Mass Statistical Analysis of Early VLF Events

Nikhil Pailoor[1] and Morris Cohen[1]

[1]Georgia Institute of Technology

November 24, 2022

Abstract

The analysis of Very Low Frequency (VLF, 3-30 kHz) radio back-scattering can be used to measure the impact of lightning on the D region of the ionosphere (60-90 km). Early/fast events are prompt and rapid changes to the D-region ionosphere associated with certain lightning flashes, causing heating, ionization, and attachment. Previous work has observed the behavior of early/fast events and their connection to specific types of lightning flashes through VLF remote sensing and lightning geolocation, but the unique nature of each event makes it difficult to broadly infer the interactions between lightning and the ionosphere using a small number of case studies.

We assembled a massive database of VLF amplitude samples for cases when high intensity lightning occurs near a transmitter-receiver path. We constructed an artificial neural network to detect and label early/fast events. With a large volume of events compiled, we charted detailed statistics of event occurrences and behavior.

We find a correlation between lightning current magnitude and event likelihood, as well as inverse correlation between event likelihood and distance to transmitter-receiver path. We further confirm the asymmetry of the peak current trends, with positive-current strokes being significantly more likely to produce an event. We find that increased distance of the lightning to the transmitter, and to a lesser extent to the receiver, decreases the probability of an ionospheric disturbance. We find that recovery time is largely not a function of the peak current. We do not find evidence that long-recovery events are a distinct class of Early/Fast events.

Mass Statistical Analysis of Early VLF Events

N. A. Pailoor[1]*, M. B. Cohen

[1]Georgia Institute of Technology

*Atlanta, Georgia

Corresponding author: Nikhil Pailoor, `npailoor3@gatech.edu`

Abstract

The analysis of Very Low Frequency (VLF, 3-30 kHz) radio back-scattering can be used to measure the impact of lightning on the D region of the ionosphere (60-90 km). Early/fast events are prompt and rapid changes to the D-region ionosphere associated with certain lightning flashes, causing heating, ionization, and attachment. Previous work has observed the behavior of early/fast events and their connection to specific types of lightning flashes through VLF remote sensing and lightning geolocation, but the unique nature of each event makes it difficult to broadly infer the interactions between lightning and the ionosphere using a small number of case studies.

We assembled a massive database of VLF amplitude samples for cases when high intensity lightning occurs near a transmitter-receiver path. We constructed an artificial neural network to detect and label early/fast events. With a large volume of events compiled, we charted detailed statistics of event occurrences and behavior.

We find a correlation between lightning current magnitude and event likelihood, as well as inverse correlation between event likelihood and distance to transmitter-receiver path. We further confirm the asymmetry of the peak current trends, with positive-current strokes being significantly more likely to produce an event. We find that increased distance of the lightning to the transmitter, and to a lesser extent to the receiver, decreases the probability of an ionospheric disturbance. We find that recovery time is largely not a function of the peak current. We do not find evidence that long-recovery events are a distinct class of Early/Fast events.

1 Introduction

Lightning is a chaotic phenomenon which has an impact on the Earth's ionosphere that is still not fully understood. One example of such effects are early/fast Very Low Frequency (VLF) events, a class of ionospheric disturbances distinguished by their occurrence within 100ms after an intense lightning stroke, affecting an area within several hundred kilometers of the stroke. These disturbances occur in the D region of the ionosphere (60 km - 90 km altitude). The D region is too high for direct balloon or aircraft observations, and too low for satellite measurements. However, by monitoring the scatter of VLF radio signals from the D region, a process known as VLF remote sensing, we can observe changes in the electron density and thickness. There is a long history of using VLF radio signals for D-region remote sensing *Silber and Price* [2016] and specifically for lightning-associated disturbances*Inan et al.* [2010].

Early/fast events were originally distinguished from a separate phenomenon, Lightning-induced Electron Precipitation Events (LEP Events), in 1983 [*Armstrong*, 1983]. LEP events carry a similar signature of a sharp perturbation in the received VLF signal, followed by a gradual recovery. However, early/fast events represent direct interactions between the lightning stroke and the ionosphere, while the perturbations in LEP events are the result of electron precipitation from the ionosphere, caused by VLF energy from the lightning stroke escaping into the radiation belt and dislodging trapped particles there. LEP events are characterized by a 1s delay between the lightning stroke and the resulting perturbations of VLF signals, as well as a significantly longer region of disturbance. In contrast, early/fast events typically have an onset delay of <100 ms. As such, *Inan et al.* [1988] coined the term "early/fast", distinguishing both the "early" onset, along with the "fast" perturbation (onset times ¡ 1s). This terminology was later used to distinguish these events from "early/slow" events, which work such as *Haldoupis et al.* [2006] suggested had a different process of generation. More recent work such as *Kotovsky and Moore* [2015] suggests that the distinction between early/fast and early/slow events may not be so clear cut when examining the total scattered field of the VLF signal, rather than just the amplitude changes. Early/fast events typically have recovery times in the range of 10-100sdriven primarily by atmospheric chemistry of recombination and attachment, although *Cotts and Inan* [2007] observed a class of "long recovery" early/fast events with recovery times ranging up to 20 minutes.

One of the first proposed mechanisms for early/fast events was direct heating from electromagnetic pulses (EMPs) produced by the lightning stroke. *Inan* [1990] used the 100 kW 28.5 kHz VLF transmitter NAU, based in Arecibo, Puerto Rico, to demonstrate the ability of VLF waves to cause heating in the lower ionosphere, and suggested that lightning discharges could cause similar effects due to VLF energy released. The authors noted that the rapid onset of early/fast events is consistent with the propagation speed and pulse-width of EMPs. However, the authors suggested that the slow recoveries of those events indicate other mechanisms contributing a role, such as ionization enhancements. Modeling efforts by *Marshall et al.* [2008] suggested that EMPs could cause significant enough changes in the local ionosphere to result in perturbations in propagating VLF signals; however, this model was unable to explain why early/fast events are overwhelmingly the result of positive current strokes.

Pasko et al. [1995] suggested that the source of ionospheric heating was not EMPs but rather the quasi-electrostatic (QE) effect from lightning clouds. This refers to a buildup of charge occurring in the cloud preceding a lightning stroke. This charge produces a strong static electric field, which under the proposed theory could couple with The intense stroke immediately removes this charge, triggering a change in the quasi-statis fields which recover in 10s of ms. Further investigation by *Inan et al.* [1996a], however, proposed that instead of of the ionosphere being heated by the quiescent QE effect, the main mechanism at play was sustained heating of the ionosphere from charges built up during thunderstorms. A sufficiently high-intensity lightning stroke, under this model, would cause a marginal change in the electrostatic field to result in a perturbation in the local ionosphere density. The sustained heating model as originally formulated predicted much smaller perturbations in VLF signals than observed. However, recently updated models that take into account the Earth's geomagnetic field have been able to predict perturbations consistent with the 0.2-1dB range typically observed. [*Kabirzadeh et al.*, 2017]

In 1989, researchers accidentally discovered a new atmospheric electrical phenomenon known as "sprites" [*Franz et al.*, 1990]. These refer to discharges of electricity that occur in the mesosphere, at the tops of clouds and extend up to 20 km higher. *Inan et al.* [1995] found a possible connection between early/fast events and sprites. This study examined six early/fast events that occurred between June 29th and July 12th, 1994, in a campaign conducted by the University of Alaska Fairbanks. The authors observed several instances of sprites occurring simultaneously to early/fast events. However, the authors also noted that sprites may not be the exclusive cause of early/fast events, as sprites are typically associated with high charge-moment positive current lightning strokes (¿ 60 kA) while early/fast events appear to occur from positive or negative strokes.

Dowden et al. [1996] proposed an explanation for the relationship between sprites and early/fast VLF events. This examined an event observed by *Fukunishi et al.* [1996], using instrumentation and data from University of Colorado's SPRITES'95 campaign [*Lyons*, 1996]. The event was caused by an intense (+326 kA) cloud-to-ground stroke that occurred 231 km from the Yuca Ridge Field Station, where a set of VLF receivers observed a perturbation in the signal received from the NLK VLF transmitter. At the same time, a set of sprites were observed above the lightning stroke, in the altitude range of 75-105 km. *Dowden et al.* [1996] suggested, based on analysis of the VLF data, that the event was caused by scattering from the sprite body itself. *Dowden and Rodger* [1997] went on to suggest that the logarithimic decay of early/fast events is a product of the vertical body of sprites, with the time scale being strongly dependent on the altitude of the plasma column.

Some more recent studies have expanded this case study process with the benefit of a larger dataset. *Salut* [2013] examined the scattering produced from 7769 lightning strokes, from which they identified 1250 events. They observed an asymmetry between positive and negative-current strokes of the same amplitude, where positive strokes were 5 times more likely to produce an event. Long recovery events in particular correlated strongly to positive strokes, and 83% of them occurred over sea rather than land. The authors also found a correlation between peak current of the lightning stroke and the size of the perturbed region of the ionosphere, extending up to 400 km from the lightning stroke. The authors found this geography consistent with the theory of early/fast events being produced by electromagnetic pulses from the lightning stroke directly heating the ionosphere.

NaitAmor et al. [2010] investigated the role that geography plays in determining how early/fast events are observed through their scattering of VLF signals. The authors found that both the raw distance from the receiver to the location of the ionospheric disturbance, and the scattering angle between the transmitter-receiver path and the transmitter-stroke path, are relevant to determining appearance and behavior of a signal perturbation at the receiver.

In this paper, we build the largest database of early/fast events yet assembled, using an automated machine-learned search process, to quantify with more specificity the relationship between path geometry, lightning peak current and polarity, and VLF early/fast event characteristics. To do this, we use lighting geolocation data as a starting point and cross-reference every stroke with a network of VLF/LF receivers across the continental US, to identify instances where an early/fast event may have occurred based on the lightning position. We train a neural-network classifier to analyze these VLF data samples and classify every candidate event as being an early/fast events (or non-event).

2 Methods

2.1 Instruments used

To accumulate a large volume of samples, we made use of Georgia Tech's network of Atmospheric Weather Electromagnetic System for Observation Modeling and Education (AWESOME) receivers stationed across North America. These receivers each use a pair of cross-looped air-core antennas that detect the magnetic field in both the north-south (N/S) and east-west (E/W) directions. The receivers have a low frequency cutoff of ∼500 Hz and a flat passband from 3-400 kHz, followed by a dropoff until a high frequency cutoff of 470 kHz due to the built-in anti-aliasing filters. The noise levels remain flat at -10 dB $fT/\sqrt{Hz}$ over the 18-30 kHz frequency band. The receivers have a 96 dB dynamic range due to the 16-bit ADC, and typically detect magnetic field signals from the femtoTesla range to the nanoTesla range. The receiver is described by *Cohen* [2018].

The receivers are used to detect scattering from the VLF signals produced by five Navy-operated transmitters as they propagate through a disturbed D-region ionosphere. These VLF signals reflect efficiently from the D-region (as well as the round) and are thus guided to global distances. Due to the long wavelengths corresponding to VLF propagation, the transmitting antennas are top-loaded dipole arrays, described in detail in *Watt* [1967]. The locations, call signs, and frequencies of these transmitters are listed in Table 1.

The transmitters make use of a Minimum Shift Key (MSK) modulation scheme with 200 baud rate. By decoding this modulation and removing ambiguities in the phase, we can separate out the signal into the components of an elliptically polarized wave as described by *Gross et al.* [2018]. The polarization technique involves combining the amplitude and phase data from both the E/W and N/S antennas and mapping them to the Major Axis, Minor Axis, Tilt Angle, and Starting Phase of a polarization ellipse.

Call Sign	Location	Frequency (kHz)
NLK	Jim Creek, Washington	24.8
NML	LaMoure, North Dakota	25.2
NAA	Cutler, Maine	24.0
NPM	Lualualei, Hawaii	21.4
NAU	Aguada, Puerto Rico	40.75

Table 1. List of VLF Transmitters detectable in North America

The data collected is sampled at both high time resolution (50 Hz) and low time resolution (1 Hz). For the purposes of Early VLF Event analysis, we used only the low resolution data. The transmitters and receivers sued are shown in Figure 1

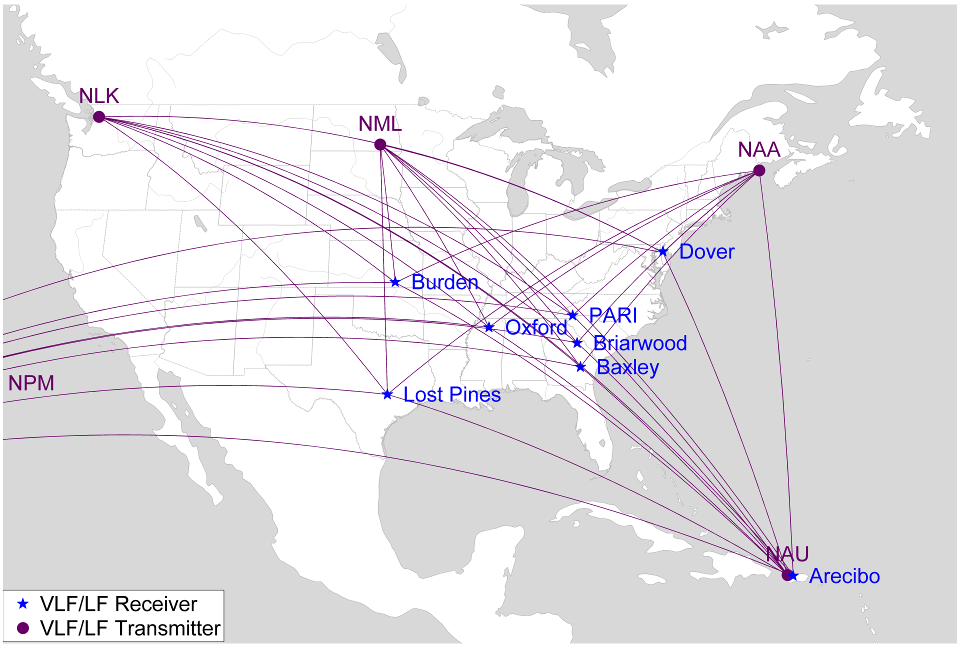

Figure 1. Georgia Tech LF Radio Lab VLF Receiver Network

2.2 Data collection

The National Lightning Detection Network (NLDN) continuously monitors lightning stroke activity in the United States. Their data includes a list of individual lightning incidents, giving precise information about geographical location, peak current, type (intracloud, cloud to ground, or ground to cloud) and polarity of each stroke [*Cummins et al.*, 1998].

Starting from September of 2017 and running to the end of June 2018, we used the NLDN to identify all lightning strokes occurring within 600 km of a transmitter-receiver path.

600 km is chosen to be fairly large, as Early/fast events more commonly occur within 100 km of the transmitter-receiver path, but by choosing a large circle we can quantify the probability as a function of distance, even if the probability of an early/fast is low. 600 km is also chosen because it excludes most lightning-induced election precipitation (LEP) events, which at these latitudes are typically poleward-displaced by several hundred km.

We screened for only the cases where the entire ionosphere (85 km) from transmitter-receiver was under nighttime conditions, as Early/fast events are known to occur almost exclusively at nighttime, if not entirely exclusively. For each path within this range, we extracted a sampled window of narrowband data.

Figure 2 shows this process. Here, a stroke occurring in upstate New York creates a potential perturbation area with a radius of 600 km. The NAA-Dover and NAA-PARI transmitter-receiver paths fall within this range, and as such we can examine the narrowband receiver data at both sites corresponding to the NAA frequency (24.0 kHz). In addition, the NLK and NML transmitters' paths to Dover (overlapping) intersect the edge of the perturbation circle, so we can examine those narrowband frequencies detected at the Dover receiver as well. However, the NAU and NPM transmitters' paths to Dover do not intersect with the perturbation circle, so we do not include the narrowband data from those frequencies. Similarly, the NAA-Arecibo path does not intersect with perturbation circle, so the 24.0 kHz narrowband data received at Arecibo is left out of out database. In summary, the data samples corresponding to this stroke would be the NAA-Dover, NAA-PARI, NLK-Dover, and NML-Dover narrowband samples. We excluded all other paths from analysis.

308,351 samples matching the above criteria were collected and stored in an Sqlite database, along with accompanying metadata such as the current of the lightning stroke, the location and geometry of the stroke and the transmitter-receiver path, and the date and time of the incident. Note that for many stroke locations there were multiple transmitter-receiver paths that went through the 600 km radius, each of which were treated as a separate sample since the geometry was different. In total, the 308,351 samples resulted from 32391 lightning strokes.

Each sample contains four channels of data, representing each component of the polarization ellipse. The samples begin 40 seconds before the stroke and end 120 seconds after, with 1 Hz resolution.

308,351 samples were collected and stored in an Sqlite database, along with accompanying metadata such as the current of the lightning stroke, the location and geometry of the stroke and the transmitter-reciever path, and the date and time of the incident.

2.3 Event Classifier

Many of the more than 1 million samples do not indicate an Early/fast event, so to manually screen out non-events would be a tedious exercise. To handle the large volume of data without this manual sorting, we constructed a classifier to identify the early/fast events automatically. To do this, a random selection of 1000 samples were manually examined and labeled as either "Events" or non-events, based on visual inspection given an understanding of previous early/fast Event observations throughout the literature. The 1000 samples were then evenly divided between training and test data.

Because a machine-learning based classifier, or really any detection algorithm, will always have a threshold for error, there is a fundamental tradeoff between building a classifier with a low false positive rate and one with a high detection efficiency. In order to

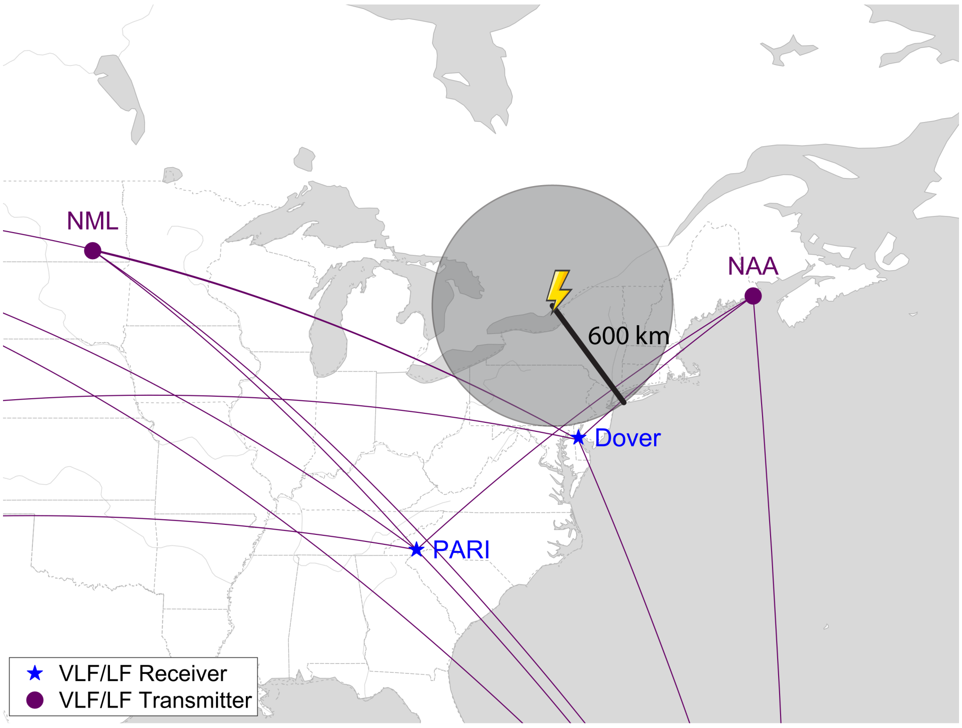

Figure 2. Hypothetical event occurring in upstate New York. Image shows several intersecting transmitter-receiver paths.

	Event	Non-Event
Classified as Event	116	22
Classified as non-Event	29	333

Table 2. Test results for classifier

accurately reflect the broader trends in the data, we chose a detection threshold that seems balances the two.

After training, the network yielded a test accuracy - that is, the percentage of test samples accurately classified - of 90%. 20% of samples classified as "Events" were false positives, while 15.9% of all actual events were classified as "non-Events". The total distribution of the classified samples is shown in Table 2.

To further test the incidence of false positives, we selected 100,000 samples from the larger database, and for each one, applied the algorithm to the data that was 120 seconds later. Nearly all of these do not have an LEP event exactly 120 seconds later, since LEP events are sparse in time, though a few may have another LEP event by coincidence. We then ran the classifier over these 100,000 non-event samples, and found that. 0.63% of these were classified as events. Since these were nearly all false positives, this allows to set a statistical significance level when applying the classifier to real data. For example, out of the 32,391 candidates, the classifier detected a total of 5548 events, or 17.1%. This is much higher than the 0.63% false positive probability indicating that most (>96%) of the selected events are likely real.

3 Observations

With this database of 5548 events, most of which are presumed to be real, we now present some statistics of early/fast event occurrence and quantify the connection to path geometry and lightning stroke properties. Some of our observations are consistent with previous findings around early/fast events, which lends some additional confidence to the automatic detection algorithm, but in addition we present several new observations using this database.

3.1 The Role of Proximity of Stroke to the Path

Figure 3 shows, on the left, the distribution of all samples (both events and non-events) as a function of (closest) distance from the stroke to the transmitter-receiver path, in 50-km bins. The total number of samples matches the >300 thousand as described earlier. As this represents all samples of lightning occurrence, there is a roughly even distribution, as expected, since lightning should be distributed evenly as a function of proximity to one of our VLF propagation paths. The right half of Figure 3 shows the fraction of samples that were classified as events, as a function of distance from lightning source to the VLF transmitter-receiver path. The red line indicates our quantification of the noise floor, or the false positive probability as discussed earlier of 0.63%. Values near or below this line are effectively too few to be measured with our current classification algorithm. There is a clear tendency for the probability of an Early/fast event to decrease as a function of distance from the transmitter-receiver path. By far the largest number of events occur within 50 km, which is consistent with *Moore* [2003]. However, there is a still a detectable and measurable quantity of early/fast events out to at least 400 km away.

3.2 The Role of Lightning Peak Current

Figure 4 shows, as 3 does with distance, the role peak current has in affecting the likelihood of an event to form. Here, as the chart on the left shows, the distribution of candidate strokes is heavily concentrated among lower intensity strokes, reflecting the distribution of peak current of NLDN-detected strokes. Meanwhile, the graph of the likelihood of events, on the right, show the strong positive relationship of peak current to the likelihood of events forming. There are no bins for lightning strokes less than 100 kA as these strokes were not collected in our search. In other words the probability of these strokes producing events is so low that it is indistinguishable from errors in the classifier. Meanwhile, strokes above 300 kA have over a 18% likelihood of forming an event, although as the figure on the left shows, the sample size of strokes in this current range is relatively small.

There is also a visibly asymmetric relationship between peak current and event likelihood, with positive current strokes being considerably more likely to form events than negative current events. In nearly every current bin, positive current strokes are twice as likely as their negative counterparts to form an event. The asymmetric behavior of event occurrence is also consistent with the work of *Salut* [2013]. However, it should be noted that these high intensity strokes are orders of magnitude less likely to occur than their lower intensity counterparts. Figure 5 displays the quantity of events in each peak current bin. While concentration of events is not as skewed towards lower current ranges as it is for the total population of samples, the majority (>80%) of events are still caused by lightning strokes with a magnitude less than 200 kA. As a higher volume of data is collected, the confidence of these trends may be tested.

Consistent with previous research [*Salut*, 2013] [*Inan et al.*, 2010], these events are limited in distance, as shown in Figure 4. Events are less likely to be detectable, if present, as the distance from the incident to the transmitter-receiver path is increased.

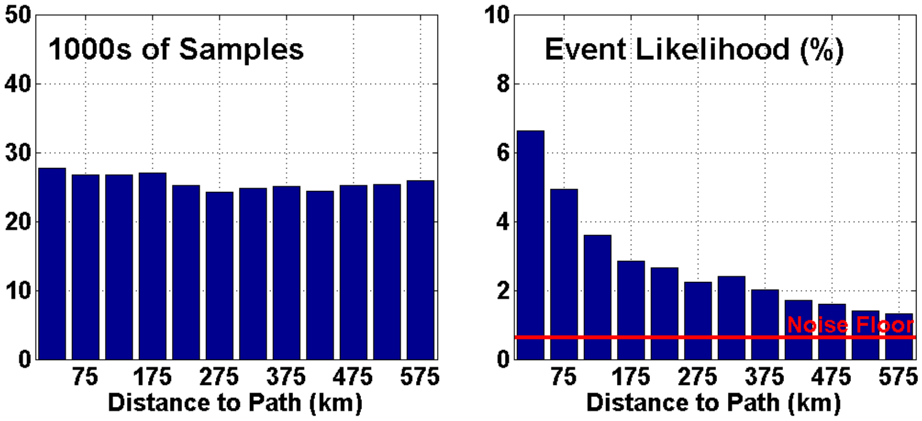

Figure 3. Changes in event occurrence and behavior over distances to transmitter-receiver path

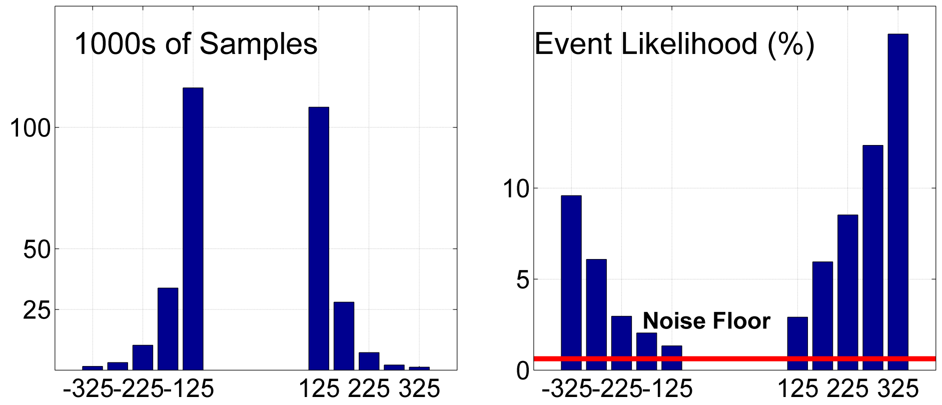

Figure 4. Changes in event occurrence and behavior over peak lightning current

Figure 6 shows that event occurrence also has an inverse relationship to both distances from the transmitter (on the left) and distances from the receiver (on the right). Note that the bins in this graph are logarithmic spaced, as the only constraint for searching for events is distance to path. A lightning stroke can occur at close distance to the transmitter-receiver path, while being at a long distance from either the transmitter or receiver. If back-scatter events are included, the only upper limit on the distance to either the transmitter or receiver is the Earth's circumference.

For the distance to transmitter, the dropoff in event likelihood for higher distances is highly visible, and is dramatic between the 750 km and 1500 km bins. One factor is that distance to transmitter is often a proxy for signal strength, and weaker signals may be perturbed by changes in the ionosphere to a lesser degree. It seems unlikely that the lightning near VLF transmitters are systematically different from the lighting that is farther away enough to demonstrate this type of relationship. As such, the dropoff in event occurrence with distance is likely a VLF propagation effect. In particular, near the transmitter, there is wider variety of propagating VLF modes, both TE and TM, particularly at high orders. As the VLF energy propagates further away from the transmitter, many of the higher order modes are attenuated away, leaving only a small number of mostly

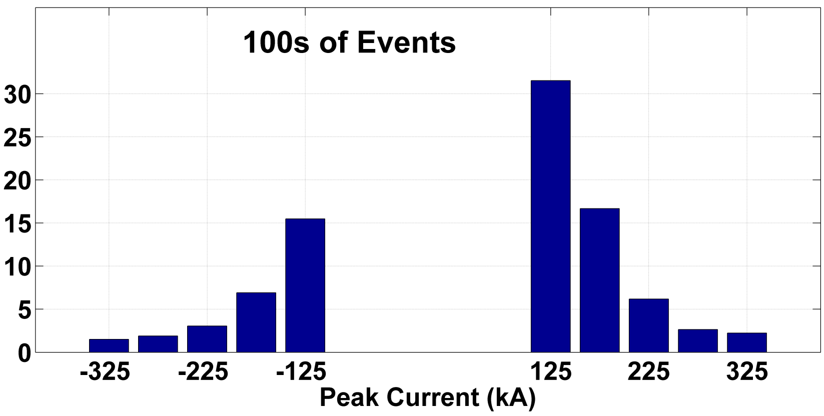

Figure 5. Distribution of events by lightning current.

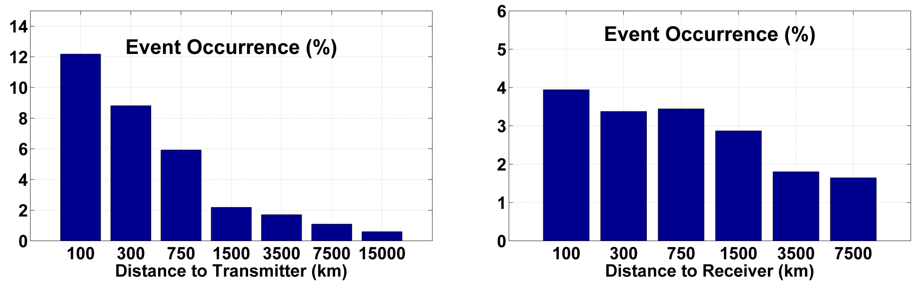

Figure 6. Changes in event occurrence and behavior as a function of distance from stroke to transmitter (left) and distance from stroke to receiver (right).

TM modes dominating the signal. Past theoretical investigations have found that lower order modes tend to be less affected by an ionospheric disruption[*Poulsen et al.*, 1993]. Samples in the higher distance bins, particularly in the 7500 km and 15000 km bins may also be more heavily contaminated by backscatter events, as these distances are much larger than most of the transmitter-receiver path lengths, and it is known that backscatter events are rare.

The dropoff in early/fast probability with distance the receiver is considerably less dramatic, although visible nonetheless. Strokes in the 1500 km bin were over 25% less likely than events in the 100 km bin. These differences may represent the same relationship to distance to path as shown in 3 as, for any given point along the transmitter-receiver path, a higher perpendicular distance to the path will also have a higher distance to the receiver. Some of the dropoff in event occurrence rate with distance may also be due to the fact that the scattered modes generated by the VLF wave against the ionospheric disturbance may fade with distance, being themselves composed of higher order modes.

3.3 Individual Path Analysis

Table 3 shows the distribution the geographic distribution of samples by transmitter-receiver path, while Table 4 shows the likelihood of those samples containing an early/fast event. Significantly fewer lightning strokes were detected along the paths between the Juneau receiver and the three mainland transmitters. This reflects the tendency of intense lightning strokes to occur closer to the tropical regions. Note that the receivers at

	Arecibo	Baxley	Briarwood	Burden	Delaware	Juneau	Oxford	PARI
NAA	4229	1128	795	7000	1709	196	219	6217
NAU	0	1779	1090	12888	11704	7490	1177	13675
NLK	8635	4203	2546	6810	7604	37	2065	9001
NML	10377	4429	2385	8415	7218	677	3006	8856
NPM	4969	2723	1381	3594	6659	0	693	5731

Table 3. Sample distribution across Transmitter/Receiver paths

	Arecibo	Baxley	Briarwood	Burden	Delaware	Juneau	Oxford	PARI
NAA	0.0192	0.0496	0.0818	0.0354	0.2165	0.0051	0.0320	0.0286
NAU	N/A	0.0090	0.0064	0.0151	0.0179	0.0264	0.0178	0.0085
NLK	0.0076	0.0193	0.0342	0.0366	0.0153	0.0811	0.0634	0.0221
NML	0.0093	0.0646	0.0683	0.1616	0.0529	0.1064	0.1277	0.0533
NPM	0.0105	0.0066	0.0014	0.0086	0.0144	N/A	0.0173	0.0112

Table 4. Event occurrence at each Transmitter/Receiver path

times have been off line for certain periods of the years, also contributing to the discrepancy in number of strokes recorded.

The NAA-Delaware path produced the highest concentration of events, with over 21% of strokes falling in range of this path resulting in events. This may be due to a combination of geographical factors, as well as the short distance of this path resulting in a stronger signal that would be more sensitive to ionospheric fluctuations.

3.4 Perturbation Intensities

Event perturbations varied depending on peak current and distance to path. Figure 7 shows the probability distributions of the major axis perturbation, plotted for varying distance to path values.

At all distances, there is an asymmetry of positive and negative perturbations. This asymmetry is more predominant at larger distance ranges, suggesting that events with a negative signal perturbation are considerably shorter in range. The predominance of positive perturbation events is consistent with previous research.[*Inan*, 1993][*Inan et al.*, 1996b][*Marshall et al.*, 2006] *Marshall et al.* [2008] suggested that this predominance is because the underlying causes of early/fast events involve a reduced electron density in the perturbed region of the ionosphere, which in turn would tends towards higher VLF signal absorption.

3.5 Recovery analysis

Due to the noisiness of the data, as well as constant changes in the background D region at nighttime, measuring recovery time is an imprecise process. To account for noise

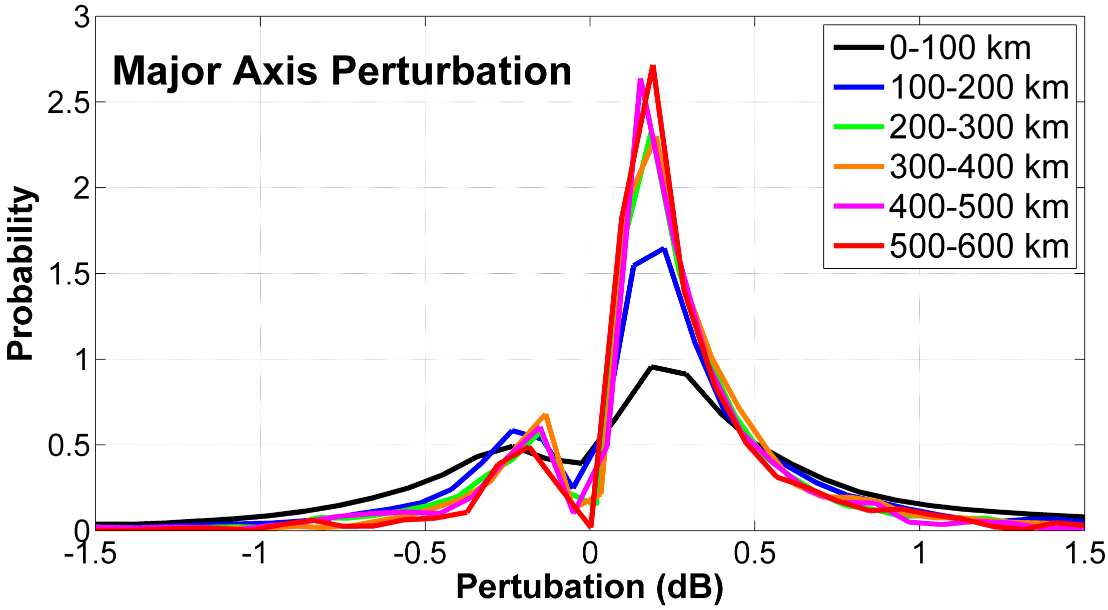

Figure 7. Changes in event occurrence and behavior over peak lightning current

and background fluctuations, we perform a least-squares linear fit for the VLF data from 10 seconds to 2 seconds before the lightning incident to a line, and did a separate linear fit to the data 2 seconds to 10 seconds after the incident. The reason the line is fit to only 8 seconds of data is to ensure contamination from multiple Early events in a row, and to account for the fact that at least some Early events may have recoveries only around 10 seconds. The data points including and immediately surrounding the incident were excluded to avoid contamination from the return stroke's direct radiation. The point where these lines intersect is defined here as the recovery time. This process is no doubt imperfect method but measuring the recovery time is itself an inexact process; when done manually it sometimes requires some amount of subjective judgment, much like defining what is or is not an Early VLF event. Our hope is defining a technique here is to determine some statistical properties with consistent criteria.

To illustrate this process, Figure 8 shows two samples from events. The first, on the left, originates from a May 31, 2018 signal at 07:05:29 UTC from NML received at Burden, corresponding to a lightning stroke of intensity 336 kA. The second, on the right, is taken from a January 22, 2018 event at 02:34:59 UTC caused by a 249 kA lightning stroke, detected at PARI from the NLK signal. Both samples are displayed in log scale after being processed through a 5-point median filter. The red lines illustrate an estimation for the background ionosphere, while the green line represents an estimate for the recovery after the initial perturbation. We observe that for the event on the left, these two lines meet at a point along the signal 47.2 seconds after the lightning occurrence. It this case it appears that the signal has recovered to the background ionospheric levels at this point. For the event on the right, however, the intersection of the green and red lines do not correspond to a point along the signal. This is because successive events are occurring, roughly at t=45 and t=75, which cause the signal to drop repeatedly below the background levels. Nonetheless, we define the recovery time at the point where the lines intersect, as this represents the recovery process from the initial event. We use this definition to minimize interference from succeeding events and other background noise. We can describe the formula for the recovery time as follows:

$$t_{recovery} = \frac{c_2 - c_1}{m_1 - m_2}$$

Where m1 and c1 represent the slope and y intercept respectively of the background line, and m2 and c2 represent the slope and intercept of the recovery line.

We note that while this method can accurately characterize the event recoveries for shorter time scales, the error scales quadratically with longer recovery times. This is a result of the inverse relationship between the recovery slopes and the recovery times, and the propagation of errors from a least squares interpolation. We attempt to account for this by estimating the error range for each sample, taking into account both the variance of the data, and the length of the estimated recovery time. The full calculation used is described below.

The general formula for a propagation of errors can be described as follows, if q is a function of variables x,...,z[*Taylor*, 1997, p.75]:

$$\delta q = \sqrt{(\frac{\partial q}{\partial x}\delta x)^2 + ... + (\frac{\partial q}{\partial z}\delta z)^2}$$

Where δq, δx, and $deltaz$ are the errors of q, x, and z respectively. Therefore, we can estimate δt_r as:

$$\delta t_r = \sqrt{(\frac{\partial t_r}{\partial m_1}\delta m_1)^2 + (\frac{\partial t_r}{\partial m_2}\delta m_2)^2}$$

Since $c = y_{avg} - mx_{avg}$, $—\frac{\partial c}{\partial m}—$ is simply equal to the average absolute values of the times of the points being sampled relative to the time of the lightning stroke, which are 20.5 and 24.5 for the background and the disturbed samples respectively.

We can therefore make the following derivations:

$$\frac{\partial t_r}{\partial m_1} = \frac{t_r(t_r + 20.5)}{c_2 - c_1}, \frac{\partial t_r}{\partial m_2} = \frac{t_r(t_r + 24.5)}{c_2 - c_1}$$

From the generalized formula of least squares linear regression error[*Taylor*, 1997, p.188],

$$\delta m = \sigma_y \sqrt{\frac{N}{\Delta}}$$
$$\Delta = N\Sigma x^2 - (\Sigma x)^2$$

Putting this together, and factoring out $\frac{sigma_y}{p}$ where $p = c_2 - c_1$ or the perturbation:

$$\delta t_r = \frac{\sigma_y}{p}\sqrt{\frac{N_1}{\Delta_1}t_r^2(t_r + 20.5)^2 + \frac{N_2}{\Delta_2}t_r^2(t_r + 24.5)^2}$$

If the lines do not intersect after the stroke, or if the calculated error was greater than both the bin size (20 seconds) or the estimated recovery time, we discarded the sample as not having a recovery time that could be estimated with confidence. If the recovery time is greater than 300s, it is treated as a long recovery event (LORE) [*Cotts and Inan*, 2007; *Haldoupis et al.*, 2013].

Figure 9 shows the distribution of event recoveries. Each bar shows the number of event samples with a Major Axis recovery time in the corresponding bin. Unlike some of the previous charts shown, this is not a quantification of distinct events, but rather different samples. Because the same event may be captured on multiple transmitter-receiver paths, different samples may display different recovery times. Excluded from this chart are event samples where the Major Axis recovery time has been found to be negative, as this is typically a consequence of either noisy data or rapidly changing background ionosphere.

Based on a Kolmogorov-Smirnov two-sample test[*Marsaglia et al.*, 2003], there is not a significant difference in the distributions for high peak current events (right panel) and more typical peak current events (left panel). The test yields a p-value of 0.5886.

A plurality of event samples have recovery times less than 20s, and each increasing time bin has a decreasing number of event samples. In addition, 0.66% of low current events and 1.19% of high current events could confidently be described are LOREs, using 300s as the threshold.

We do not see a bifurcation of this distribution, meaning that we cannot see an easily separable category of LORE events distinct from ordinary events. This suggests that LORES may not be fundamentally different types of events but rather at the extremes of a continuum, with similar underlying physics. As mentioned previously, however, this method of calculating recovery times becomes increasingly imprecise for higher recovery times, meaning further investigation is necessary.

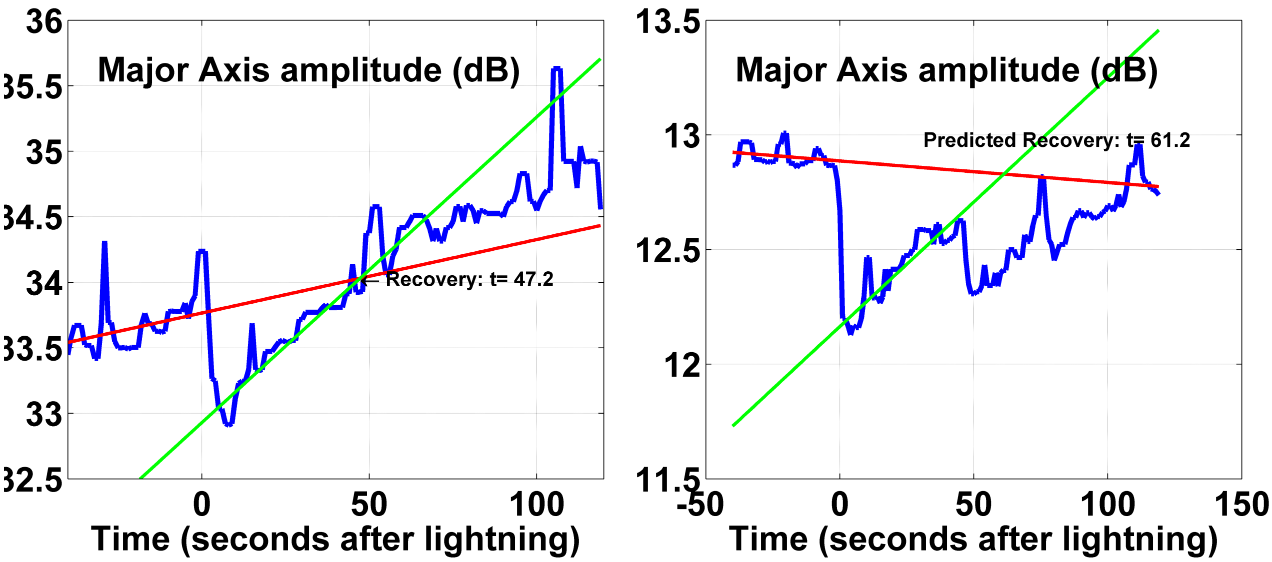

Figure 8. Two events, with the recovery time estimate shown.

4 Conclusion

The constant fluctuations in the ionosphere, and the inherent randomness in lightning properties and distribution mean that the behavior of each early/fast is unique. How-

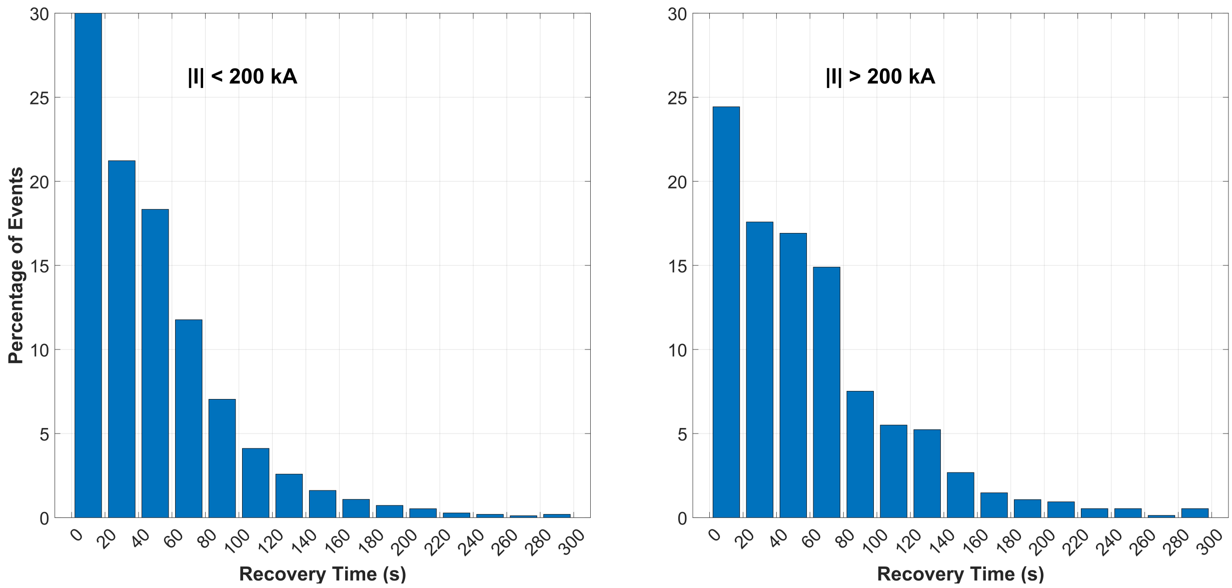

Figure 9. Distribution of Event Recoveries

ever, by collecting a large enough sample, we can still isolate patterns in the occurrence and appearance of these events.

Our efforts represent a starting point towards a more comprehensive, data-driven approach towards early/fast event analysis. Working with a larger dataset means that standardized measurements of event perturbation and recovery time will be imprecise and potentially skewed by noise. Nonetheless, our observations seem to confirm much of the evidence from previous research, particularly in the asymmetric likelihoods of positive vs negative current strokes producing events, as well as the much greater proportion of positive-perturbation events.

Acknowledgments

This work is supported by the National Science Foundation under awards AGS-1451210417 and AGS-1653114 (CAREER) to the Georgia Institute of Technology, and by the Defense Advanced Research Projects Agency (DARPA) through US Department of the Interior award D19AC00009 to the Georgia Institute of Technology.

References

Armstrong, W. C. (1983), Recent advances from studies of the trimpi effect, *Antarctic Journal of the United States*.

Cohen, M. (2018), Broadband VLF/LF/MF Radio Reception and Remote Sensing with the AWESOME Instrument, .

Cotts, B. R. T., and U. S. Inan (2007), VLF observation of long ionospheric recovery events, *Geophysical Research Letters*, *34*(14), 10.1029/2007gl030094.

Cummins, K. L., M. J. Murphy, E. A. Bardo, W. L. Hiscox, R. B. Pyle, and A. E. Pifer (1998), A combined TOA/MDF technology upgrade of the u.s. national lightning detection network, *Journal of Geophysical Research: Atmospheres*, *103*(D8), 9035–9044, 10.1029/98jd00153.

Dowden, R. L., and C. J. Rodger (1997), Decay of a vertical plasma column: A model to explain VLF sprites, *Geophysical Research Letters*, *24*(22), 2765–2768, 10.1029/97gl02822.

Dowden, R. L., J. B. Brundell, and W. A. Lyons (1996), Are VLF rapid onset, rapid decay perturbations produced by scattering off sprite plasma?, *Journal of Geophysical Research: Atmospheres*, *101*(D14), 19,175–19,183, 10.1029/96jd01346.

Franz, R. C., R. J. Nemzek, and J. R. Winckler (1990), Television image of a large upward electrical discharge above a thunderstorm system, *Science*, *249*(4964), 48–51, 10.1126/science.249.4964.48.

Fukunishi, H., Y. Takahashi, M. Kubota, K. Sakanoi, U. S. Inan, and W. A. Lyons (1996), Elves: Lightning-induced transient luminous events in the lower ionosphere, *Geophysical Research Letters*, *23*(16), 2157–2160, 10.1029/96gl01979.

Gross, N. C., M. B. Cohen, R. K. Said, and M. Gołkowski (2018), Polarization of Narrowband VLF Transmitter Signals as an Ionospheric Diagnostic, *Journal of Geophysical Research: Space Physics*, *123*(1), 901–917, 10.1002/2017ja024907.

Haldoupis, C., R. J. Steiner, Á. Mika, S. Shalimov, R. A. Marshall, U. S. Inan, T. Bsinger, and T. Neubert (2006), "early/slow" events: A new category of VLF perturbations observed in relation with sprites, *Journal of Geophysical Research*, *111*(A11), 10.1029/2006ja011960.

Haldoupis, C., M. Cohen, E. Arnone, B. Cotts, and S. Dietrich (2013), The VLF fingerprint of elves: Step-like and long-recovery early VLF perturbations caused by powerful ±CG lightning EM pulses, *Journal of Geophysical Research: Space Physics*, *118*(8), 5392–5402, 10.1002/jgra.50489.

Inan, U. S. (1990), VLF heating of the lower ionosphere, *Geophysical Research Letters*, *17*(6), 729–732, 10.1029/gl017i006p00729.

Inan, U. S. (1993), VLF signatures of lightning-induced heating and ionization of the nighttime D-region, *Geophysical Research Letters*.

Inan, U. S., D. C. Shafer, W. Y. Yip, and R. E. Orville (1988), Subionospheric VLF signatures of nighttime d region perturbations in the vicinity of lightning discharges, *Journal of Geophysical Research*, *93*(A10), 11,455, 10.1029/ja093ia10p11455.

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, *Geophysical Research Letters*, *22*(24), 3461–3464, 10.1029/95gl03507.

Inan, U. S., V. P. Pasko, and T. F. Bell (1996a), Sustained heating of the ionosphere above thunderstorms as evidenced in "early/fast" VLF events, *Geophysical Research Letters*, *23*(10), 1067–1070, 10.1029/96gl01360.

Inan, U. S., A. Slingeland, V. P. Pasko, and J. V. Rodriguez (1996b), VLF and LF signatures of mesospheric/lower ionospheric response to lightning discharges, *Journal of Geophysical Research: Space Physics*, *101*(A3), 5219–5238, 10.1029/95ja03514.

Inan, U. S., S. A. Cummer, and R. A. Marshall (2010), A survey of ELF and VLF research on lightningionosphere interactions and causative discharges, *Journal of Geophysical Research*.

Kabirzadeh, R., R. A. Marshall, and U. S. Inan (2017), Early/fast VLF events produced by the quiescent heating of the lower ionosphere by thunderstorms, *Journal of Geophysical Research: Atmospheres*, *122*(12), 6217–6230, 10.1002/2017jd026528.

Kotovsky, D. A., and R. C. Moore (2015), Classifying onset durations of early VLF events: Scattered field analysis and new insights, *Journal of Geophysical Research: Space Physics*, *120*(8), 6661–6668, 10.1002/2015ja021370.

Lyons, W. (1996), The sprites95 field campaign: Initial results–characteristics of sprites and the mesoscale convective systems that produce them, in *Preprints,* *18th Conf. on Severe Local Storms, San Francisco, CA, Amer. Meteor. Soc*, pp. 442–446.

Marsaglia, G., W. W. Tsang, and J. Wang (2003), Evaluating kolmogorov's distribution, *Journal of Statistical Software*, *8*(18), 10.18637/jss.v008.i18.

Marshall, R. A., U. S. Inan, and W. A. Lyons (2006), On the association of early/fast very low frequency perturbations with sprites and rare examples of VLF backscatter, *Journal of Geophysical Research*, *111*(D19), 10.1029/2006jd007219.

Marshall, R. A., U. S. Inan, and T. W. Chevalier (2008), Early VLF perturbations caused by lightning EMP-driven dissociative attachment, *Geophysical Research Letters*, *35*(21), 10.1029/2008gl035358.

Moore, R. C. (2003), Early/fast VLF events produced by electron density changes associated with sprite halos, *Journal of Geophysical Research*, *108*(A10), 10.1029/2002ja009816.

NaitAmor, S., M. A. AlAbdoadaim, M. B. Cohen, B. R. T. Cotts, S. Soula, O. Chanrion, T. Neubert, and T. Abdelatif (2010), VLF observations of ionospheric disturbances in association with TLEs from the EuroSprite-2007 campaign, *Journal of Geophysical Research: Space Physics*, *115*(A7), 10.1029/2009ja015026.

Pasko, V. P., U. S. Inan, Y. N. Taranenko, and T. F. Bell (1995), Heating, ionization and upward discharges in the mesosphere, due to intense quasi-electrostatic thundercloud fields, *Geophysical Research Letters*, *22*(4), 365–368, 10.1029/95gl00008.

Poulsen, W. L., T. F. Bell, and U. S. Inan (1993), The scattering of VLF waves by localized ionospheric disturbances produced by lightning-induced electron precipitation, *Journal of Geophysical Research*, *98*(A9), 15,553, 10.1029/93ja01201.

Salut, M. M. (2013), On the relationship between lightning peak current and Early VLF perturbations, *Journal of Geophysical Research.*

Silber, I., and C. Price (2016), On the Use of VLF Narrowband Measurements to Study the Lower Ionosphere and the Mesosphere–Lower Thermosphere, *Surveys in Geophysics*, *38*(2), 407–441, 10.1007/s10712-016-9396-9.

Taylor, J. R. (1997), *An introduction to error analysis*, second ed., University Science Books.

Watt, A. D. (1967), *VLF Radio Engineering*, Elsevier, 10.1016/c2013-0-02069-5.