5).

polarity. In this context, the polarity is defined with respect to the waveform bank, which in turn is referenced to negative CG discharges. A negative polarity is assigned if the field deflection corresponding to the ground wave is positive; that is, the electric field is directed toward the Earth (atmospheric electricity sign convention). For CG discharges, this convention assigns a negative polarity to events that effectively lower negative charge to the ground. A polarity based on the received field is assigned to each event. Since the network does not distinguish between cloud and ground discharges, for the subset of reported events that happen to be cloud pulses, the assigned polarity does not relate to an Earth-referenced channel current as it does for CG discharges.

[61] Figure 9 shows the percentage of NLDN-referenced sferics that, when cross correlated with the daytime waveform bank in Figure 3, yielded a higher (normalized) cross correlation with the correct polarity (dotted lines). The dip near 2500 km with Santa Cruz (SC) data is likely due to a larger attenuation for subsequent hops over the Rocky Mountains, coupled with the fact that the corresponding entries in the waveform bank were derived using data from Taylor. Similarly, the dip near 4000 km at Juneau (JU) and Chistochina (CH) is a result of the higher attenuation seen in Figure 8d over several paths at this distance.

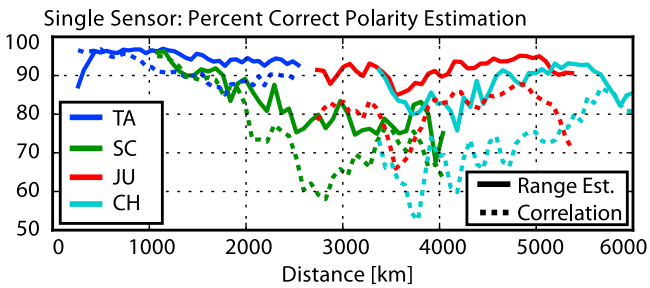


Figure 9. Percentage of sferics that reported a correct polarity based on the range estimation error (solid line) and the peak correlation coefficient (dashed line).

[62] The solid curves in Figure 9 show, for the same data set, the percentage of sferics that had a lower range estimation error associated with the correct polarity. The higher value for these curves suggests a robust method for choosing the polarity at the central processor. Once an initial discharge location is estimated, the accuracy of the range estimate for each polarity may also be determined. Evaluating the range contribution to the χ^2 in equation (1) for both polarities, the polarity which yields the lowest overall chi-squared value is chosen. When necessary, the appropriate zero-crossing level is then selected.

[63] Figure 9 used NLDN-reported negative subsequent CG strokes propagating under a daytime ionosphere. Table 1 summarizes the percentage of discharges that yielded a correct polarity using the range estimation for a variety of source types and propagation conditions, again using NLDN data as a reference for discharge coordinates and source type classification. The polarity estimation improves under a nighttime ionosphere, as the lower attenuation presents a more sustained zero-crossing profile that aids in the range estimation used for the polarity determination. First negative strokes (SO = 1) have a poorer performance, partly due to the fact that the waveform bank is tuned to subsequent strokes. First positive and cloud discharges have a poorer performance still, likely due to a more irregular source functions for these discharge types.

Table 1. Percentage of NLDN-Correlated Discharges That Yield a Correct Single-Sensor Polarity Estimation Based on the Range Estimation Error of Each Polarity

	TA	SC	JU	CH
−CG, SO > 1, Day[a]	95	85	92	89
−CG, SO > 1, Night	97	94	95	93
−CG, SO = 1, Day	86	77	86	84
+CG, Day	70	71	69	70
Cloud, Day	76	64	70	58

[a]Stroke order (SO) refers to the order of a stroke in a multistroke negative CG flash (so discharges with SO > 1 are preceded by a dart leader that tends to propagate in an existing discharge channel). TA, Taylor, Indiana; SC, Santa Cruz, California; JU, Juneau, Alaska; and CH, Chistochina, Alaska.

5.2. Full-Cycle Errors

[64] The above method corrects for the sometimes high percentage of discharges that give a higher cross correlation with the opposite polarity (see Figure 9), though the occurrence rate for an incorrect single-sensor polarity estimate is kept to $\lesssim$25% by using the range estimate. Since the zero-crossing levels are ~30 μs apart, such polarity errors introduce distance errors on the order of 10 km. A smaller subset of sferics that must still be considered are those that yield a peak cross correlation with an offset greater than one adjacent zero-crossing level. For example, if a sferic that has propagated 4000 km generates a maximum correct-polarity cross correlation with a waveform bank entry corresponding to 2000 km, the reported zero-crossing level will be 1. It is unlikely that the attenuation is sufficiently low such that the ground wave perturbs from the noise to a degree where the true first zero-crossing is selected from the waveform. A more likely scenario is that the actual zero-crossing level selected is 3, just after the second ionospheric hop. Even though the selected polarity is correct, the reported zero-crossing level is off by 2 levels. As a result, the corrected delay will be off by ~60 μs.

[65] Figure 10 empirically quantifies the occurrence rate of these "full-cycle" offset errors for daytime sferics. Each panel plots the distance distribution of a true zero-crossing level, as determined by the delay after the d/c line using NLDN stroke data, for a specific reported level $L = L_R$. The events in Figure 10 (top), for example, all reported a peak cross correlation of the correct polarity with a waveform bank entry whose ground wave was at least 25% of the peak amplitude (i.e., $L = 1$). At Taylor, the correct zero-crossing was chosen for ~99% of the events. The yellow curve, labeled $L = 0$, represents delays <35 μs, i.e., a zero-crossing before the ground wave. The red curve represents delays near the third zero-crossing level; physically, the zero-crossing after the second ionospheric hop. At distances >3500 km, where the ground wave is severely attenuated in the noise, most of the sferics with a reported zero-crossing level of 1 (associated with the Juneau site) actually correspond to the third level. It should also be emphasized that these mismatched $L = 1$ events represent only a few percent of events at Juneau. Comparing the Juneau (JU) curves in Figure 10 (top and middle), most of the events correctly correlate to the $L = 2$ level. Figure 10 (middle and bottom) shows a similar pattern. For most reported $L = 3$ events closer than 4500 km (Figure 10, bottom), for example, the reported zero-crossing delay corresponds to the first level. Once the propagation distance to the sensor is estimated at the central processor, the correct delay offset can be recovered due to the relatively nonoverlapping distance distribution of the same-polarity zero level curves.

5.3. Peak Current

[66] The peak current estimate follows an analogous procedure to the arrival time estimate whereby the measurement is referred to a reference value based on the propagation distance and profile. The peak amplitude measurement corresponds to the peak of the ground wave and, beyond a certain distance, transitions to correspond to the peak of the first ionospheric reflection and then to subsequent peaks in the waveform. It was shown in Figure 6c that the peak amplitude, while derived from various features depending on distance, nevertheless varies monotonically with the NLDN-reported peak current at each distance.

[67] The NLDN calculates peak current assuming the transmission line model [*Cummins et al.*, 1998b], which predicts that the peak current will be linearly proportional to the peak in the radiation field [*Uman*, 2001, p. 333]. In order to convert the measured amplitude (SS) to an estimate of the peak current I_{peak}, the NLDN uses a distance-(d-)dependent correction factor,

$$I_{peak} = (SS)(C)\left(\frac{d}{I}\right)\exp\left(\frac{d-I}{A}\right). \quad (4)$$

C is a proportionality factor that converts raw measured units to kA and A (in km) gives the e-folding distance, or the attenuation rate due to losses. I gives the normalization distance (in km) at which the conversion C is calculated. In the NLDN, A is set to 1100 km.

[68] At large distances, instead of an inverse square power fall-off, to first order one would expect power to fall of with the inverse of distance due to energy guiding in the Earth-ionosphere waveguide. For the equivalent long-range normalization factor, we use following conversion factor:

$$I_{peak} = (SS)(C)\left(\frac{d}{I}\right)^{\frac{1}{2}}\left[\frac{\sin(d/R_0)}{(d/R_0)}\right]^{\frac{1}{2}}\exp\left(\frac{d-I}{A}\right), \quad (5)$$

where R_0 is Earth's radius. The factor in square brackets accounts for the focusing effect of the spherical waveguide. Except for an angle-dependent transmitter gain factor and the normalization at distance I, the inverse of equation (5) matches the functional dependence on distance of the attenuation at a single mode and frequency for the horizontal magnetic field given by *Wait* [1960]. At short distances, equation (5) does not reflect the expected $1/d$ amplitude dependence of the ground wave, and so a separate normalization factor may be necessary for distances $\lesssim$500 km. Over the United States, peak current estimates given by equation (5) using Taylor data ranged from 0.55 (16th percentile) to 0.97 (84th percentile) of the NLDN-reported peak current value (for subsequent negative strokes under a daytime ionosphere, sample size 132,916 events). The resulting range 0.97/0.55 = 4.9 dB agrees well with the amplitude correlation range estimate given in Figure 6c. The tendency to underestimate the peak current, even with the distance dependence in equation (5), was due in part to saturation at that sensor.

[69] Using the averaged curves in Figure 6c and a corresponding nighttime plot, we found an empirical conversion factor of $C = 5.0 \times 10^{-3}$ kA/pT (with $I = 100$ km) and a daytime (nighttime) e-folding distance of 2820 (5640) km. As expected, the e-folding distance is shorter under a daytime ionosphere as compared to a nighttime ionosphere.

6. Results

[70] This section analyzes geolocation results using data from Taylor, Santa Cruz, and Juneau. The baseline distances between these sensors are 2559 km (JU–SC), 3981 km (JU–TA), and 3187 km (SC–TA). Synoptic data were recorded for 1 min out of every 15 at each station. Due to mismatched down times, synoptic data was available

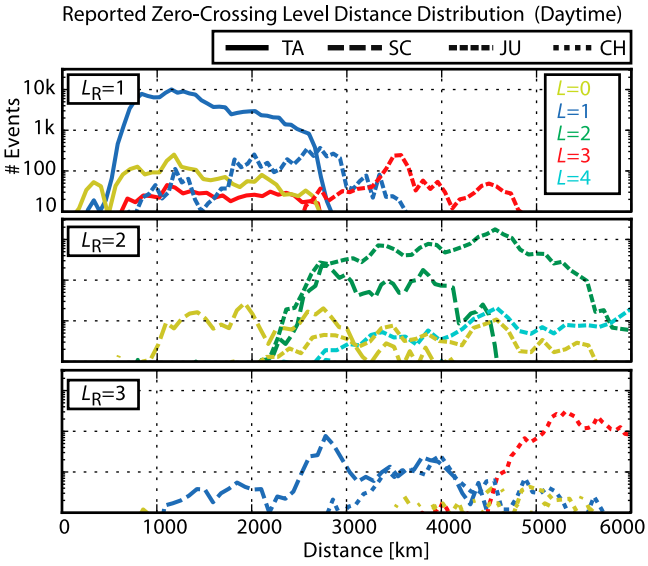


Figure 10. Distance distribution for three reported zero-crossing levels. For each reported level L_R, the distance distribution of each 'true' zero-crossing level is shown in a different color. True zero-crossing levels are determined by the delay after the d/c line, using NLDN stroke data as a reference. In each panel, data from two sensors are chosen as representative cases; not all data from each sensor are shown. The $L = 0$ level is defined as any zero-crossing delay less than 30 μs.

from 2000–2400 UT on 5 September and 0000–0700 UT, 1000–1600 UT, and 2000–2400 UT on 6 September 2007, which covers several day/night profiles to each sensor. To allow for easy labeling and to contrast with the NLDN stroke data used as a reference, this section refers to this trial network as the Stanford Lightning Detection Network (SLDN).

[71] A loose range accuracy factor of $\alpha = 0.9$ was used in equation (3) to ensure that a large number of candidate sensor combinations were considered. The following σ values were used in calculating χ^2 in equation (1):

$$\sigma_R = 0.2R$$

$$\sigma_t = 5\ \mu\text{s}$$

$$\sigma_\theta = 3^\circ.$$

[72] To account for the increased azimuth uncertainty at high VLF amplitudes at close range seen in Figure 8b, σ_θ was increased linearly from 3° at 1000 km to 10° at 100 km.

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

$$\chi^2_R : 2.0$$

$$\chi^2_t : 1.0$$

$$\chi^2_\theta : 1.0.$$

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 events from other source classes that may not yield an accurate cross-correlation result.

[74] Sections 6.1 and 6.2 investigate the detection efficiency and location accuracy by comparing SLDN to NLDN stroke location results. For this comparison, an individual event in the SLDN data set is considered matched if it satisfies a simultaneous temporal and spatial coincidence of <180 μs and <60 km, respectively to an NLDN ground stroke. Both data sets were limited to the latitude range [25, 55] and the longitude range [−120, −75] so that the statistics are not skewed due to the lack of NLDN's coverage far beyond the continental United States and Canada and to minimize the geometric effects of SLDN. The NLDN data set is also restricted to the synoptic recording schedule of the three sensors in the network. Using these criteria, out of a total of 28,685 SLDN-located events, 11,383 (40%) matched to NLDN strokes.

[75] Figure 11 demonstrates the spatio-temporal relationship between the matched and unmatched strokes with respect to the NLDN data set. The matched data set exhibits a tight clustering to the NLDN ground strokes with 99% of events occurring within 20 km of an NLDN first stroke, and a median location error of <2 km. The step in the "matched" time difference CDF (subplot b) corresponds to events matched to NLDN first strokes. Matched events after the first stroke are matched to subsequent strokes of negative CG flashes.

[76] The large number of unmatched strokes is not randomly distributed with respect to the NLDN data set. Ninety percent of all unmatched SLDN events occur within 60 s and 80 km of an NLDN CG flash (Figure 11a), indicating the events are coincident with storms reported by NLDN. For unmatched events occurring within 7 km and 1 s from an NLDN first stroke, over 70% occur after the NLDN stroke and over 75% have a small negative reported polarity (Figures 11b and 11c, green lines), which is consistent with subsequent negative CG strokes or large M components [*Kitagawa et al.*, 1962] that NLDN may have missed.

[77] Events whose location is greater than 7 km from an NLDN flash have a storm-level spatiotemporal correlation with the NLDN-reported discharges but no temporal correlation on the time scale of a flash. The SLDN-reported peak current for these events provide insight on their source types. By comparing video records to weak NLDN-reported discharges, *Biagi et al.* [2007] found that ~87% of small ($|I_p| < 10$ kA) negative discharges were correctly identified as CG strokes, and less than 7% of small positive NLDN reports ($|I_p| < 10$ kA) were CG strokes. Nearly 70% of unmatched events >7 km from an NLDN flash have an SLDN-reported weak positive polarity. If we assume that the SLDN-reported peak polarity has a similar bias based on source type, the larger percentage of positive unmatched discharges in this category suggests that many of these events are large pulses produced by in-cloud processes (subpanel c, red lines). The lower range accuracy exhibited by the higher skew in the range contribution to χ^2 (Figure 11d) for these reports further suggests a different category of discharges that do not correlate well with the waveform bank,

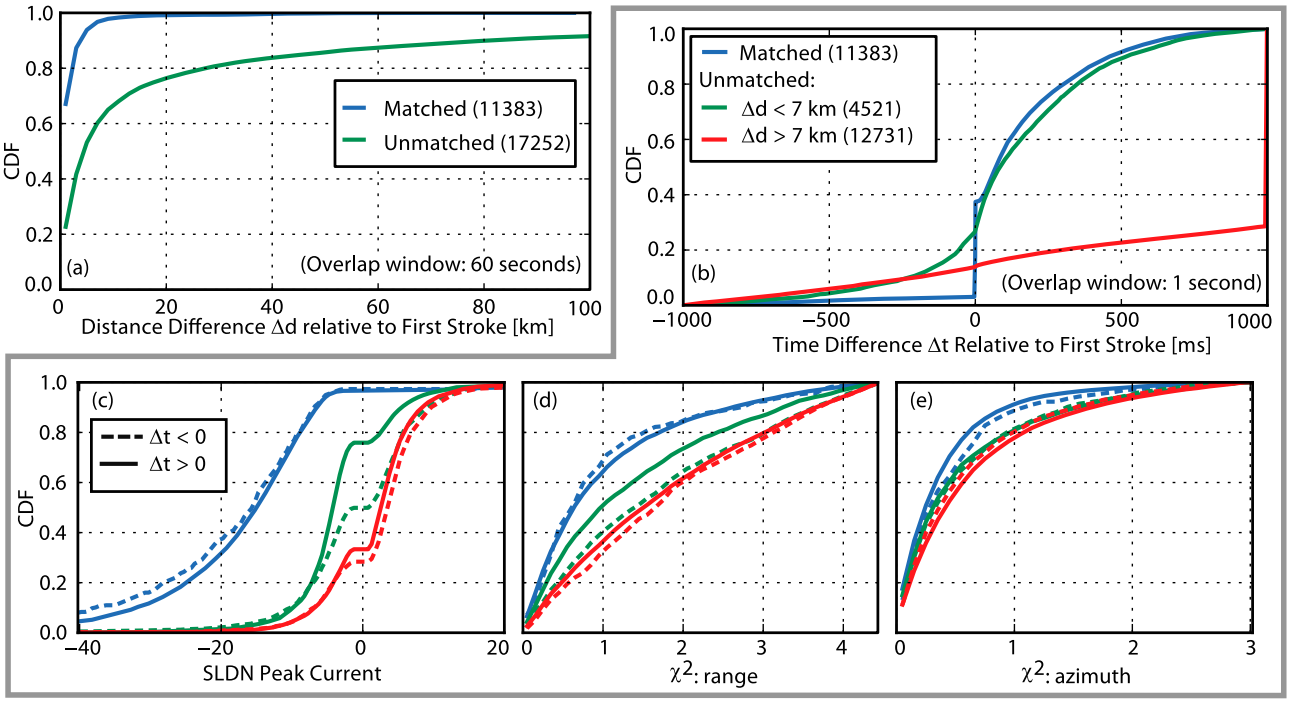


Figure 11. Temporal and spatial correlations between NLDN and SLDN geolocation data. (a) Distance between matched (blue) and unmatched (green) SLDN events and NLDN first negative CG strokes using a 60 s comparison window. (b) Time difference between matched (blue) and unmatched (green and red) SLDN events and NLDN first negative CG strokes using a 1 s comparison window. Green (red) curve shows the time difference for unmatched SLDN events that occur less than (greater than) 7 km from a first negative CG NLDN stroke within the 1 s comparison window. Unmatched events that do not occur within 1 s of an NLDN first stroke are assigned to $\Delta t = 1$. (c) SLDN peak current CDFs for events corresponding to the three categories from Figure 11b. The dotted (solid) line shows the CDF for SLDN events that occur before (after) the first NLDN stroke, within a 1 s comparison window. (d) Range and (e) azimuth error contribution to overall χ^2 value in equation (1), using the same color and line legend as Figure 11c.

consistent with the poor polarity estimation (which is related to the range estimate accuracy; see Figure 9) of the last row in Table 1. The distribution of the azimuth's contribution to χ^2 for the unmatched events has only a slightly higher skew than the matched events, lending further evidence that the unmatched events correlate to real, though weaker, discharges. In addition, for lower peak current events, which have a lower SNR at each receiver, the ability of each sensor to estimate the correct polarity diminishes. Since the range estimation is closely tied to the polarity estimate, it is reasonable to assume that the polarity estimation for these events will be less robust. Many of the small negative events may therefore be small cloud discharges with an incorrect polarity estimation.

[78] The SLDN long-range network, using the VLF signature only, does not 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 has a modified mode excitation spectrum in accordance with the height-gain function [e.g., *Wait*, 1962]. The mode structure will additionally be affected by the dipole's orientation, especially under an anisotropic ionosphere [*Pappert and Bickel*, 1970]. The degree to which one can determine the discharge type using long-range VLF measurements is left to be determined in future work.

6.1. Detection Efficiency

[79] This section looks more closely at the factors determining the detection efficiency (DE) of SLDN, as referenced to NLDN. 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 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 NLDN strokes matched to SLDN events divided by the total number of NLDN strokes.

[80] Figure 12a (top) shows a histogram of all NLDN strokes during the synoptic recording periods. Since only ground discharges reported by the NLDN are included here, and all low-current (<15 kA [*Cummins and Murphy*, 2009]) positive-polarity events are labeled as cloud discharges,

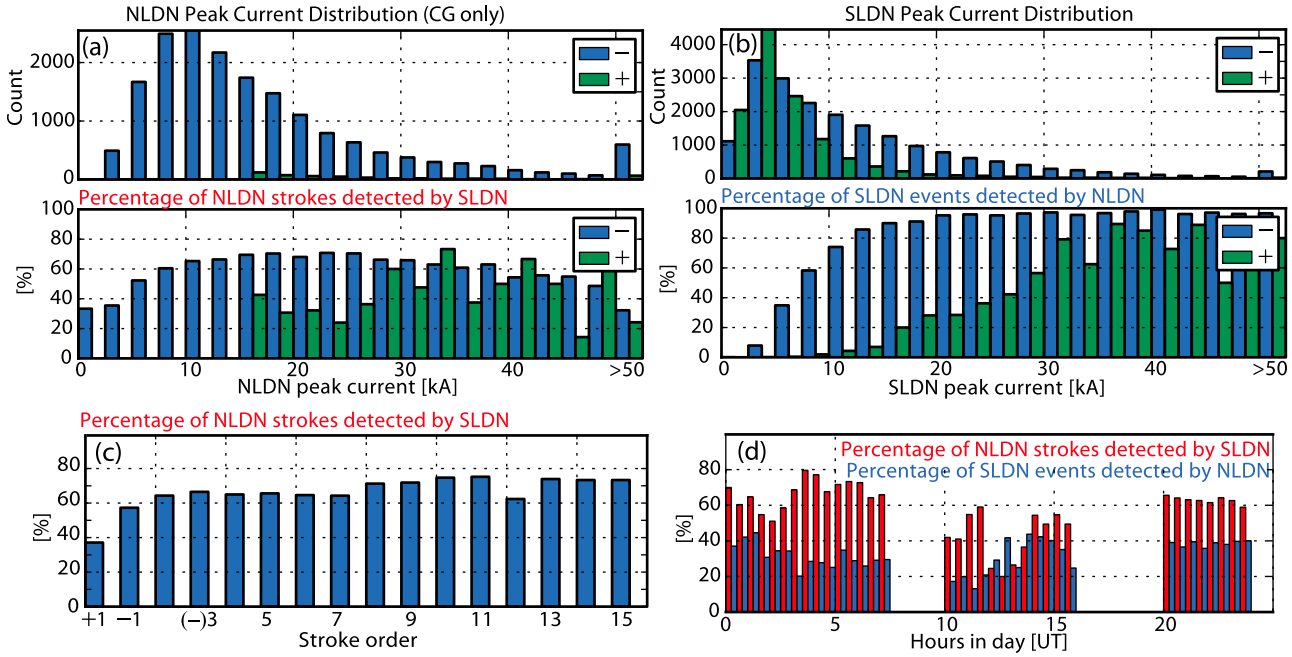


Figure 12. Stroke detection efficiency using NLDN data as a reference. (a) NLDN peak current histogram and percentage of NLDN strokes matched to SLDN events, distributed according to the NLDN-reported peak current in 2.5 kA bins. Negative (positive) NLDN CG strokes are binned in the blue (green) bars. (b) SLDN-reported peak current histogram and percentage of SLDN events matched to NLDN. The color coding is the same as Figure 12a but with the SLDN reported polarity. (c) Percentage of NLDN strokes that are matched to SLDN by stroke order. The left-most bin represents positive CG strokes; bins 1–15 represent negative CG strokes, with bins 2–15 corresponding to subsequent strokes in a negative CG flash. (d) Percentage of NLDN (SLDN) strokes detected by SLDN (NLDN), shown in red (blue), plotted in 30 min increments between 0000 and 2400 UT. The blank segments correspond to time ranges when simultaneous data from three sensors was not available.

there are no low positive peak current events from NLDN. Figure 12a (bottom) shows the percentage of NLDN 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 sensor. The dip in DE for the lower peak current values is expected as the SNR degrades at the Juneau sensor.

[81] Figure 12b plots the converse: a peak current distribution of SLDN strokes on Figure 12b (top) with the percent in each bin that were matched to NLDN on Figure 12b (bottom). The high correlation rate for large (≳20 kA) discharges coupled with the 90% stroke detection efficiency of NLDN suggests very few spurious events in the SLDN. The spatiotemporal relationship of the many weak, unmatched events discussed in connection with Figure 11 is in evidence here by the low percentage of weak SLDN events detected by NLDN, especially weak positive events. During the time of this data set, the NLDN was reporting 50% detection efficiency threshold was ~5 kA [*Cummins and Murphy*, 2009].

[82] Figure 12c plots the percentage of NLDN ground strokes reported by SLDN versus stroke order. The NLDN-referenced DE is relatively low for positive discharges (+1), is higher for first negative ground strokes (−1), and remains relatively flat for negative subsequent strokes. These results are consistent with the use of the empirical waveform bank tuned to the latter class of discharges, and the increased variability of source waveforms for discharges that pioneer a new channel where the range estimation may not be as consistently reliable.

[83] The NLDN-referenced DE and converse are plotted versus time of day in Figure 12d. Both the detection efficiency and number of events not detected by NLDN increase under a nighttime ionosphere, between 0300 and 1100 UT, consistent with the improved polarity estimation from Table 1 for nighttime paths and lower attenuation at night. Both parameters remain stable during the daytime hours, between 1400 and 2400 UT. There is a slight dip in performance during the day to night transition, between 0000 and 0300 UT. There is a sharp decrease in performance during the night to day transition, between 1100 and 1400 UT, possibly due to a more dramatic shift in waveform structure during this latter transition period. Introducing partial day/night waveform banks would likely improve performance during these transition periods.

6.2. Location Accuracy and Peak Current

[84] Figure 13a shows a scattergram of SLDN-determined peak current against the NLDN peak current for the entire data set. In each quadrant, the magnitude of the slope is close to unity, consistent with the VLF amplitude correlation

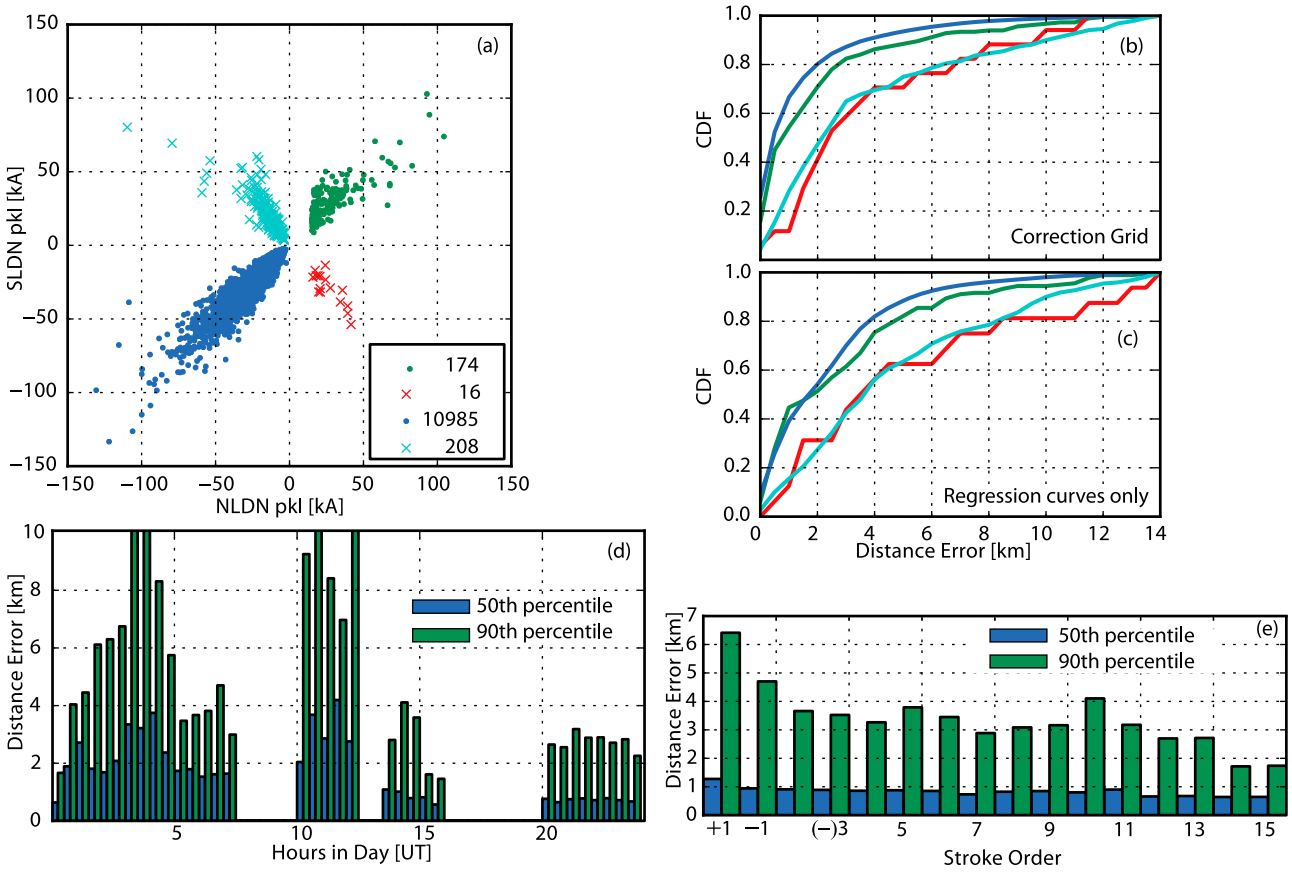


Figure 13. Comparison of peak current and location between NLDN strokes and SLDN events. (a) Peak current scatterplot, with the NLDN (SLDN) peak current reported on the x (y) axis. The blue and green points correspond to strokes where the SLDN gave the same polarity as NLDN; the cyan and red points correspond to strokes where SLDN gave the opposite polarity. (b) Global CDF of the difference in location between matched NLDN strokes and the corresponding SLDN event, where the colors correspond to the polarity matching quadrants in Figure 13a. (c) Location error CDFs using the same color legend as Figures 13a and 13b. Geolocation results were obtained using single regression curves for the threshold and zero-crossing delay at all three sensors. (d) The 50th (90th) percentile location accuracy in blue (green) in 30 min increments between 0000 and 2400 UT. (e) The 50th (90th) percentile location accuracy in blue (green) versus NLDN stroke order, using the same ordering as Figure 12c. The location error in Figures 13b, 13d, and 13e are taken from SLDN geolocation results that used the full empirical correction grid.

with NDLN-reported peak current seen in Figure 6c. Out of 11,383 matched CG strokes, 11,159 (98%) were identified with the same polarity by both networks. While 91% of the NLDN-reported positive CG discharges had a polarity match to SLDN data, the relatively large number of negative CG events in this data set skews the bias score for positive events. The positive bias score — the number of positive SLDN events divided by number of positive NLDN events, where a score greater than (less than) 1 indicates a systematic overestimation (underestimation) of a positive polarity — is 2.01. The negative bias score, on the other hand, is 0.98, indicating a good overall estimation of the number of negative CG events.

[85] The location accuracy, defined as the position difference between NLDN and SLDN locations, for the matched strokes is analyzed in Figures 13b–13e. With the exception of Figure 13d, all statistics are dominated by daytime events due to the relatively large number of discharges in the late local afternoon hours. The location accuracy CDF curves in Figures 13b and 13c highlight the benefit of a full empirical correction grid. The overall median location accuracy for correctly identified polarities (blue and green curves) is $\lesssim$1 km when the full empirical correction grids for the arrival time correction factors are used. This value increase to ~2 km when the averaged zero-crossing delays are derived from the waveform banks (Figure 13c). Without the full correction grids, the location accuracy is still less than the feature size of a compact cell and is therefore useful for high-resolution tracking of thunderstorm activity. Nevertheless, the full correction grid results, which improve the location accuracy by a factor of ~2 and are used in Figures 13d and 13e, may be constructed on a global scale. In the absence of a reference network, one could use redundant arrival time measurements with four or

more sensors to minimize a global χ^2 value over many discharges, as was done by *Lee* [1989]. There is also a good opportunity for accurate VLF propagation modeling to be employed in order to theoretically generate the correction grids over 'new' propagation paths.

[86] The dependence of the location accuracy with the time of day (Figure 13d) and stroke order (Figure 13e) follows the discussion related to DE. Due to a stable ionosphere, the daytime results have a lower median error of $\lesssim$1 km (1400–2400 UT). The location errors for nighttime paths are considerably larger, with a median stroke distance error of ~2 km. 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, ones that consider the time since sunset, for example, may improve the location estimates. As with the DE, the performance degrades considerably during the night-to-day transition, starting at ~1100 UT. The decrease performance immediately following the onset of nighttime across the United States, after 0300 UT, may be due to a slower transition to the nighttime profile at odds with a binary decision at each sensor to determine the day/night profile at each distance based on the sun's location. Considering Figure 13e, the location accuracy is worse for positive strokes (+1), and improves with first negative (−1) and the more regularly shaped subsequent stokes.

[87] The spatial variation of the detection efficiency and location accuracy depends largely on the geometrical configuration of the network. Within the interior of the network, the location accuracy and DE are relatively uniform. The single-sensor performance characteristics shown in Figures 5, 6, 9, and 10 suggest that the usable detection radius for each sensor is ~6000 km. While we expect the per-sensor performance to extend out to this range, this sensor configuration does not afford a real-world test of the network performance for events out to these distances. The only portion in the United States that extends to 6000 km from any of the sensors is southern Florida. This region was not included in the network comparison study because of the elongated error ellipses resulting from the SLDN's geometry: this region is along the baseline between Juneau (and Chistochina, had it been used) and Taylor.

7. Discussion

[88] The waveform banks, regression curves, and correction matrices were derived using NLDN-referenced sferics from the United States, the same region over which the performance metrics in section 6 were evaluated. However, the waveform comparison in Figure 2 suggests that an empirically derived waveform bank and regression curve set may be applied to a variety of propagation paths throughout the globe, with a location accuracy penalty consistent with the time deviation of the zero-crossing delay due to propagation effects. Only a small subset of paths used in the trial network were actually represented in the waveform bank, yet it was successfully applied to a multiplicity of paths to geolocate discharges throughout the United States. In an operational global network, custom correction matrices may be established by using redundant arrival-time measurements to minimize an aggregate cost function, as was done by *Lee* [1989]. Furthermore, if the phase structure is sufficiently modified by a particular path profile to the point where a single entry for the specific day/night split is insufficient at that sensor, one could build up an expanded set of waveform banks (using referenced sferics from geolocation results derived from at least three other sensors) for the offending propagation directions.

[89] Results in this work were also derived using only an all-day and all-night waveform bank and a detailed analysis of the impact from this type of interpolation are outside the scope of this paper. Figure 2a suggests that for a mixed day/night path, the zero-crossing delays take approximately a linear dependence on the day/night percentage. For long-distance propagation paths, in the absence of a profile with the specific day/night percentage, a linear interpolation may therefore be used between the all-day and all-night profile. It is important to note, however, that the entire waveform will not assume a smooth transition between the daytime and nighttime shapes, especially for shorter propagation distances where sustained tweeks [*Shvets and Hayakawa*, 1998] are found under a nighttime ionosphere.

[90] Finally, it should be emphasized that the canonical waveform bank shown in Figure 3 reflects an 'averaged' negative subsequent stroke source type. Discharges with current sources that deviate significantly from this canonical form may not correlate as well with these waveform bank entries; the resulting drop in system performance for these classes of discharges was partially characterized in sections 5 and 6. Previous studies have categorized source types based on the time scale of the ground wave and the ray hops, including *Schonland et al.* [1940], *Caton and Pierce* [1952], and *Chapman and Pierce* [1957]. An extension of the waveform bank and/or correction curves at the central processor may improve the network's performance with first ground strokes and cloud pulses.

8. Conclusion

[91] We have developed a new paradigm for long-range lightning detection and geolocation. 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 for subsequent negative CG discharges 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 widely separated sensors capable of measuring the arrival azimuth of individual sferics, each received waveform may be compared to an azimuth- and local time–dependent stored waveform bank. The result of this correlation gives an estimate of the sferic range, polarity, and, crucially, allows the identification of a low time variance feature of the waveform for use in a geolocation algorithm at the central processor.

[92] Using this new technique for extracting arrival time and correcting for second-order propagation effects empirically, three sensors were used to geolocate events in the continental United States for several synoptic periods in a 28 h time span. By using NLDN as a reference data set, discharge 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 NLDN-referenced CG

stroke detection efficiency of 40–60% was measured, and a significant amount of tightly clustered lightning activity was additionally detected that was not registered by NLDN and that may be from many weak subsequent CG strokes, M components, and cloud pulses. This location accuracy and detection efficiency either compare favorably or are significantly better than the performance metrics reported by the existing long-range networks. The ability of this approach to correctly identify polarity (negative bias score of 0.98) and to provide "reasonable" estimates of peak current are also unique functions of this technique. Since the network does not distinguish between ground and cloud pulses, we emphasize that many of the peak current estimates — those corresponding to (uncategorized) cloud pulses — do not necessarily correspond to a channel-based current as they do for CG strokes.

[93] **Acknowledgments.** This work has been supported at various times by a Stanford Electrical Engineering Departmental Fellowship, National Science Foundation grant ANT-0538627-002, NASA Graduate Student Researchers Program Fellowship NNM06AA04H, and a collaboration agreement with Vaisala, Inc. Some of the data acquisition was supported by the Office of Naval Research via grant N00014-06-1-1036. We thank the three anonymous reviewers for thoroughly reading the manuscript and providing valuable recommendations to improve the quality of this paper.

References

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, *J. Geophys. Res.*, *112*, D05208, doi:10.1029/2006JD007341.

Burke, C. P., and D. L. Jones (1992), An experimental investigation of ELF attenuation rates in the Earth-ionosphere duct, *J. Atmos. Terr. Phys.*, *54*, 243–250.

Burke, C. P., and D. L. Jones (1995), Global radiolocation in the lower elf frequency band, *J. Geophys. Res.*, *100*(D12), 26,263–26,272, doi:10.1029/95JD02735.

Butler, K. E., and R. D. Russell (2003), Cancellation of multiple harmonic noise series in geophysical records, *Geophysics*, *68*, 1083–1090, doi:10.1190/1.1581080.

Caton, P., and E. Pierce (1952), The waveforms of atmospherics, *Philos. Mag.*, *43*, 393–409.

Chapman, J., and E. T. Pierce (1957), Relations between the character of atmospherics and their place of origin, *Proc. Inst. Radio Eng.*, *45*, 804–806.

Cherna, E. V., and E. J. Stansbury (1986), Sferics rate in relation to thunderstorm dimensions, *J. Geophys. Res.*, *91*(D8), 8701–8708, doi:10.1029/JD091iD08p08701.

Christian, H. J., et al. (1999), The lightning imaging sensor, in *11th International Conference on Atmospheric Electricity*, edited by H. J. Christian, pp. 746–749, NASA Marshall Space Flight Cent., Huntsville, Ala.

Christian, H. J., et al. (2003), Global frequency and distribution of lightning as observed from space by the optical transient detector, *J. Geophys. Res.*, *108*(D1), 4005, 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. Geophys. Res.*, *108*(D24), 4779, 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 Trans. Geosci. Remote Sens.*, *44*, 1504–1510.

Cohen, M., U. Inan, and E. Paschal (2010a), Sensitive broadband ELF/VLF radio reception with the AWESOME instrument, *IEEE Trans. Geosci. Remote Sens.*, *48*, 3–17.

Cohen, M. B., U. S. Inan, R. K. Said, and T. Gjestland (2010b), Geolocation of terrestrial gamma-ray flash source lightning, *Geophys. Res. Lett.*, *37*, L02801, doi:10.1029/2009GL041753.

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 Trans. Electromagn. Compat.*, *51*, 499–518, doi:10.1109/TEMC.2009.2023450.

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 Trans. Electromagn. Compat.*, *40*, 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. Geophys. Res.*, *103*(D8), 9035–9044, doi:10.1029/98JD00153.

Davies, K. (1990), *Ionospheric Radio*, Peter Peregrinus Ltd., London.

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

Fishman, G. J., et al. (1994), Discovery of intense gamma-ray flashes of atmospheric origin, *Science*, *264*, 1313–1316, doi:10.1126/science.264.5163.1313.

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

Horner, F. (1954), The accuracy of the location of sources of atmospherics by radio direction finding, *Proc. Inst. Electr. Eng., Part III*, *74*, 383–390, doi:10.1049/pi-3.1954.0091.

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. Geophys. Res.*, *90*(A8), 7457–7469, doi:10.1029/JA090iA08p07457.

Inan, U. S., T. F. Bell, V. P. Pasko, D. D. Sentman, E. M. Wescott, and W. A. Lyons (1995), VLF signatures of ionospheric disturbances associated with sprites, *Geophys. Res. Lett.*, *22*, 3461–3464, doi:10.1029/95GL03507.

Inan, U. S., M. B. Cohen, R. K. Said, D. M. Smith, and L. I. Lopez (2006), Terrestrial gamma ray flashes and lightning discharges, *Geophys. Res. Lett.*, *33*, L18802, doi:10.1029/2006GL027085.

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, *J. Atmos. Oceanic Technol.*, *23*, 1082–1092.

Jerauld, J., V. A. Rakov, M. A. Uman, K. J. Rambo, D. M. Jordan, K. L. Cummins, and J. A. Cramer (2005), An evaluation of the performance characteristics of the U.S. National Lightning Detection Network in Florida using rocket-triggered lightning, *J. Geophys. Res.*, *110*, D19106, doi:10.1029/2005JD005924.

Kemp, D. T. (1971), The global location of large lightning discharges from single station observations of elf disturbances in the Earth-ionosphere cavity, *J. Atmos. Terr. Phys.*, *33*, 919–927, doi:10.1016/0021-9169(71)90091-2.

Kemp, D. T., and D. Llanwyn Jones (1971), A new technique for the analysis of transient ELF electromagnetic disturbances within the Earth-ionosphere cavity, *J. Atmos. Terr. Phys.*, *33*, 567–572.

Kitagawa, N., M. Brook, and E. J. Workman (1962), Continuing currents in cloud-to-ground lightning discharges, *J. Geophys. Res.*, *67*(2), 637–647, doi:10.1029/JZ067i002p00637.

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

Lagouvardos, K., V. Kotroni, H. Betz, and K. Schmidt (2009), A comparison of lightning data provided by ZEUS and LINET networks over western Europe, *Nat. Haz. Earth Syst. Sci.*, *9*, 1713–1717.

Lee, A. C. L. (1986), An experimental study of the remote location of lightning flashes using a VLF arrival time difference technique, *Q. J. R. Meteorol. Soc.*, *112*, 203–229.

Lee, A. C. L. (1989), The limiting accuracy of long wavelength lightning flash location, *J. Atmos. Oceanic Technol.*, *6*(1), 43–49, doi:10.1175/1520-0426(1989)006<0043:TLAOLW>2.0.CO;2.

Lee, A. C. L. (1990), Bias elimination and scatter in lightning location by the VLF arrival time difference technique, *J. Atmos. Oceanic Technol.*, *7*, 719–733.

Leon-Garcia, A. (1994), *Probability and Random Processes for Electriacal Engineering*, 2nd ed., Addison-Wesley, Boston, Mass.

Le Vine, D. M., J. C. Willett, and J. C. Bailey (1989), Comparison of fast electric field changes from subsequent return strokes of natural and triggered lightning, *J. Geophys. Res.*, *94*(D11), 13,259–13,265, doi:10.1029/JD094iD11p13259.

Llanwyn Jones, D., and D. T. Kemp (1970), Experimental and theoretical observations on the transient excitation of Schumann resonances, *J. Atmos. Terr. Phys.*, *32*, 1095–1100.

Lyons, W. A. (1996), Sprite observations above the U.S. High Plains in relation to their parent thunderstorm systems, *J. Geophys. Res.*, *101*(D23), 29,641–29,652, doi:10.1029/96JD01866.

Macgorman, D. R., D. W. Burgess, V. Mazur, W. D. Rust, W. L. Taylor, and B. C. Johnson (1989), Lightning rates relative to tornadic storm evolution on 22 May 1981, *J. Atmos. Sci.*, *46*, 221–251.

McDonald, T. B., III, M. A. Uman, J. A. Tiller, and W. H. Beasley (1979), Lightning location and lower-ionospheric height determination from two-station magnetic field measurements, *J. Geophys. Res.*, *84*(C4), 1727–1734, doi:10.1029/JC084iC04p01727.

McRae, W. M., and N. R. Thomson (2000), VLF phase and amplitude: Daytime ionospheric parameters, *J. Atmos. Sol. Terr. Phys.*, *62*, 609–618.

Morgan, R. R. (1968), World-wide VLF effective conductivity map, *Rep. 8013F-1*, Westinghouse Electric, Cranberry Township, Pa.

Nyman, D. C., and J. E. Gaiser (1983), Adaptive rejection of high-line contamination, *SEG Tech. Prog. Expanded Abstr.*, *2*(1), 321–323, doi:10.1190/1.1893897.

Pappert, R. A., and J. E. Bickel (1970), Vertical and horizontal VLF fields excited by dipoles of arbitrary orientation and elevation, *Radio Sci.*, *5*(12), 1445–1452, doi:10.1029/RS005i012p01445.

Pearson, K. (1901), On lines and planes of closest fit to systems of points in space, *Philos. Mag.*, *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, *J. Atmos. Oceanic Technol.*, *26*, 145–166.

Peter, W. B., and U. S. Inan (2007), A quantitative comparison of lightning-induced electron precipitation and VLF signal perturbations, *J. Geophys. Res.*, *112*, A12212, doi:10.1029/2006JA012165.

Petersen, W. A., and S. A. Rutledge (1998), On the relationship between cloud-to-ground lightning and convective rainfall, *J. Geophys. Res.*, *103*(D12), 14,025–14,040, doi:10.1029/97JD02064.

Rafalsky, V. A., A. P. Nickolaenko, A. V. Shvets, and M. Hayakawa (1995), Location of lightning discharges from a single station, *J. Geophys. Res.*, *100*(D10), 20,829–20,838, doi:10.1029/95JD01532.

Rodger, C., J. Brundell, R. Dowden, and N. Thomson (2004), Location accuracy of long distance VLF lightning location network, *Ann. Geophys.*, *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, *Ann. Geophys.*, *23*, 277–290.

Rodger, C. J., J. B. Brundell, R. Holzworth, and E. Lay (2009), Growing detection efficiency of the world wide lightning location network, in *Coupling of Thunderstorms and Lightning Discharges to Near-Earth*, *Am. Inst. Phys. Conf. Ser.*, vol. 1118, pp. 15–20, Am. Inst. of Phys., College Park, Md.

Sao, K., and H. Jindoh (1974), Real time location of atmospherics by single station techniques and preliminary results, *J. Atmos. Terr. Phys.*, *36*, 261–266, doi:10.1016/0021-9169(74)90046-4.

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.

Sazhin, S. S., and M. Hayakawa (1992), Magnetospheric chorus emissions —A review, *Planet. Space Sci.*, *40*, 681–697.

Schonland, B. F. J., J. S. Elder, D. B. Hodges, W. E. Phillips, and J. W. van Wyk (1940), The wave form of atmospherics at night, *Proc. R. Soc. A*, *176*, 180–202.

Shafer, D. C. (1994), Spread-spectrum VLF remote sensing of the ionosphere, Ph.D. thesis, Stanford Univ., Stanford, Calif.

Shvets, A. V., and M. Hayakawa (1998), Polarisation effects for tweek propagation, *J. Atmos. Terr. Phys.*, *60*, 461–469.

Tapia, A., J. A. Smith, and M. Dixon (1998), Estimation of convective rainfall from lightning observations, *J. Appl. Meteorol.*, *37*, 1497–1509.

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), 1703–1706, doi:10.1029/1999GL010845.

Thomson, N. R., M. A. Clilverd, and W. M. McRae (2007), Nighttime ionospheric D region parameters from VLF phase and amplitude, *J. Geophys. Res.*, *112*, A07304, doi:10.1029/2007JA012271.

Uman, M. A. (2001), *The Lightning Discharge*, Dover, Mineola, N. Y.

Wait, J. R. (1960), Terrestrial propagation of very-low-frequency radio waves: A theoretical investigation, *J. Res. Natl. Bur. Stand., Sect. D*, *64*(2), 153–203.

Wait, J. R. (1962), Excitation of modes at very low frequency in the Earth-ionosphere wave guide, *J. Geophys. Res.*, *67*(10), 3823–3828, doi:10.1029/JZ067i010p03823.

Wait, J. R. (1970), *Electromagnetic waves in Stratified Media*, Pergamon, New York.

Wait, J. R., and K. P. Spies (1964), Characteristics of the Earth-ionosphere waveguide for VLF radio waves, *NBS Tech. Note U.S.*, *300*, 93 pp.

Watt, A. D. (1967), *V.L.F. Radio Engineering*, Elsevier, New York.

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, doi:10.1029/RS021i006p00964.

K. L. Cummins, Department of Atmospheric Sciences, University of Arizona, Tucson, AZ 85721, USA.

U. S. Inan and R. K. Said, Department of Electrical Engineering, Stanford University, 350 Serra Mall, Rm. 356, Stanford, CA 94305, USA. (rsaid@stanford.edu)