EVALUATING THE ROLE OF LIGHTNING-INDUCED ELECTRON PRECIPITATION TOWARDS RADIATION BELT ELECTRON LOSS

A Dissertation
Presented to
The Academic Faculty

By

Nikhil Pailoor

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Engineering
Department of Electrical and Computer Engineering

Georgia Institute of Technology

Dec 2022

© Nikhil Pailoor 2022
EVALUATING THE ROLE OF LIGHTNING-INDUCED ELECTRON PRECIPITATION TOWARDS RADIATION BELT ELECTRON LOSS

Thesis committee:

Dr. Morris Cohen
Electrical and Computer Engineering
Georgia Institute of Technology

Dr. Sven Simon
Earth and Atmospheric Sciences
Georgia Institute of Technology

Dr. David Anderson
Electrical and Computer Engineering
Georgia Institute of Technology

Dr. Paul Steffes
Electrical and Computer Engineering
Georgia Institute of Technology

Dr. Andrew Peterson
Electrical and Computer Engineering
Georgia Institute of Technology

Date approved: December 1, 2022
ACKNOWLEDGMENTS

I owe a great deal of gratitude to my advisor, Dr. Morris Cohen, for providing support and advice throughout this process. His ability to explain the basics of VLF remote sensing and the dynamics of space physics were invaluable, especially during my early time in the lab, for growing comfortable with doing research in this field.

Without a doubt, this has extended to the LF Radio lab environment. Over my years, many of my fellow lab members have graduated, while new faces have arrived. But consistently, it has remained an environment of support, trust, and friendship. I thank Jackson McCormick, Evan Worthington, Marc Higginson-Rollins, Nathan Opalinsky, Parker Singletary, Lee Thompson, Nick Gross, Ted Slevin, Kevin Whitmore, David Richardson, Charles Topliff, Shweta Dutta, Rodrick Gray, and Liam Smith. In addition to my LF Lab compatriots, Alex Akins and Amoree Hodges have equally and indispensably been part of the Van Leer 515 family, and deserve a great thanks as well.

The others on my defense committee deserve a great thanks as well. Dr. Sven Simon’s courses in plasma physics have provided a bedrock of knowledge for my work. Dr. David Anderson’s machine learning course, and his advice during meetings, was very useful in improving the LEP classifier I used. Dr. Paul Steffes has been a consistent source of support and input through my time at Van Leer’s fifth floor. Dr. Andrew Peterson has helped develop and advance my understanding of electromagnetics.

I also thank Dr. Mark Golkowski and Dr. Vijay Harid of UC Denver, who provided key insights that were essential for my work, such as the Semi-Analytic Model.

Finally, I thank my mother and father, for instilling a deep love of learning and exploration from a young age.
CONTENTS

Acknowledgments ........................................................................................................ iii

List of Tables ................................................................................................................ vii

List of Figures ................................................................................................................ viii

Summary ........................................................................................................................ xi

Chapter 1: Introduction and Background ................................................................. 1

1.1 Research purpose .................................................................................................... 1

1.2 VLF interactions with the radiation belts ............................................................... 2

1.2.1 Lightning-generated whistler waves .................................................................. 3

1.3 VLF remote sensing ............................................................................................. 8

1.3.1 VLF transmitters and receivers ....................................................................... 10

1.3.2 Lightning data ..................................................................................................... 11

1.4 Contributions ......................................................................................................... 12

Chapter 2: Database of LEP Events ....................................................................... 14

2.1 Past work: Early/fast detection ............................................................................. 14

2.1.1 Database of early/fast events .......................................................................... 15

2.1.2 VLF data format ............................................................................................... 16
Chapter 2: LEP Classification

2.1 Artificial Neural Network classifier

2.1.3 Artificial Neural Network classifier ........................................... 19

2.1.4 Applicability to LEP events .................................................. 21

2.2 LEP Classification ................................................................. 21

2.2.1 Training set preparation .................................................... 22

2.2.2 Neural network structure .................................................... 26

2.2.3 Effect of sample width ....................................................... 27

2.2.4 Loss function weighting .................................................... 28

2.2.5 Comparison to alternative approaches ................................. 30

2.2.6 High resolution analysis .................................................... 30

Chapter 3: LEP Modeling ............................................................. 33

3.1 Past modeling work ............................................................... 33

3.1.1 LEP Occurrence Rate ....................................................... 33

3.1.2 Whistler wave raytracing ................................................... 34

3.1.3 Modeling VLF signal perturbations ...................................... 35

3.2 LEP Combined Model ........................................................... 37

3.2.1 Whistler-Induced Particle Precipitation model ...................... 38

3.2.2 Semi-Analytic Model ....................................................... 40

3.2.3 VLF Propagation Model .................................................. 43

3.3 Precipitation Calculations ...................................................... 44

3.3.1 Path-integrated precipitation ............................................. 45

3.3.2 Linear relation between precipitation and field perturbation ..... 46

3.3.3 Total election precipitation .............................................. 49
LIST OF TABLES

1.1 VLF transmitters operating in the United States, by location and frequency. 10

2.1 Test results for classifier ................................. 21

4.1 Distribution of LEP events according to components of the polarization ellipse that are disturbed. There are 25961 events in total. ................. 54

4.2 Likelihood of individual field components to be perturbed in an event at different peak current ranges. ........................... 57
## LIST OF FIGURES

1.1  The general shape of Earth’s magnetic field lines, with different L shells shown and labeled. Courtesy Dan Golden. 

1.2  Depiction of radiation belts in the near-Earth space environment. 

1.3  The path of a propagating whistler-mode wave through the radiation belt. From Lauben (1998) 

1.4  Motion of electrons in the Earth’s Radiation Belts. Cited from Zong, 2022 

1.5  Illustration of gyro-resonance between trapped particle and incoming polarized electric field. Figure from Sousa (2018) 

1.6  An an illustration of a typical LEP event. From Ben Cotts and Cohen, n.d. 

1.7  Network of receivers operated by Georgia Tech’s LF Radio Group 

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

2.2  Adapted from Gross et al. (2018). Geometric setup for a VLF receiver’s antenna orientation (black axes), wave propagation axes (red axes), and polarization ellipse (blue ellipse) of the magnetic flux density. The angle $\theta_{az}$ gives the nominal arrival direction from the source. The green markings show the major axis and minor axis of the ellipse, $\tau$ is the tilt angle of the ellipse from $\phi_1$, and $\chi$ is the ellipticity angle. The red vector on the ellipse shows the start phase and rotation sense of the ellipse. 

2.3  Example of an event during manual classification process. This event occurred on April 10, 2018, observed at the Baxley receiver reading the NAA signal.
2.4 Example of an non-event during manual classification process. This sample occurred on January 28, 2018, taken from the Baxley receiver reading the NLK signal. ................................................................. 20

2.5 An LEP event, shown through the minor axis of the polarization ellipse. Note the delay between the lightning sferic and the onset of the perturbation. This event occurred on April 3, 2019 detected at the Delaware receiver from the NAA signal. ................................................................. 23

2.6 Test accuracy over each training batch for the three networks compared. ................................................................. 27

2.7 Test accuracy over training batch for different time widths ................................................................. 28

2.8 Positive Predictive Value (PPV) versus True Positive Rate (TPR). Labels of each point indicate the classifier weight assigned to Events. Non-Event weight was standardized to 1.0. ................................................................. 29

2.9 High resolution image of an LEP Event detected on March 26, 2019, at the Burden receiver from the NAA transmitter. ................................................................. 31

3.1 A flowchart showing the construction of our disturbed ionosphere table, built as a function of lightning latitude, L shell, and longitudinal displacement. 38

3.2 A flowchart for the signal perturbation calculation, given the locations of a transmitter, receiver, and lightning stroke. ................................................................. 39

3.3 Example differential electron flux output from WIPP for L=2.5, simulation with source lightning of 30 degrees N. ................................................................. 40

3.4 Ambient and disturbed electron density profile for D region at L=2.5. Disturbed profile is the result of a lightning stroke at 30 N latitude, and shows the disturbance at t=5s. ................................................................. 42

3.5 Time evolution of the ionosphere disturbance. The x axis shows the change in the electron density compared to the ambient profile. ................................................................. 42

3.6 Flowchart showing the process of modeling and using real data to estimate electron precipitation. ................................................................. 45

3.7 The gridspace of precipitation flux for an event caused by a lightning stroke 33 N, 87 W (shown in red dot). ................................................................. 46

3.8 The relationship between path-integrated electron flux and perturbation of each polarization ellipse component. ................................................................. 48
4.1 Left: histogram of candidate strokes by peak current Right: Event occurrence rate for each peak current bin

4.2 Histogram of nonzero perturbations for each polarization ellipse component, from left to right, major axis, minor axis, tilt angle, and start phase.

4.3 Percentile magnitudes of field component perturbations for different peak current ranges.

4.4 A histogram of total electron precipitation of the events in our LEP database.

4.5 A histogram of the standard deviation of precipitation measurements across different polarization ellipse components, scaled to the median measurement.

4.6 A histogram of the standard deviation of precipitation measurements across different path estimates, scaled to the median measurement.

4.7 Total electron precipitation for each day in the LEP database.

5.1 Pitch angle distribution for 33 keV, 226 keV, and 1575 keV electrons.

5.2 Comparison of integrated electron flux on High and Low LEP days, with both global average and North America averages shown.

5.3 Cumulative Distribution Function (CDF) of the total electron count over North America.
SUMMARY

Lightning-induced electron precipitation (LEP) is one of the principle mechanisms for electrons to be drained from the radiation belts. An LEP event progresses with Very Low Frequency (VLF) radio wave propagation from lightning, trans-ionospheric propagation, and wave-particle gyroresonance interaction with energetic radiation belt electrons. Scattered electrons then precipitate onto the ionosphere, and this disturbance is detected of these events through VLF signals scattering off the disturbed ionosphere. This research attempts to quantify the role of LEP events, through several steps. First, we build a massive database of LEP events observed within the continental US (CONUS) by a network of VLF receivers. To do this, we employ the use of an Artificial Neural Network (ANN) classifier, in order to automatically detect LEP events from VLF signals. Second, we apply a model to estimate the total number of precipitating electrons, which we can then sum up over all LEP events to quantify lightning’s contribution within CONUS. We draw on a cascading trio of models to construct of LEP events, including a ray tracing code for whistler wave propagation, a model of electron deposition into the ionosphere, and finally a model of VLF propagation. Finally, we examine data from the Van Allen Probes, both to investigate the correlation between electron distributions and the occurrence of LEP events, and to provide a reference for the total number of electrons available in the belts to be removed. We find that LEP events within CONUS appear to be capable of removing a substantial fraction (up to 0.1%-1%) of radiation belt electrons between 33 keV and 1000 kA.
CHAPTER 1
INTRODUCTION AND BACKGROUND

1.1 Research purpose

The Earth’s radiation belts are a major topic of interest for space research and exploration. The radiation belts are layers of charged particles within the Earth’s magnetosphere that surround the planet. Satellites orbiting within the belts must be equipped with proper shielding to avoid being damaged by charged particles, which results in greater weight and operating costs. Manned missions must take caution to minimize the amount of human exposure to these hazards. The 1962 Starfish Prime nuclear test in space caused a measured disruption to the belts (Narin, 1962), one of several instances of manmade activities dramatically increasing the particle distribution in the near-Earth environment. Space weather events such as strong solar storms can also expand the radiation belts and reduce satellite lifetimes (Hudson et al., 2008). Understanding the mechanisms by which these changes occur could provide insight into topics such as active mitigation of the belts. By understanding the natural phenomena that can change the belts, we can potentially develop artificial means of altering the radiation belt (Song et al., 2022), in order to protect satellites from artificial or manmade expansions of the radiation belt, as well as facilitate safer manned space travel.

One particular mechanism for radiation belt electron loss is lightning-induced electron precipitation, caused when lightning-generated Very Low Frequency (VLF) radiation enters the radiation belt and dislodges electrons. The dislodged electrons then precipitate back to Earth, entering the ionosphere, a region of charged particles at the top of the atmosphere.

This thesis focuses on Very Low Frequency (VLF) (3-30 kHz) radio remote sensing as a method for detecting and quantifying the role of LEP events in removing electrons
from the radiation belt. VLF remote sensing takes advantage of the ionosphere’s properties in reflecting incident VLF signals, allowing detection of the signals by receivers at global distances from the VLF transmitter. A perturbation in the detected signal will therefore result from a change in the ionosphere along the transmitter-receiver path.

The goal of this thesis is to use VLF remote sensing to detect a large number of LEP events, and thereby quantify the contribution of LEP events to radiation belt dynamics. The LEP events are analyzed to find patterns in event occurrence and behavior. Finally, by modeling the process by which each event occurs, we estimate the total number of electrons precipitated for each event and extrapolate the overall role of LEP events as a loss mechanism.

1.2 VLF interactions with the radiation belts

The radiation belts are made of layers of charged particles, or plasma. They are thought to surround the Earth primarily in two regions, or ”belts”. Because the plasma is stratified along the Earth’s magnetic field lines, which roughly follows a dipole pattern, the locations of these regions are typically measured by magnetic altitude, or L value. The L value roughly refers to a magnetic field line which, at the equator, is L Earth radii from the center, as shown in Figure 1.1. The inner belt lies from L = 1.1 to L = 2.5, while the outer
belt ranges from $L = 3$ to $L = 6$, as depicted in Figure 1.2. Since Allen et al. (1958) first identified the radiation belts over 60 years ago, researchers have been investigating what the dominant mechanisms are that grow and shrink the belts, and what mechanisms add or remove particles. Walt and MacDonald (1964) showed that Coulomb collisions between particles play a role in the lower magnetic altitudes ($L < 1.25$) where particle density is higher. At higher $L$ values, VLF radio waves form the dominant mechanism.

The three principle sources for these VLF waves are plasmaspheric hiss (Lyons et al., 1972), man-made VLF signals from transmitters (Inan et al., 1978), and lightning-generated radiation (Helliwell, 1965). The degree to which each of these sources is predominant remains a topic of active research. Abel and Thorne (1998a) suggested that VLF radiation generated by lightning flashes play a significant role in affecting the electron population in the range beyond the inner radiation belt. However, this study assumed uniform behavior and occurrence of lightning-generated waves propagating through the radiation belt. In reality, the behavior of lightning-generated radiation is variable and dynamic.

Understanding the extent to which lightning-generated VLF whistler waves contribute to electron loss provides an insight towards the ability to artificially drain electrons from the radiation belt using VLF transmitters, as both of these sources interact with the radiation belt through the same fundamental wave-particle interactions. Many of the techniques used in the past to model lightning-induced electron precipitation (LEP) events have been extended to describe the precipitation due to transmitters (Kulkarni et al., 2008).

1.2.1 Lightning-generated whistler waves

LEP events are caused by leakage of a lightning stroke’s electromagnetic radiation through the ionosphere into the magnetosphere, the region dominated by the Earth’s magnetic field but well above where neutral particles exist (roughly $>1000$ km). Cloud-to-ground (CG) lightning discharges typically contain strokes of high intensity current ranging from tens to hundreds of kiloamps, which release electromagnetic energy in a broadband impulse
across the entire VLF range. When electromagnetic waves in the VLF frequency range propagate through the plasma in the lower radiation belts, the plasma conditions allow a type of plasma wave known as the whistler-wave. Ordinarily wave propagation is not possible below a frequency known as the plasma frequency, which is well above the VLF range. But the whistler mode is possible only because of the geomagnetic field through the plasma. This propagation mode occurs when the frequency of the wave is above the ion gyrofrequency and below the electron gyrofrequency, the gyrofrequency being the rate at which a charged electron or ion orbits the geomagnetic field, to be defined in the next section. Under these conditions, left hand circularly polarized waves do not propagate, while right hand circularly polarized waves propagate. The following, known as a dispersion relation, relates the wave frequency $\omega$ to the wavenumber $k$:

$$\omega = \frac{c^2 \left\| \Omega_e \right\|}{\omega^2_{pe}} k^2$$
And the electron gyrofrequency $\Omega_e$ is obtained from the magnetic field $B_0$, electron charge $q_0$, and electron mass $m_e$:

$$\Omega_e = \frac{q_0 B_0}{m_e}$$

$\omega_{pe}$ is the plasma frequency for electrons, and $c$ is the speed of light. Since the phase velocity $v_p$ is defined as $\frac{\omega}{k}$, we have the following relationship for the phase velocity as a function of frequency:

$$\frac{v_p}{c} = \sqrt{\frac{\Omega_e}{\omega_{pe}}} \sqrt{\omega}$$

This results in a dispersive waveform where group velocity grows with the square root of frequency within the VLF band. When played as an audio signal, as early VLF recordings often were, they create a "whistle" sound, hence the name "whistler wave".

Figure 1.3 shows the whistler-wave propagation behavior, where waves may "leak" through the ionosphere into the magnetic fields, where they may reflect and scatter multiple times within the radiation belts. Once in the radiation belt, these waves begin interactions with the existing charged particles.

Electrons in the near Earth environment have their movement constrained by the magnetic field. This leads to three principal motions: parallel motion, gyromotion, and drift motion. Gyromotion is a consequence of the Lorentz force from the magnetic field. The Lorentz force is defined by the characteristic equation $\ddot{x} = \frac{q}{m}(\dot{x} \times \vec{B})$ where $q$ and $m$ are the particle charge and mass respectively, and $\vec{B}$ is the magnetic field. Because acceleration must therefore be perpendicular to the current velocity of the particle, the magnetic field causes a circular path of motion around the field lines. Parallel motion is caused by an external force to the magnetic field, most commonly an electric field running parallel to the magnetic field. The interactions between the magnetic and electric fields, along with gradients and curvatures in the magnetic fields, also produce various drift motions, per-
Figure 1.3: The path of a propagating whistler-mode wave through the radiation belt. From Lauben (1998)

perpendicular to the magnetic fields. Figure 1.4 illustrates these three types of motion in the radiation belts.

As an electron moves polewards along the magnetic field, it will encounter increasingly stronger magnetic field, which cause the gyration velocity to increase. Because the magnetic field cannot change the overall kinetic energy in an electron, the parallel velocity must reduce as the magnetic field grows stronger, due to forces caused by the spatial gradient of the magnetic field. When the parallel velocity hits zero, the electron’s energy will then be contained entirely in gyration. The location this happens in is known as the particle’s “mirror point”. The gradient forces will then reverse the parallel velocity, sending it on a trajectory back towards the equator. This property is known as the “mirror force”. The radiation belt consists of particles continuously going through this pattern of motion, constrained by the magnetic field.

Particles with higher parallel velocities at the equator will have a mirror point further down along the magnetic field. Some particles will have a high enough equatorial velocity
parallel to the magnetic field that the mirror point is located near or below the top of the Earth’s atmosphere at ~100 km. Instead of mirroring, these particles are absorbed into the Earth’s ionosphere and are lost from the radiation belts altogether. The ratio of the parallel and perpendicular velocities at the equator determines whether or not a particle’s trajectory results in precipitation. This ratio can be expressed in terms of a pitch angle, where

\[ \frac{v_{\text{perpendicular}}}{v_{\text{parallel}}} = \tan(\theta) \]

and \( \theta \) is the pitch angle. At pitch angles smaller than the critical angle \( \theta_c \), the parallel velocity will be large enough that the particle will be lost. The critical angle can be expressed in terms of the L value with the following relation:

\[ \sin^2(\theta_c) = \frac{1}{\sqrt{4L^6 - 3L^5}} \]

The range of pitch angles less than the critical angle is known as the loss cone. For the L=2 shell, the critical angle is roughly 16.3°.

When whistler-mode waves enter the magnetosphere, the plasma conditions force them to assume right hand circular polarization. Particles whose gyro-motion is in resonance with the incoming wave experience a constant electric field. Figure 1.5 illustrates this phenomenon, known as gyroresonance. More precisely, gyroresonance occurs when the
parallel velocity \( v_z = \frac{\Omega_e - \omega}{k_z} \), where \( \Omega_e \) is the electron gyrofrequency equal to \( \frac{eB_0}{m} \), and \( B_0 \) is the background magnetic field. Under this condition, the phase of electrons in gyromotion becomes bounded. Electrons in gyroresonance experience an acceleration in their parallel velocities, raising or lowering their pitch angle and potentially shifting them into the loss cone. The result is an increase in electron precipitation, ultimately observed in the LEP Event.

1.3 VLF remote sensing

LEP events have been observed directly through satellites (Voss et al., 1984) (Inan et al., 2007), but they are more commonly observed through ground-based VLF measurements (Helliwell, 1965), using the technique of VLF remote sensing.

When electromagnetic waves dislodge particles from the radiation belt, the particles precipitate into the ionosphere, changing the electron density in a geographic area. The region of the ionosphere from 60 km to 90 km, also known as the D region, reflects several modes of VLF waves. The successive reflections allow the VLF waves to travel long distances, well over the horizon, under a propagation condition known as the Earth-ionosphere waveguide (EIWG) (Cummer et al., 1998) (Thomson, 1993).
For a remote receiver observing a particular VLF signal from a distant transmitter, perturbations in the signal generally correspond to changes in the ionosphere somewhere along the transmitter-receiver path. Because LEP events result in a disturbance in the D region of the ionosphere, they can be detected using VLF receivers placed in locations such that the transmitter-receiver path passes through the disturbed region.

For a received signal, the perturbation due to the LEP event typically occurs 100-200 ms after the initial lightning stroke due to the distance the wave must travel, and because the whistler wave dispersion relation results in a wave speed much slower than the normal speed of light. Chemical processes of electron attachment, recombination, and detachment in the ionosphere cause the signal to return to a steady state over a period of 30s to several minutes (Pasko & Inan, 1994).
1.3.1 VLF transmitters and receivers

A limitation for using VLF remote sensing is the availability of narrowband VLF signals. Because VLF wavelengths are 10-100 km, VLF transmitters must be physically large and energy-demanding (Watt, 1967). There are 5 transmitters most useful across North America, Puerto Rico, and Hawaii, all operated by the US Navy. Their frequencies and call signs are described in Table 1.1. They are designed as top-loaded dipoles with bandwidths in the range of 100 Hz, and radiated power between 400 and 1000 kW.

Table 1.1: VLF transmitters operating in the United States, by location and frequency.

<table>
<thead>
<tr>
<th>Call Sign</th>
<th>Location</th>
<th>Frequency (kHz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NLK</td>
<td>Jim Creek, Washington</td>
<td>24.8</td>
</tr>
<tr>
<td>NML</td>
<td>LaMoure, North Dakota</td>
<td>25.2</td>
</tr>
<tr>
<td>NAA</td>
<td>Cutler, Maine</td>
<td>24.0</td>
</tr>
<tr>
<td>NPM</td>
<td>Lualualei, Hawaii</td>
<td>21.4</td>
</tr>
<tr>
<td>NAU</td>
<td>Aguada, Puerto Rico</td>
<td>40.75</td>
</tr>
</tbody>
</table>

To detect the signals generated by these transmitters, we use the Georgia Tech Low Frequency (LF) Radio Group’s network of VLF receivers. Each receiver collects the VLF signals using two wire-loop antennas, designed to capture the two horizontal components of the magnetic field. The two antennas are perpendicular to one another, such that one is oriented along the North/South axis, and the other along the East/West axis (Cohen et al., 2010).

The VLF signals travel as elliptically-polarized waves. This means that over one period of the wave’s propagation (a time equal to 1/frequency), the direction of the magnetic field will rotate, with the full range of magnetic field directions forming an ellipse. The axes of this ellipse will typically be roughly aligned with the radial (direction of wave propagation) and azimuthal (perpendicular to wave propagation, parallel to earth’s surface) directions. The cross-looped magnetic field antenna design of the VLF receiver allows for the signal to be decomposed into its radial and azimuthal components.
This thesis uses data from eight VLF receivers placed around the continental United States and in Puerto Rico. Paired with the transmitters, we have transmitter-receiver network that allows us to observe the ionosphere over a large section of the continental United States. Figure 1.7 displays the locations of the transmitters and receivers, as well as, in the purple lines, the transmitter-receiver paths. A disturbance in the ionosphere over any part of these paths can, in principle, be detected at the receiver.

1.3.2 Lightning data

LEP events, as their name suggests, originate from high intensity lightning strokes. Lightning is constantly occurring throughout the world, and lightning strokes release unique broadband VLF radio signatures (Uman & McLain, 1969) that can propagate at global scale distances. Using a network of detection antennas, the National Lightning Detection Network (NLDN) has been a source of lightning data in the continental United States since
Our investigation of LEP events begins with a search of lightning strokes with a peak current with an absolute value greater than 100 kA that occurred at nighttime, defined at an 80 km altitude, from November 4, 2017 to August 13, 2019. Previous research by Gołkowski et al. (2014) suggests 100 kA is a threshold required to produce detectable LEP events, although it’s likely that at least some strokes below 100 kA may also have triggered LEP events. LEP events are overwhelmingly a nighttime phenomena, as the ionospheric attenuation of whistler waves is much higher during the daytime (Němec et al., 2008).

The region of the ionosphere disturbed by LEP events has, in the northern hemisphere, a northwards displacement from the lightning stroke. The extent of this displacement varies at different latitudes, and continental United States typically ranges about 6-8 degrees in latitude (Lauben et al., 1999). We approximate this at 700 km, and we set a 400 km radius from the displacement point to account for variance. We term this region the “area of disturbance” associated with the lightning stroke, and filter our search of lightning to strokes where the area of disturbance intersects with a transmitter-receiver path in our network.

This initial search, carried out from November 4, 2017 to August 13, 2019, results in 157,780 candidate lightning strokes. Across all transmitter-receiver paths that intersect the hypothetical area of disturbance produced by these strokes, there are 755,524 narrowband VLF samples in which an LEP event in theory may be visible.

This initial set of lightning strokes provides us with a database of candidate events. As a given area of disturbance from a lightning stroke can intersect with multiple transmitter-receiver paths, each candidate event will have one or more candidate samples of VLF data.

### 1.4 Contributions

This thesis reports the following contributions towards the research of lightning-induced electron precipitation:
• We demonstrate the use of machine learning techniques to assemble a large database of LEP events via automatic classification. We document the design and training of an artificial neural network (ANN) classifier, building on previous networks trained on similar VLF disturbances, and show a reasonable accuracy.

• We use our database of events to obtain previously unexplored patterns in LEP occurrence and behavior, such as the tendency for LEP events to exclusively appear as phase changes in the VLF signals rather than amplitude changes.

• Using direct measurements of the radiation belts from the Van Allen Probes, we show that days with a high occurrence of LEP events have a statistically significant difference in electron distribution at the corresponding regions of the radiation belt.

• Through modeling techniques, we demonstrate a linear relationship between the number of electrons precipitated from a given LEP events, and the magnitude of the perturbation in the VLF signal. We show that this relationship generally holds under different assumptions for the background ionosphere. This means that a measurement of VLF perturbation can be used to estimate the total precipitation count of the corresponding LEP event.

• We provide an estimate for the total annual electron precipitation caused by lightning-induced whistler waves, and show the relative impact of this phenomenon compared to other radiation belt loss mechanisms.
CHAPTER 2
DATABASE OF LEP EVENTS

Our work towards detecting and analyzing LEP events at a large scale builds on our previous work, in Pailoor and Cohen (2022), in assembling a large database of early/fast events. This involved the training of an artificial neural network (ANN) classifier, which was trained on a manually labeled sample of early/fast events.

Due to the similar waveforms at a low resolution level, we employed a transfer learning approach by using the early/fast event classifier to assist with collecting a large training set of LEP events. This allowed us to develop a classifier for LEP events with a greatly reduced amount of manual labeling needed.

2.1 Past work: Early/fast detection

Early/fast events are a disturbance in the D region of the ionosphere caused by direct coupling between a high intensity lightning stroke and the surrounding ionosphere (Armstrong, 1983). Their name originates from the manner in which VLF signals are perturbed, with the abrupt change in the signal occurring within 100 ms of the lightning stroke (“early”) and the onset of the full disturbance being within 1 s (“fast”). 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 (2017) 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 10-100s driven 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.
The occurrence and behavior of early/fast events is dependent on the specific conditions of the lightning stroke, as well as the geometry of the transmitter, receiver, and stroke location. To account for these variables, we assembled a large database of events, and trained a neural network classifier to automatically detect events based on the VLF waveform.

2.1.1 Database of early/fast events

Starting from September of 2017 and running to the end of June 2018, we used the NLDN (Cummins et al., 1998) 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. This criteria also excludes many LEP Events, as LEP Events undergo a polewards displacement of several hundreds of kilometers north of the lightning stroke location. However, there is still a possibility of LEP Events falling alongside the transmitter-receiver paths. We have excluded samples with a scattering angle, defined as the angle between the transmitter to lightning stroke azimuth and the lightning stroke to receiver azimuth, greater than 90 degrees. This excludes potential back-scatter events, which are a rare but not fully understood phenomena (Marshall et al., 2006).

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 (Inan et al., 1988). For each path within this range, we extracted a sampled window of narrowband data.

Figure 2.1 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.

300,355 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 300,355 samples resulted from 91616 lightning strokes.

2.1.2 VLF data format

The transmitters make use of a 200 baud Minimum Shift Key (MSK) modulation scheme. Here, frequency variances of $\pm 50$ Hz (one fourth of the baud rate) from the center frequency define a bit of "1" or "0" being communicated. In MSK communications with this frequency variance, the transmitter uses a bit period of 5 ms, within which the phase rises or falls by 90 degrees. This serves to limit the overall bandwidth by ensuring continuous phase during the bit sequence. By decoding this modulation and removing the modulated phase shifts, we can recover an effective continuous wave (CW) phase, as if the transmitter were sending a monotone single frequency signal. After removing ambiguities in the phase, we can describe the signal using the four components of an elliptically polarized wave. This
Figure 2.1: Hypothetical event occurring in upstate New York. Image shows several intersecting transmitter-receiver paths.

combines the amplitude and phase data from both the E/W and N/S antennas into the Major Axis, Minor Axis, Tilt Angle, and Starting Phase of the incoming signal (Gross et al., 2018). Figure 2.2 shows the structure of the polarization ellipse, where the vertical and horizontal axes represent the orientation of the receiver’s loop antennas, typically in the north-south and east-west directions, and the red axes shows the direction away from $\hat{r}$ and orthogonal to $\hat{\phi}$ the source. The blue curve shows an example of how the magnetic field might oscillate given the amplitudes and phases of the two components, which is typically an ellipse. The four components of the ellipse are then extracted: (1) Major axis, or the longest diameter of the ellipse, (2) Minor axis, of the shortest diameter of the ellipse, (3) Tilt angle, which captures the rotation of the major axis counterclockwise from the $\phi$ direction, and (4) Start phase, which captures where along the ellipse is the magnetic field at $t = 0$, and which direction it rotates.

The receivers collect VLF narrowband data in both a high resolution (50 Hz) and low
resolution (1 Hz) format. Although the high resolution data contains useful information about the onset delay and onset time that characterize early/fast events and can help distinguish them from related phenomena, we ultimately chose to primarily use low resolution data for detection. The reason for this is that the high resolution data contains a large amount of noise, due to the signatures of other lightning strokes. These VLF signatures, also known as sferics, are very short in duration, but cover a broadband range of frequencies, and reach values far higher than the background signal. This can make it difficult to detect changes in the ionosphere, which are more visible from the “background” narrow-band data from the VLF transmitter. Because the event occurs over a period of several tens of seconds, using the high resolution data for an ANN classifier would also require a far larger training set.
2.1.3 Artificial Neural Network classifier

Many of the 300,355 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. An example of this is shown in Figure 2.3. The 40 second window provides sufficient window to visually observe the perturbation, in this case occurring largely on the Minor Axis, Tilt Angle, and Starting Phase channels. In contrast, Figure 2.4 shows a ”non-event”. While this sample shows a strong sferic at the $t = 0$ mark, the lightning stroke does not appear to have impacted the ionosphere over the transmitter-receiver path sufficiently to affect the signal afterwards.

![Figure 2.3: Example of an event during manual classification process. This event occurred on April 10, 2018, observed at the Baxley receiver reading the NAA signal.](image)

The 1000 samples were then evenly divided between training and test data. We used a training set of this size given the relatively low number of features in our data, with each sample only containing 160 data points. We evenly divided the samples between training and test data, so that our validation results could provide a more accurate picture of the
Figure 2.4: Example of an non-event during manual classification process. This sample occurred on January 28, 2018, taken from the Baxley receiver reading the NLK signal.

The architecture of the network employed a series of three fully connected layers, with an additional fully connected layer used for prediction. These layers contained 1000, 1000, and 2000 nodes respectively, and used a linear rectifier (ReLU) activation function. The prediction layer used a softmax activation function. The input layer takes in 40 seconds of data from each of the four channels, resulting in each sample containing 160 data points.

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 low false negative rate. In order to 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 2.1.

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...
Table 2.1: Test results for classifier

<table>
<thead>
<tr>
<th>Event</th>
<th>Non-Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classified as Event</td>
<td>116</td>
</tr>
<tr>
<td>Classified as non-Event</td>
<td>29</td>
</tr>
</tbody>
</table>

later. Nearly all of these do not have an event exactly 120 seconds later, since events are sparse in time, though a few may have another 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, or a ”noise floor” when applying the classifier to real data. For example, out of the 91,616 candidates, the classifier detected a total of 6500 events, or 7.1%. This is much higher than the 0.63% false positive probability indicating that most (>96%) of the selected events are likely real.

2.1.4 Applicability to LEP events

Despite different mechanisms, early/fast events display perturbations of a similar magnitude and duration as LEP events(Gołkowski et al., 2021). At a low resolution, key differences in the waveforms, such as the onset delay between the lightning stroke and perturbation, are typically not clearly visible. However, the expected path geometries in relation to the causative lightning stroke will be different, as the region of disturbance for LEP events is displaced from the lightning stroke.

Therefore, we can use our early/fast event classifier as a starting point for LEP detection, by starting from a different list of stoke-transmitter-receiver combinations.

2.2 LEP Classification

As discussed earlier, our dataset of candidate events is assembled using narrowband VLF data samples, paired with high intensity lightning strokes in the continental US, filtered
to include samples with diurnal and geographic conditions that could plausibly result in events.

Not all lightning strokes, even intense strokes in the ideal geography, generate a noticeable LEP event. Due to the noisy nature of VLF data, the perturbations associated with large-scale changes in the ionosphere are difficult to analytically separate from those caused by sferics, background static, or instrumentation artifacts. Previous case studies have relied on manual classification, which results in a limited sample size with narrow geographic and temporal scope. It would be too time consuming to rely on manual classification over the entire set of candidates within the continental US over even a period of a single year, so we aim to create an automated process for event detection. Recent work has shown promise in the use of Artificial Neural Networks (ANNs) to classify VLF waveforms (Wang et al., 2020) (Konan et al., 2020).

We present our approach to training and deploying an ANN classifier for LEP events, using our early/fast event classifier as a starting point.

2.2.1 Training set preparation

As stated in Section 1.3, the AWESOME receivers produce VLF data in two resolutions, a high resolution sampling of 50 Hz, and a low resolution sampling of 1 Hz. We show a comparison of these resolution formats for the minor axis variable in Figure Figure 2.5. In the high resolution format, the onset duration is more clearly visible, occurring between 300 to 800 ms of the lightning stroke. Due to the noise, however, the precise time of the onset is difficult to determine clearly. Furthermore, other lightning sferics and short-term ionospheric disturbances are visible and shape the data. Because the full event lasts over a period of tens of seconds, up to several minutes, we choose to primarily use the low resolution data in the interest of minimizing noise, but do use the high resolution data to check that we have not included many events with small onset delay (which might indicate an early/fast event).
Figure 2.5: An LEP event, shown through the minor axis of the polarization ellipse. Note the delay between the lightning sferic and the onset of the perturbation. This event occurred on April 3, 2019 detected at the Delaware receiver from the NAA signal.
As with early/fast events, we use the polarization ellipse format for the data, and initially set our sample width to span from 10s before the lightning stroke to 30s after. This means we have four channels of input - major axis, minor axis, tilt angle, and start phase - and each channel contains 40 data points, or features, 1 for each second of the sample. Therefore, each sample contains 160 features. Each channel of the sample is normalized such that the maximum value is scaled to 1, and the minimum value is at 0. This is because background variations in the ionosphere, and different VLF signal strengths, mean that the raw values of different samples will have very different averages and ranges. Additionally, for the major axis, we convert the data to decibel scale before normalization, so that effects of sferics and background ionospheric fluctuations are diminished. For the start phase, we apply a phase unwrapping prior to normalization, so that wrap-arounds when the phase value is near the 360 degree mark are not misclassified as large perturbations in the signal.

We begin by applying a simple perturbation detector. Since a defining characteristic of LEP Events is an initial sharp perturbation following a lightning stroke, we can examine the data preceding and proceeding the stroke to see if such a perturbation has occurred. We can define this using the variance and average of the data points immediately prior and immediately after the stroke occurrence. For our perturbation detector, we define the incident field average using the average of data points ranging from 12 to 2 seconds before the stroke, and define the post-stroke average using the average of data points 2 to 12 seconds after the stroke. We exclude the time points including the stroke and directly surrounding it to avoid the waveform of the sferic distorting the average or variance.

The perturbation detector serves as a quick way to initially filter the dataset of candidates to remove samples with no visible perturbation, or highly noisy samples where meaningful patterns cannot reliably be observed. However, it is not sufficient as a classifier.

After filtering out candidates that fail the perturbation detector, we randomly select 500 samples, and manually examine them for waveforms indicative of an LEP event. We find
that 54 of them are events, and 446 are non-events. This ratio suggests that the population, even when filtered through a perturbation detector, is still heavily imbalanced towards non-events. In order to effectively train the classifier, we need a much larger number of event samples. To achieve this, we make use of our existing early/fast classifier.

We run our early/fast classifier over the database, and select 500 samples that the classifier labels as events. We find within this set of samples that 298 are actual events, and the other 202 are false positives. This suggests that the early/fast classifier is a considerable improvement over a random selection of samples, but has significant shortcomings in detecting LEP events. In addition to the false positives, we must consider the possibility that the early/fast classifier may be missing events. To address this, we select 500 additional samples that the early/fast classifier had labeled as non-events. We find that 7 of these samples are events, and the other 493 were (accurately) labeled non-events.

Putting these three sets of 500 samples together, we have a dataset of 1500 samples, which includes 359 events and 1141 non-events. This provides us with a much larger number of event samples than we would have obtained by manually labeling 1500 random samples. However, we note that the majority of events in this dataset are those detected by the early/fast classifier. Although false negatives were rare, we must consider a possible source: the differences in waveforms between early/fast events and LEP events. Early/fast events are more likely to result in a positive amplitude perturbation (signal strength increased), while LEP events can result in either a positive or negative perturbation. To correct for this, we take each data channel after the preprocessing, and ”invert” the data, resulting in a new sample with values of ”1-x”, where x is the original sample after normalizing. This inverted sample is then added as an additional sample in the dataset, effectively doubling the dataset size and training the classifier to recognize both positive and negative polarity events.

Although we initially used an early/fast event classifier in the data collection process, our aim is to train a new classifier, built from LEP events, that can improve on the shortcom-
ings of the early/fast classifier and provide an accurate classification over the full dataset of candidates.

2.2.2 Neural network structure

The early/fast classifier was built using a traditional artificial neural network (ANN) structure, using multi-layer perceptrons, also known as ”fully connected layers”. A fully connected layer acts upon its inputs through a linear transformation, and a nonlinear activation function. In this case, we chose our activation function to be a Rectified Linear Unit (ReLU), equivalent to \( f(x) = \max(0, x) \). ReLU has long been a popular activation function in deep learning due to sparse activation, scale invariance, and efficient computation of gradients (Glorot et al., 2011).

Recent research has indicated that that 1D Convolutional layers can improve classifier accuracy (Kiranyaz et al., 2021). The mechanism for this improvement comes from the use of convolutional kernels to match the shape of specific waveforms in a 1D signal. The use of 1D convolution has been growing in recent years in deep learning applications, beginning with efforts to classify and detect Electrocardiogram (ECG) signals (Kiranyaz et al., 2014), and later to applications such as structural damage detection using wireless sensors (Avci et al., 2018) and fault detection in modular multilevel converters (Abdeljaber et al., 2018).

To compare the effectiveness of different network structures for our purposes, we trained and tested three different networks: a predominantly convolution-based network, a fully connected network, and a ”hybrid” network including both convolutional and fully connected layers. Note that even in the convolution-based network, a fully connected layer is still used to sort the data into one of the two result classes.

The classifier’s performance can be measured using a metric called the loss function, which is a function that assigns a penalty to incorrect classification outputs. The classifier ”learns” by carrying out a gradient descent optimization in order to minimize the the loss function output for the training set. Initially, the loss function outputs for both the training
and the test set will decrease as the classifier attempts to minimize loss for the training set. However, because the classifier is not learning from errors from the test, it will eventually ”overfit” (Tetko et al., 1995) to the specific characteristics of the training set, and the loss function output for the test set (test loss) may increase. We use a cross entropy loss function, which will be described in more comprehensive detail further down.

We present the comparison of the test performance over the course of classifier training in Figure 2.6. The largely convolutional network (CNN) performed significantly better than either the fully connected network (FC) or the hybrid network. Note that for the fully connected network and the hybrid network, the test loss begins overfitting a certain number of training batches (∼2200 and ∼2050) respectively. Also note that while the CNN performed better than its rivals, it converged significantly slower, and does not appear to have reached its peak after 2500 training batches.

2.2.3 Effect of sample width

The widths of the samples in time also have an effect on the classifier performance. Our database stores low resolution data ranging from 40 seconds before the stroke to 120 seconds after. The full 160 second long sample is impractical for use in the classifier, as the
main event behavior occurs very close to the t=0 region. Additionally, because the data preprocessing normalizes each channel between 0 and 1, other anomalies and events that occur before and after the stroke can greatly distort the sample.

The early/fast classifier used a 40 second window, ranging from 10 seconds before the stroke to 30 seconds after. For the LEP event classifier, we present variations of that use 40 second, 60 second, and 80 second sample widths, ranging from t=-10:30, t=-20:40, and t=-30:50 respectively. We find that a 60 second sample with produced the best results, as shown in Figure 2.7.

2.2.4 Loss function weighting

Finally, we note that the classifier’s loss function can be subject to several chosen parameters. In this classifier, we have used a cross entropy loss measurement, which research has indicated provides robust and accurate optimization for datasets where the labeling process may be subject to error (Zhang & Sabuncu, 2018). Rather than assign a strict, discrete class to each sample passed through, the classifier provides an output vector with dimension 2, with each dimension representing one of the two classes (Non-Event and Event). The cross entropy of an output $x$ corresponding to a sample with a label $y$ is as follows:
Figure 2.8: Positive Predictive Value (PPV) versus True Positive Rate (TPR). Labels of each point indicate the classifier weight assigned to Events. Non-Event weight was standardized to 1.0.

\[ l(x, y) = -w_y \log \frac{\exp(x_y)}{\exp(x_0) + \exp(x_1)} \]

For example, if we have an LEP Event (y=1) which the classifier produces the output [0,1], the loss function should return the value \(-w_y \log \frac{e^1}{1+e} = -w_y(1 - \log(1 + e)) \approx 0.31w_y\). However, if the classifier had instead returned the output [1,0], the loss function would return \(-w_y \log \frac{1}{e^1} = w_y(0 - \log(1 + e)) \approx 1.31w_y\). We see that \(w_y\) acts as a weight to each class, allowing us to decide how much the classifier should “punish” getting wrong each class. By setting a relatively higher weight for Events than Non-Events, we can trade off the True Positive Rate (TPR, defined by \(\frac{TP}{P}\)) and the Positive Predictive Value (PPV, defined by \(\frac{TP}{PP}\)). Figure 2.8 shows this trade off, with the green dots. If the relative weight assigned to Events is low, we can achieve a high Positive Predictive Value (most samples classified as events will really be events), but this comes at a cost of a low TPR (most events will be missed). Likewise, by increasing the weight assigned to events, we can capture most events and achieve a high TPR, but at the cost of reducing our PPV (a sample classified as an event is less likely to really be an event).

Effectively, this allows us to tune our network, and choose whether we want a classifier
that will identify nearly all events, but also collect some false positives, or one that will be selective in how it classifies events, reducing false positives but potentially missing events.

Ultimately, we set our weight value equal to 2.0 to balance the TPR and PPV effectively, using the convolutional network described previously as well as preprocessing our data samples to have a 60s with. Combining these choices, we achieved a network that lowered test loss to 0.335, and achieved a TPR of 0.92 and a PPV of 0.79. As TPR is equal to 1-FNR, this represents an improvement of the PPV by 32.6 from the original classifier, and a lowering of the FNR by 38.5%.

2.2.5 Comparison to alternative approaches

The process described so far has involved a transfer learning approach, using a classifier originally trained on early/fast events to detect candidate samples, which are then added to the training dataset to train a new classifier. Because of the low occurrence of LEP events within the broader population of candidate samples, if we were to start from scratch it would require considerably more manual classification in order to get a statistically significant number of events.

To illustrate this, we manually labeled an additional 1000 samples randomly selected from the population of candidate samples. Adding to our existing pool of 500 random samples, this gave us a dataset of 1500 samples, equal in size to the one produced with the assistance of the early/fast classifier. As shown in the red dots in ??, this classifier had a significantly worse performance over a wide range of weight values. The blue dot in the figure shows the performance of the perturbation detector on its own.

2.2.6 High resolution analysis

We note that the survey parameters used to search for LEP events in some cases allow for cases where both the stroke, and the assumed region of perturbation, are in close proximity to the transmitter-receiver path. This is particularly more likely in paths that span over a
greater North-South range. In this cases, there is a chance that detected perturbations may be the result of Early/Fast events, a phenomena that at low resolutions appears similar to LEP events.

To gauge the incidence of Early/Fast events in our database, we examined the corresponding high resolution samples of 145 events. The high resolution data is sampled at 50 Hz. At this resolution, it is possible to detect the onset delay between the initial lightning sferic, and the perturbation that occurs as a result of electron precipitation. This onset delay is the result of whistler-mode VLF waves having to propagate large distances along the magnetic field lines in order to scatter energetic electrons in the radiation belts, and typically ranges from 200 ms to 2.5 s (Peter, 2004).

In our survey of high resolution samples, 59 out of 145 had a clear, identifiable onset delay, while 8 had an identifiable perturbation within 200 ms of the lightning stroke. 78 of the samples were too noisy to clearly distinguish the precise time of perturbation. Figure ?? is an example of an event, with the major axis shown, that has a clearly identifiable onset delay.
The ability to accurately classify a dataset of this size allows us to map in detail the occurrence of events across different points of time, allowing us to find key patterns and relationships to direct observations of precipitating electrons from satellites.
While a database of LEP events provides key information about the occurrence of electron precipitation from lightning strokes, it is not sufficient to provide a comprehensive picture of the extent to which lightning-induced whistler waves impact the radiation belt.

Because we lack direct measurements of the radiation belts at the exact times and locations corresponding to the LEP events as observed on the ground, we must rely on modeling to describe the processes of wave-particle interaction and precipitation into the ionosphere.

3.1 Past modeling work

LEP events have been the subject of many observational and modeling studies, largely analyses of one or a small number of individual LEP events. LEP events can be observed through ground-based VLF measurements (Peter, 2005) (Clilverd, 2004), ground-based optical emissions (Doolittle & Carpenter, 1983) (Marshall et al., 2011), and on satellites (Inan et al., 2007).

3.1.1 LEP Occurrence Rate

While LEP events are well understood on an individual level, quantifying the role of all lightning in radiation belt dynamics is more difficult. A number of different approaches have been presented. One of the key limitations to these past efforts have been lack of knowledge of how often LEP events occur.

Burgess and Inan (1993) estimated that in a given minute, LEP events result in 1e-6% (one millionth of a percent) of radiation belt flux being lost. This study reached this estimate by assuming lightning generated whistlers on average occurred six times a minute, based on data from Laaspere et al. (1963), and using satellite observations from Voss et
al. (1998) to provide an estimate for the average fraction of electrons removed in a given whistler. This estimate would suggest that lightning-generated whistlers represent a similarly significant mechanism as the plasmaspheric hiss (Lyons & Thorne, 1973). These ground-based measurements of whistler waves, however, relied on the assumption that whistler waves travel within concentrated ducts of the magnetic field lines, and therefore can be reliably observed at the conjugate point on the Earth. Lauben et al. (1999), however, showed that whistler waves can also travel in a non-ducted mode, and as a result may not be detected on Earth. Using this information, Rodger and Clilverd (2002) estimated the role of LEP events by, instead of starting with the occurrence rate of ground-detected whistlers, using the occurrence rate of LEP events, which was estimated at 0.3 events/minute. This study used the AE-5 Inner Zone Electron Model (Teague & Vette, 1972) to approximate the number of electrons in a flux tube at L=2.23, and calculated that LEP events at this L shell were a more significant loss mechanism than all others.

While these studies used a combination of ground-based LEP detection as well as case studies from satellite data, other approaches have focused on attempting to model the propagation and energy distribution of whistler waves, to gain a more precise insight towards the effects of whistlers on the radiation belt electron distribution.

3.1.2 Whistler wave raytracing

One prominent effort to model the impact of lightning-generated whistler waves in the radiation belt was described by Abel and Thorne (1998a). This approach used the HOTRAY code (Horne 1989) to trace the propagation paths of 4.5 kHz lightning-generated whistlers. This model also assumed a wave intensity of 10 pT, uniform throughout the 1.2 < L < 4 region of the radiation belts, and an occurrence rate of six whistlers per minute, based on estimates from Burgess and Inan (1993). This approach did not incorporate Landau damping, which reduces the intensity of certain whistlers in the outer regions of the radiation belts (Horne & Thorne, 1994)(Sazhin, 1991). Next, the model uses a formula given by
Lyons et al. (1971) to the bounce-averaged pitch angle scattering diffusion constant for the full range of pitch angles affected. Finally, in Abel and Thorne (1998b), the model derives the changes to the electron distribution using the approach from Walt and Farley (1976).

Bortnik (2003) built on this work to provide a more accurate model incorporating the distribution of whistler wave power within magnetosphere. This approach used the Stanford VLF Ray Tracing Code from Inan and Bell (1977) to trace the paths of whistlers, incorporating Landau damping using the formulation from Brinca (1972). Rather than assuming a single frequency with uniform power density, this model used an initial power distribution derived from Uman (1984)[p. 6], given as:

\[ S(\omega) = \frac{1}{Z_0} \left( \frac{\mu_0 h_e I_e}{2\pi} \right)^2 \left( \frac{\sin \theta}{R} \right)^2 \frac{\omega^2(a-b)^2}{(\omega^2 + a^2)(\omega^2 + b^2)} \]

Following up, Bortnik et al. (2006) developed a wave-particle interaction model to calculate the precipitated differential electron flux as a function of time and energy. This approach implemented the gyroresonance equations from Bell (1984), which provide a formula for \( \frac{d\alpha}{dt} \) as a function of the electric field and the electron momentum, where \( \alpha \) is the equatorial pitch angle. For each time, frequency, latitude, and resonance, the model calculates the resonance velocity using the formula from Chang and Inan (1983), and using this calculates the total pitch angle change \( \Delta \alpha \).

3.1.3 Modeling VLF signal perturbations

To model LEP observations end-to-end, we need a model of both the ionospheric disturbance caused by precipitated electrons, and another of the subionospheric VLF wave propagation through the Earth-ionosphere waveguide passing underneath the LEP-disturbed ionosphere.

The primary region of the ionosphere relevant to subionospheric VLF propagation is the D region (60-90 km altitude). The D region is too low for satellite measurements, and too high for balloon measurements. However, Wait and Spies (1964) derived a two-
parameter model for this region, using a combination of data from sounding rockets and VLF propagation measurements. This model presents electron density as the following function of altitude:

\[ N_e(h) = 1.43 \times 10^{13} e^{-0.15h} e^{\beta(h-h')} m^{-3} \]

Where \( h' \) and \( \beta \) are variable parameters, roughly describing the height and gradient of the D region respectively. There are other models that parameterize the model with more than two parameters, which are likely closer to the physical reality of the D-region 
McCormick2021, however when using single-frequency transmitters there is not sufficient information to constrain the D-region beyond 2 parameters, as the problem becomes ill-posed. As such, we will continue to use the Wait-Spies ionosphere here.

Poussard and Corcuff (2000) presents an early example of LEP modeling, treating the disturbed region as a triangularly ramped section where \( h' \) and \( \beta \) both decrease linearly towards the center of the disturbance. To simulate the perturbation of VLF signals, the model uses two approaches, a mode solving algorithm built from work by Morfitt (1980), and a Finite-Difference Time-Domain (FDTD) method using code from Berenger1994. Mode solving algorithms work by dividing the the cross-section of the ionosphere traversed by the VLF signal into slabs, and solve for the the different modes of VLF waves in each slab. FDTD methods, on the other hand, work by considering the total electric and magnetic fields, and numerically simulating their progression by implementing a discretized version of Maxwell’s Equations.

Peter (2007) attempted a more rigorous approach to modeling the ionospheric density enhancement, by taking the pitch angle and energy distribution of the precipitated electrons, and using a Monte Carlo simulation from Lehtinen et al. (2001) produce a map of energy deposition as a function of L shell and altitude. The model calculated electron density changes based on the assumption from Rees1963 that one electron-ion pair is produced for every 35 eV of deposited energy. Peter (2007) modeled the VLF signal perturbation using
3.2 LEP Combined Model

We now describe our modeling approach, building on previous work, in particular Peter (2007)'s effort to model the full process of an LEP event, from the whistler wave propagation in the radiation belt, to the wave-induced particle precipitation, to the precipitation deposition in the ionosphere, to the perturbation of subionospheric VLF signals propagating in the Earth-ionosphere waveguide. The aim of our modeling work is to connect the perturbations observed in our LEP events to the number and energy distribution of precipitated electrons.

We use the Whistler-Induced Particle Precipitation model (WIPP) from Bortnik et al. (2006) to carry out ray tracing for the whistler waves and calculate the precipitated electron flux. To calculate ionospheric density changes, we employ the Semi-Analytic Model (SAM) described below. Finally, we model the perturbation of subionospheric VLF signals using the Long Wavelength Propagation Code (LWPC)(Ferguson et al., 1989), a modesolving approach.

We illustrate an overview of our LEP modeling process in Figure 3.1 and Figure 3.2. Starting with the latitude of a lightning stroke, we run the WIPP code to produce a series of electron precipitation calculations over a range of L shells, as shown in Figure 3.1. We then feed each of these precipitation calculations into the SAM, to produce a table of ionospheric profiles. Next, in Figure 3.2, we show the process of modeling a specific event as observed from a given transmitter and receiver. We calculate the great circle path (GPC) between the transmitter and receiver, and discretize this path into separate sections. For each section of the path, we calculate the corresponding L shell that intersects with the ionosphere over the point. We then look up the disturbed profile for this section from our table of profiles, given the location of the lightning stroke. We combine the profiles over the full path, and use this as the input to the LWPC. The LWPC generates an estimate for the signal perturbation in
Figure 3.1: A flowchart showing the construction of our disturbed ionosphere table, built as a function of lightning latitude, L shell, and longitudinal displacement.

3.2.1 Whistler-Induced Particle Precipitation model

The WIPP treats each lightning stroke as a short, vertical current (8.5 kA) (Bortnik et al., 2006), and describes the VLF wave power at the bottomside of the ionosphere as a function of distance from the source. The WIPP uses data from Helliwell (1965) to calculate the attenuation for a whistler wave traveling through the nighttime ionosphere, resulting in a wave power specification at the topside of the ionosphere, 1000 km in altitude. This provides a "weight" for each ray as it is traced, using the Stanford Ray tracing code. The WIPP carries out ray tracing for 130 frequencies between 200 Hz to 60 kHz, generating for each frequency 41 rays over a spread of 20 degrees in latitude centered around the latitude of the lightning stroke. The model then interpolates between the adjacent rays, both in frequency and in latitude. This incorporates the Landau damping effect, as modeled by
Brinca (1972). In total, the model interpolates 120 million ray paths. The model integrates over these rays to produce calculations of the pitch angle changes for resonant electrons, for each L shell. In our implementation, we run this calculation for 101 L shells evenly spaced between 1.0 to 3.5. For each L shell, the model integrates over all frequencies over the 200-60,000 Hz range to calculate the differential electron flux over a 10 second period, sampled at 0.02s intervals. The energy range is from $10^{1.5}$ to $10^{7.5}$ eV, with 2000 logarithmically-spaced samples. Figure 3.3 shows one of these flux calculations, corresponding to a lightning center latitude of 30 degrees N and L shell of L=2.5.

In theory, this process can be repeated for every event in our database. However, this would be excessively time consuming and ultimately redundant. Instead, we run the WIPP model over a range of latitudes, and approximate each event by fitting it to a previously modeled latitude. We run the WIPP over a range of 18 center latitudes, starting from 15 degrees N to 66 degrees N, spaced every 3 degrees. In total, this gives us a lookup table of
3.2.2 Semi-Analytic Model

For the process of ionospheric deposition, we use a recently developed method known as the Semi-Analytic Model (SAM). The SAM uses a 6 species ionospheric chemistry model, taking in as inputs the differential electron flux along with a background ionosphere, as well as day/night status.

The SAM obtains the initial electron pair generation rate ($I_s$) using the following equation from Rees (1963):

$$I_s = \frac{F}{\Delta \xi_{ion}} \left( \frac{\xi_0}{r_0} \right) \frac{\gamma'[\frac{z}{R}]}{n_{\mu}(R)} n_{\mu}(z)$$

Where $F$ is the electron flux (obtained from WIPP and the Bortnik model), $\xi_0$ is the average initial energy of the particles, $\Delta \xi_{ion}$ is the energy needed to produce an electron-ion pair (35 eV), $z$ is the atmospheric depth and $R$ is the lowest depth of penetration, with a value of $4.57 \times 10^6 \xi_0^{1.75} g/cm^2$. $r_0 = R/\rho$, where $\rho$ is the atmospheric density at the lowest depth of penetration. $n_{\mu}$ is the number density of ionizable atoms at a given penetration depth. $\gamma'[\frac{z}{R}]$ is the energy dissipation function, as calculated by Grün (1957).
After generating the electron pair generation rate, the model begins computing the atmospheric chemistry dynamics. This makes use of the 6 species ionospheric chemistry model from Glukhov et al. (1992) to calculate the electron density changes over the full range of the D region (60-120 km) over the period of time desired.

For a background ionosphere, we use the two-parameter nighttime assumption from Wait and Spies (1964), setting \( h' = 85 \) and \( \beta = 0.4 \). The SAM generates electron density profiles, at 99 altitudes linearly spaced from 60 to 120 km. Each simulation runs from 0 to 5 seconds, in 0.02s increments. Figure 3.4 shows the total effect of the precipitated electrons on the density profile. Although the changes are relatively slight, they have the impact of lowering the height of the ionosphere, increasing the attenuation for subionospheric VLF waves.

We can observe the time evolution of the disturbance in Figure 3.5. Each curve shows the difference in electron density at a given time compared to the ambient profile. We can see that, at elevations above 90 km, the disturbance peaks by \( t=1s \). However, beyond this time, we can see the diffusion of electrons into the lower altitude regions of the ionosphere. Between \( t=2.5 \) and \( t=5 \), we see that the disturbance begins to recede at the lowest sections of the ionosphere, due to chemical recombination processes removing free electrons.

The 1812 precipitation distributions produced by the WIPP model allow us, for a given lightning latitude, to calculate the disturbed ionosphere profile over the range of L shells at the same longitude as the lightning stroke, which correspond to a range of latitudes along the line of longitude containing the stroke. This effectively captures the latitudinal variation of the ionospheric disturbance, but we must also calculate the longitudinal variation. To do this, we apply the scaling relationship from Bortnik (2003)[Equation 1] to each calculation precipitated flux. We run the SAM code for each of this new precipitation distributions. Together, this provides a table of disturbed ionospheres, as a function of lightning latitude, L shell, and longitudinal displacement.
Figure 3.4: Ambient and disturbed electron density profile for D region at L=2.5. Disturbed profile is the result of a lightning stroke at 30 N latitude, and shows the disturbance at t=5s.

Figure 3.5: Time evolution of the ionosphere disturbance. The x axis shows the change in the electron density compared to the ambient profile.
3.2.3 VLF Propagation Model

After modeling the changes in the ionosphere as a result of electron precipitation, we must simulate the VLF signal perturbations that result from scattering off this disturbed region. We employ the Long-Wavelength Propagation Capability (LWPC) model (Ferguson, 1998) for this. The LWPC is a mode-solving Earth-ionosphere waveguide model, which treats the surface of the Earth and the ionosphere’s D region as plates in a parallel-plate waveguide. The model breaks the propagation path into a series of segments that are assumed to be homogeneous. Using the "MODESRCH" algorithm from Morfitt and Shellman (1979), the model finds the propagation modes compatible with each segment. These modes are then connected between the segments using the mode-conversion algorithm from Ferguson and Snyder (1980).

To run the LWPC, we segment the transmitter-receiver path, and for each segment, assign an ionospheric profile from the table of ionospheres produced by the SAM. Unlike with the SAM and WIPP, we run the LWPC separately for each event in our database, as the each combination of transmitter, receiver, and stroke provides a unique geometry that cannot necessarily be translated from one another. For each event, we run an LWPC simulation for the ambient profile, followed by a simulation of the disturbed ionosphere every 0.4 seconds of the disturbance evolution, for a total of 25 runs.

The LWPC calculates the magnetic field amplitude and phase every 20 km along the path, for both the azimuthal and radial components of the magnetic field. We take these values at the distance corresponding to the receiver, and calculate the polarization ellipse using the method described in Gross et al. (2018), which yields a major and minor axis, a tilt angle, and a start phase.

We first calculate the tilt angle $\tau$:

$$\tau = \frac{1}{2} tan^{-1}\left[ \frac{2\gamma}{1 - \gamma^2} cos(\psi_0) \right]$$
Where
\[ \gamma = \frac{|-B_r|}{|B_\phi|} \]

and
\[ \psi_0 = \angle \left[ \frac{-B_r}{B_\phi} \right] \]

With \( \tau \) known, we can calculate the major axis and minor axis components of the magnetic field as follows:

\[
\begin{bmatrix}
B_{maj} \\
B_{min}
\end{bmatrix} =
\begin{bmatrix}
\cos(\tau) & -\sin(\tau) \\
\sin(\tau) & \cos(\tau)
\end{bmatrix}
\begin{bmatrix}
B_r \\
B_\phi
\end{bmatrix}
\]

Finally, we can obtain the start phase as simply \( -\angle B_{maj} \).

For each disturbed ionosphere run, we calculate the field perturbation as the difference between the field components for that run and those for the ambient run. For each of the field components, we choose the disturbed run with the maximum perturbation, and define this to be the modeled perturbation for the event.

### 3.3 Precipitation Calculations

As described in the last section, we have developed a method to simulate the resulting VLF signal perturbation from a given distribution of precipitated electrons. However, our goal is to invert the process, and to infer, as best we can, the electron precipitation given a VLF signal perturbation.

The electron precipitation is distributed unevenly over time, location, and energy, so it is not possible to infer the entire spatio-temporal precipitation structure without a very large number of observations of the same LEP event. As such, our goal here is to infer only the total number of precipitated electrons from each LEP events, as a function of energy. Since the spatio-temporal pattern is captured by the model, we can infer the rest of it with a single VLF observation. This process works as long as the spatio-temporal precipitation
Figure 3.6 provides an overview of the full process. Our model provides both a simulation of the signal perturbation and the total electron precipitation, from which we ultimately derive a relationship between the two. We then use this relationship, paired with our data for signal perturbation, to estimate total electron precipitation.

### 3.3.1 Path-integrated precipitation

We employ a similar method to Peter (2007) to define and calculate the total precipitation along a VLF transmitter-receiver path. We refer to it here as $\Gamma$, defined as follows:

$$
\Gamma = \int_{t_{\text{txlocation}}}^{t_{\text{rxlocation}}} \int_{0s}^{5s} \int_{30\text{keV}}^{1000\text{keV}} \Phi(E, t, l) dE dt dl
$$

Our method differs only slightly from Peter (Peter, 2007) in that we integrate over a
wider energy range. Peter (2007) integrated between 100 keV and 300 keV, while we integrate from 30 keV to 1000 keV. Our widened energy range is in light of recent work by Green et al. (2020) suggesting that lightning-generated whistlers can affect electrons up to 1 MeV, and other recent work indicating that precipitated electrons within as low as 30 keV may deposit significant energy in the D region (Marshall & Cully, 2020).

Figure 3.7 illustrates the process of calculating the precipitation metric for a given event. For a given lightning stroke, we calculate the integrated precipitation flux across a range of latitudes and longitudes, shown in the grid. We then integrate once more over the transmitter receiver path, using the points on this grid that overlap with the path.

### 3.3.2 Linear relation between precipitation and field perturbation

The precipitation metric $\Gamma$ serves as a simplified description of the electron precipitation, but in order to use our model results to obtain estimates for $\Gamma$ from the measured field perturbation, we must establish a one-to-one relationship between $\Gamma$ and the magnitude of
the perturbation in one or more of the field components.

We do this by turning once more to the Semi-Analytic Model and the LWPC. We choose a single event, in this case from an event observed from the NML-PARI VLF signal, originating from a lightning stroke at 32.60 N, 85.31 W. This stroke occurred on May 23, 2018 at 7:49:15 UTC. We adjust the input to the SAM, by scaling the precipitated electron distribution by various levels. Each of these scaled precipitation flux values will result in a different path-integrated precipitation metric.

For each scaling of the precipitation flux, we run a new LWPC simulation. We observe the resulting perturbation to each of the four components of the polarization ellipse.

Figure 3.8 shows how, as the path-integrated electron precipitation is increased, each of the polarization ellipse components respond. However, we cannot assume that this relationship is always the same, it may be ionosphere dependent, and as such we run this for four different ambient D-region ionosphere models representative of typical nighttime conditions at mid latitudes. For each of the four polarization ellipse parameters, there is a linear relationship between the path-integrated flux at least part way through as the \( \Gamma \) is increased. The start phase appears to be the most reliable metric, as it takes on a roughly linear relationship with the path integrated flux metric under a variety of ionospheres. The minor axis, on the other hand, appears to become extremely sensitive at higher electron flux levels, particularly for an ionosphere of \( h' = 88 \). This may be due a result of the radial component being much smaller than the azimuthal component, with LWPC often treating it as zero. The major axis and tilt angle are fairly linear until \( \Gamma = 1.5 \times 10^5 \) el/m. However, in our case, the major axis typically changes by \(<2\, \text{dB}\), and the tilt angle by \(<2^\circ\), so for practical purposes the LEP events we observe are within the linear ranges for these two parameters.

As a result, for each observed LEP event, we neglect the minor axis, and estimate the path-integrated electron flux using each of the other three parameters. We do this with the following relationship:
Figure 3.8: The relationship between path-integrated electron flux and perturbation of each polarization ellipse component.
\[ \Gamma_i = B_{i \text{observed}} \times \frac{\Gamma_{\text{modeled}}}{B_{i \text{modeled}}} \]

Where \( B_i \) refers to the disturbance to a given component of the polarization ellipse, and \( \Gamma_i \) is the inferred path-integrated precipitation based on parameter \( i \). Thus, for a single LEP observed on one transmitter-receiver path, we actually have three estimates of \( \Gamma \). And while each measurement may be noisy on its own, the average of all three is likely more robust to measurement error. It also provides an opportunity to check the self-consistency of our estimates, as we will do in the next section.

While this provides us with an estimate for the path-integrated precipitation metric for each event, an additional step is required to calculate the total electron precipitation over the region of disturbance, particularly for cases where the same LEP event was observed by multiple transmitter-receiver paths, as we discuss in the next subsection.

3.3.3 Total electron precipitation

The transmitter-receiver path for a given observed LEP event covers only a subset of the total region of disturbed ionosphere. The calculation of the total precipitation must be done uniquely for each event, as it is a function of the geography of the transmitter-receiver path along with the geographic distribution of precipitation, both unique for each lightning stroke.

The WIPP output provides the differential electron flux over a range of L shells, which can be mapped to a range of latitudes. Combined with the longitudinal scaling function, we map the precipitated electron flux over the full two-dimensional area of disturbance, as Figure 3.7 shows. We then integrate over this region to obtain a model of total number of precipitated electrons that would be observed by a particular transmitter-receiver path.

This number does not necessarily reflect the actual precipitation count. However, we do assume that the spatial distribution of precipitated electrons is roughly constant, there-
fore, the ratio of the path-integrated precipitation metric to the total number of precipitated
electrons is constant for a given geometry of transmitter-receiver path and lightning stroke.
We can therefore estimate the total number of precipitated electrons as follows:

\[ N_{e,i} = \Gamma_i \times \frac{N_{\text{modeled}}}{\Gamma_{\text{modeled}}} \]

To estimate the total precipitation for each event, we first separate each narrowband sig-
nal sample in which the event was detected - corresponding to a distinct transmitter-receiver
path. Each of these samples contains up to four estimates for the total precipitation, using
either the major axis, minor axis, tilt angle, or start phase components of the polarization
ellipse. We ignore the minor axis estimates, due to the unreliability of this measurement
shown in Figure 3.8. We define the precipitation estimate for a given transmitter-
receiver path observation as the median precipitation estimate of the major axis, tilt angle,
and start phase estimates. Having defined a precipitation estimate for each path observation
of a given event, we can then define the precipitation event for the event as a whole as the
median of the precipitation estimates for all the paths.
4.1 LEP distribution and behavior

Our collection process resulted in 157,780 candidate lightning strokes, visible over 755,524 narrowband VLF samples. We ran our classifier over each of these samples, and identified 25,961 where an event was visible. These 25,961 map to 18,109 unique events. This occurrence rate (11.5%) is similar to the 10.8% occurrence rate found by manually examining 1000 random samples.

The scale of our database allows for a uniquely comprehensive analysis of the conditions that produce LEP events, and the way the disturbed regions of the ionosphere affect narrowband VLF signals.

4.1.1 Distribution of Events

Our database allows us to look at the distribution of LEP events by parameters such as lightning current and geography in relation to the transmitter-receiver path.

Figure 4.1 shows the distribution of events by lightning current. Because we only examined candidate strokes with $|I| < 100kA$, there is a ”gap” in the center. The errorbars reflect the True Positive Rate (as an upper bound error) and the Positive Predictive Value (as a lower bound error) of the classifier. We see a positive correlation between peak current magnitude and the likelihood of an event to occur. There is an asymmetry in the event occurrence, with positive current strokes generating an event 16.2% of the time while negative current strokes generate an event 10.4% of the time. While this difference is larger than the error range, it is not as significant as the variance by stroke magnitude.
4.1.2 Amplitude and phase disturbance analysis

LEP events have long been reported using the amplitude disturbance. It is also known, however, that the phase may be disturbed, either in addition to, or instead of, the amplitude. As such, some analyses that consider only amplitude may be missing some LEP events. In our case, we are converting our amplitude and phase measurements into polarization ellipse, following the method described by Gross et al., 2018. With the polarization ellipse technique, the amplitude and phase on each of two horizontal magnetic field components are together converted into a single ellipse that has four components: A major axis, a minor axis, a tilt angle, and a start phase.

We examine the disturbance using the “perturbation detector” method described in Section 2.4, where we examine the average values for each of the four polarization variables in the 10 seconds preceding the stroke, and compare it to the post-stroke average 2-12 seconds after. For each variable this value is greater than the sum of the variances of the data before and after the stroke, we consider the disturbance to be nonzero. We note that while this method is resilient to noise, it is sensitive to changes in the event recovery time; if the event recovers within 12 seconds of the lightning stroke, the size of the scattered field can be underestimated.
Table 4.1 shows the distribution of events according to which components of the polarization ellipse is disturbed. Each entry in the table shows the fraction of events that have measurable scattered field components corresponding to the combination described in the row and column name. Each row and column represents a pair of scattered field components, with the rows referring to the amplitudes (major axis and minor axis), and the columns referring to the tilt angle and starting phase. The "Row total" and "Column total" sums up the rows, and columns, respectively.

The top-right entry in the table shows a count of 0, as all identified LEP events must be disturbed in at least one of the four components. On the other hand, 10.3% of events had measurable disturbances on all four ellipse components (‘both’ row and ‘both’ column). Looking at the distribution of amplitude disturbances (row total), we can see that both the major and minor axes are disturbed in roughly equal proportion: 16.5% and 22.7%, with 25.5% being disturbed on both.

However, 35.3% of events had no amplitude disturbance, meaning that many of these events would be missed if looking only at amplitude to identify LEP events. Of these, most had a disturbance exclusively in the start phase. In fact, this category (disturbance only in start phase) is the largest one, representing 30.1% of the LEP events in our database. For a single-mode TEM wave, the start phase corresponds to the phase velocity of the wave. To the extent that VLF waves may sometimes be dominated by a single mode, we may interpret this result, very roughly, to mean that some LEP events cause a change in group velocity of the propagating VLF waves, but do not change the attenuation rate by much. Gross et al. (2018) identified these types of LEP events, and suggested that these are due to the wide region of disturbance causing the phase velocities of the predominant signal propagation mode to be affected in a roughly equal way, without changing the signal strength of the combined modes.

More generally, it can be seen that 84.9% of events disturb the start phase (59.5% + 25.4%). The start phase can be thought of as a common-mode phase disturbance between
Table 4.1: Distribution of LEP events according to components of the polarization ellipse that are disturbed. There are 25961 events in total.

<table>
<thead>
<tr>
<th>ROW TOTAL</th>
<th>Angle/phase disturbance</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Start phase</td>
<td>Tilt angle</td>
<td>Both</td>
<td>Neither</td>
</tr>
<tr>
<td>Neither axis</td>
<td>35.3%</td>
<td>30.1%</td>
<td>0.96%</td>
<td>4.2%</td>
</tr>
<tr>
<td>Amplitude</td>
<td>Major axis only</td>
<td>16.5%</td>
<td>8.4%</td>
<td>1.6%</td>
</tr>
<tr>
<td>Disturbance</td>
<td>Minor axis only</td>
<td>22.7%</td>
<td>11.1%</td>
<td>3.1%</td>
</tr>
<tr>
<td>Both axes</td>
<td>25.5%</td>
<td>9.9%</td>
<td>3.2%</td>
<td>10.3%</td>
</tr>
<tr>
<td>COLUMN TOTAL</td>
<td>100%</td>
<td>59.5%</td>
<td>8.9%</td>
<td>25.4%</td>
</tr>
</tbody>
</table>

the two components of the polarization ellipse.

The tilt angle, however, was disturbed only 34.3% of the time. This may be a consequence of the fact that LEP ionospheric disturbances are relatively smooth spatially compared to wavelength. As such, the effect on VLF propagation is primarily a change in the attenuation rate and/or wavenumber of each mode, and less so on mode conversion on a sharp boundary as is often seen in the day/night terminator (Westerlund, 1968), or some Early/Fast events (Poulsen et al., 1993). In addition, we will show later that the tilt angle disturbance, when it does occur, does not appear to be related to the peak current of the lightning stroke triggering the LEP event.

Now that we have analyzed the occurrence of disturbances, we can examine their distribution. Figure 4.2 shows histograms of nonzero disturbances in all four polarization variables. Note that we present the major axis disturbances in log scale, while the minor axis is in absolute scale. This is because while the major axis amplitude is normalized in calculations to always have a positive value, the minor axis can take on negative or zero values, and as such is easier to examine in linear units.

The major axis is more likely to see negative perturbations (meaning a reduction) than positive perturbations (meaning an increase). This may be a consequence of the precipitation of electrons decreasing the ionosphere’s effective height, which when applied to a single-mode propagation theory results in an increased absorption of the signal and, on balance, a decreased amplitude at the receiver (Inan & Carpenter, 1987), since at lower
altitudes there are more collisions and thus higher absorption. However, in reality VLF propagation is complicated by multi-mode propagation so depending on the makeup of waveguide modes, may result in an increase in signal strength depending on how the constructive and destructive interference of the modes works. But this single-mode picture may explain why a reduction in amplitude is more common.

The starting phase more commonly decreases but sometimes increases. Using a similar single-mode simplified picture, the reduction in reflection height increases the incidence angle of propagating modes since the waveguide cutoff increases. This would cause an increase in the phase velocity along the waveguide, and thus a lower phase. But, as in the previous case, the single-mode picture may only help explain the imbalance, and the multi-mode propagation muddies this simplified picture.

The minor axis and tilt angle appear to have a more symmetric distribution between positive and negative changes.

Our database also allows us to better understand the role of peak current in determining the disturbances seen the VLF signals. Figure 4.3 shows our analysis of perturbations as a result of lightning strokes of different peak current ranges. Within each panel, each bar
Figure 4.3: Percentile magnitudes of field component perturbations for different peak current ranges.

represents a different percentile value of the respective (absolute) field component perturbations. The database from which this is drawn is event samples corresponding to lightning strokes with a peak current (magnitude) within the range indicated on the horizontal axis. There are correlations between the magnitude of perturbations within the major axis, minor axis, and starting phase, particularly for the 50th and 75th percentiles of the disturbances. However, we do not see a significant variation in tilt angle change with respect to peak current. This logically follows from the assumption that changes to the tilt angle, which represent the effects of field modes arriving at different scattering angles, may be largely determined by geographical factors rather than the intensity of the electron precipitation.

The role of peak current in determining the likelihood of individual field component perturbations appears to largely be confined to the major axis. Table 4.2 shows, for each field component, the likelihood of an event to have a perturbation in that polarization variable. We see a 19% increase in likelihood for the major axis to perturb between events formed by strokes with $|I| < 150kA$ and those from strokes with $250kA < |I| < 300kA$. The other three field components do not appear to have this relationship, however, even if the likelihood of an event to occur in the first place is more strongly determined by peak
Figure 4.4: A histogram of total electron precipitation of the events in our LEP database.

Table 4.2: Likelihood of individual field components to be perturbed in an event at different peak current ranges.

<table>
<thead>
<tr>
<th>Peak current magnitude</th>
<th>Likelihood of field component perturbation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Major Axis</td>
</tr>
<tr>
<td>100-150 kA</td>
<td>40.89%</td>
</tr>
<tr>
<td>150-200 kA</td>
<td>49.71%</td>
</tr>
<tr>
<td>200-250 kA</td>
<td>57.90%</td>
</tr>
<tr>
<td>250-300 kA</td>
<td>59.90%</td>
</tr>
</tbody>
</table>

4.2 Electron precipitation estimates

4.2.1 Distribution of Total Electron Precipitation

We apply our combined model to this database 18,109 unique LEP events, consisting of 25,961 narrowband VLF signatures. Each of these VLF signatures can be modeled using the approach described in Chapter 3.

We therefore have a precipitation estimate for each event in our database. Figure 4.4 shows the distribution of these estimates. Note that the LEP events can vary significantly,
Figure 4.5: A histogram of the standard deviation of precipitation measurements across different polarization ellipse components, scaled to the median measurement over several orders of magnitude, in estimated precipitated electrons. However, this range is consistent with the $10^{16}$ electron estimate from Peter (2007).

4.2.2 Self Consistency

We have two means of self-consistency checks to establish that our precipitation estimates are reasonable. For each LEP observation on a single transmitter-receiver path, we have three estimates of the total precipitation using each of the major axis, tilt angle, and start phase.

For each transmitter-receiver path for which we have an LEP event, we calculate the standard deviation between the three estimates, and then normalize that to the median estimate. Figure 4.5 shows the distribution of this quantity. There can be up to an order of magnitude variance among the three estimates. While significant, this is much less than $\sim 3$ orders of magnitude variance in the total precipitated flux as shown by Figure 4.4. As such, the individual estimates from the three channels appear to fairly well correlated compared to the overall dataset.

A second self consistency check comes from the fact that a given event may be ob-
Figure 4.6: A histogram of the standard deviation of precipitation measurements across different path estimates, scaled to the median measurement observed over multiple transmitter-receiver paths. These observations should be generally in agreement subject to measurement error. There are 4677 cases where the same LEP event is observed on multiple paths.

Figure 4.6 show the same calculation as in Figure 4.5, except applied to the multiple paths estimating the same LEP event. Once again, the $<1$ order of magnitude variance between multiple paths is less than the $\sim3$ orders of magnitude variance within all LEP events, indicating that the multiple path estimates were reasonably self-consistent.

However, since there is still some variance for the many estimates of total precipitation, we apply our earlier technique of taking the median among the three polarization parameter estimates, and if applicable, among the multiple path estimates, and use that result as our best guess of the total precipitation for a given LEP event. This approach likely minimizes the error due to measurement noise and model uncertainty.

4.2.3 Cumulative daily precipitation from LEP events

The Georgia Tech LF Receiver Network provides comprehensive coverage over large parts of the continental United States, particularly in the Southeast, but it does not provide com-
plete coverage (see Figure 1.7).

Therefore, when estimating the total number of LEP events occurring in a day, we must adjust for the fact that there are likely large numbers of LEP events not captured in our database. To extrapolate our results to the total precipitation induced by lightning in the continental United States, we make use of the National Lightning Detection Network (NLDN) (Cummins et al., 1998). The NLDN provides detailed information of all lightning strokes in the continental United States, including location, time, and peak current.

For each day, we note the total number of lightning strokes in the NLDN that occurred at nighttime with a peak current greater than 100 kA (or less than -100 kA) as $N_{\text{stroke total}}$, and compare this number to $N_{\text{stroke indb}}$, the number of strokes whose potential region of disturbance was measurable by the Georgia Tech LF Receiver network. This allows us to make two precipitation estimates for each day: the observed precipitation, defined as the sum of all precipitation estimates for events in our LEP database that occurred during that day; and the projected precipitation, defined by multiplying the observed precipitation by $\frac{N_{\text{stroke total}}}{N_{\text{stroke indb}}}$.

We show these estimates in Figure 4.7, which shows the observed precipitation and projected precipitation over the range of days in the LEP database. We ignore days for which $N_{\text{stroke indb}} = 0$. 60
Figure 4.7: Total electron precipitation for each day in the LEP database.
5.1 Satellite measurements

Although our main method of observing LEP events so far has been through the use of ground-based VLF remote sensing, we do have access to direct measurements of the radiation belts to help corroborate and contextualize our results.

We use data from the Van Allen Probes, formerly known as the Radiation Belt Storm Probes. These were a pair of satellites orbiting the Earth in operation from 2012 to 2019. Each probe was equipped with a number of instruments used to measure characteristics of the radiation belts, including the Magnetic Electron Ion Spectrometer (MagEIS). This instrument contained four separate units used to measure electrons with energies from 30 keV to 4 MeV (Boyd et al., 2019): LOW (30-200 keV), M75 and M35 (200 keV to 1 MeV), and HIGH (1-4 MeV). The M75 and M35 units differ in their orientation from the spacecraft’s spin axis; as the names suggest, the M35 unit points 35 degrees from the spin axis while the M75 unit points 75 degrees from the spin axis (the LOW and HIGH units also point 75 degrees from the spin axis). The spacecraft’s spin period (11s) allows the measurements to resolve pitch angles of measurements. Pitch angle measurements for electrons at the L3 data level are divided into 11 bins, centered at 8.19°, 24.55°, 40.91°, 57.27°, 73.64°, 90°, 106.36°, 122.73°, 139.09°, 155.45°, and 171.82°. In this data format, each measurement represents a 3 minute epoch, meaning there are 480 epochs over each 24 hours day. The Van Allen probes follow a highly eccentric equatorial orbit, taking them through L shells 6.75 to as low as 1.06, which means that the data collected broadly captures the inner radiation belt, the slot region, and much of the outer radiation belt.

Satellite data, including data from the Van Allen Probes, has been used in many previ-
ous attempts to detect and quantify lightning-induced precipitation. These approaches have limits, given that satellites can only observe a localized section of the radiation belt, both in time and space. However, they provide a method of directly observing the context of the radiation belt environment.

5.1.1 Past satellite studies

The first satellite observation of an LEP event was presented by Voss et al. (1984), which used the low Earth orbit satellite S81-1 to detect the precipitation of energetic electrons in bursts correlating one-to-one with intense VLF sferics. The study estimated that a single LEP event could empty up to 0.001% of the electrons in the radiation belt at L=2.3, based on the assumption that the electron flux in this L shell was $10^8 \text{ el cm}^{-2}\text{s}^{-1}$.

From 2004-2010, the DEMETER satellite orbited the Earth at a 710 km altitude circular polar orbit (Parrot et al., 2006). DEMETER was equipped with the Instrument for Detection of Particles (IDP) which allows it to measure electron flux from 72.9 keV to 2.35 MeV with 4 s time resolution and 17.8 keV energy resolution (Sauvaud et al., 2006). Analysis from the DEMETER satellite over 2006-2008 shows a correlation between changes in flux among 126 keV electrons and average nighttime lightning flash intensity over the year (Gemelos et al., 2009). This occurred in the range of $2 < L < 3$, consistent with the assumption that whistler waves from LEP events are a dominant mechanism in electron flux changes in that region of the radiation belts. However, this study was only able to correlate changes in electron flux with the occurrence of high intensity lightning, as sourced from the National Lightning Detection Network (NLDN) (Cummins et al., 1998), rather than with actual occurrence of LEP events. Recent work by Martinez-Calderon et al. (2020), using data from the Van Allen Probes in comparison with NLDN’s dataset, has supported this correlation between lightning occurrence and measured changes in electron flux.

Meredith et al. (2007), meanwhile, utilized data from the Combined Release and Radiation Effects Satellite (CRRES) to compare the impacts from the plasmaspheric hiss and
from terrestrial lightning-generated whistler waves on the radiation belt at $L=2.5$. The CR-RES orbited from a range of 305 km to 35,768 km and was equipped with the Plasma Wave Experiment, which allowed it to measure electric fields in the range of 5.6 Hz to 400 kHz (Anderson et al., 1992). Rather than directly calculating the changes in pitch angle using the equations from Bell (1984) as Bortnik et al. (2006) did, Meredith et al. (2007) used the PADIE (Pitch Angle and energy Diffusion of Ions and Electrons) code (Glauert & Horne, 2005) to compute the pitch angle diffusion rates for electrons as a result of different detected wave frequencies. The result found that lightning generated whistler waves were not a significant mechanism for electron losses in the radiation belt.

Green et al. (2020) estimated the occurrence of LEP events using electromagnetic field sensor data from the Van Allen Probes, defined in this case as the percentage of time during which whistler waves could be detected in the sensor data. This approach found a nighttime whistler presence 30% of the time. Building on these results, Claudepierre et al. (2022) used the wave spectrum of lightning-generated whistlers to estimate the impact on electron lifetimes, using the Full Diffusion Code (Ni et al., 2008).

5.1.2 Relationship between LEP occurrence and electron distribution

We examined the MagEIS data measured from January 1, 2018 to July 16, 2019, and from the longitude ranges from 60 degrees West to 130 degrees West, roughly covering the longitude range of the contiguous United States. For each of these 567 days, we used data from when the probes passed between the $L=1.5$ and $L=2.5$ shells.

To test the relationship between LEP events and the electron distribution in the inner radiation belt, we partitioned the set of days in the time scope into "High LEP" and "Low LEP" days, based on whether, in the database of LEP events described in Chapter 2, the percentage of candidate lightning strokes that result in detectable LEP events is greater or less than 10%. There are a total of 218 Low LEP days, and 349 High LEP days.

To define the statistical significance of any differences observed between the High/Low
LEP partition of days, we take a series of 100 random partitions of the days in this period. For each of these partitions, we can examine the differences in the average flux between the two subsets of the data. The define our error range as average of these difference.

There are 20 distinct energy bins that contain electron flux data, with centers ranging from 33 keV to 4092 keV. We observe, at various energy bands, a significant difference between the average flux on High LEP and Low LEP days.

We show three of these energy bins in Figure 5.1. Specifically we show the 33 keV bin (top graph) as an example of low energies. Here, High LEP days correlate to lower electron flux levels for 90 degree pitch angles. For higher energy electrons, the inverse trend appears to exist, where High LEP days have higher levels of electron flux for 90 degree pitch angle electrons, as can be seen in the distribution for 226 keV (middle graph). Beyond 1 MeV, such as for the 1575 keV band (bottom graph), the electron flux levels are so low that any clear trends are not visible beyond the ordinary variance between partitions.

These distributions make clear the differences between High LEP and Low LEP days at higher pitch angles, but the differences at low pitch angles may be difficult to observe due to the lower differential electron flux values.

Figure 5.2 shows the integrated electron flux on both High LEP and Low LEP days, averaged for both the global range of longitudes, and limited to the longitudes spanning North America. We see that within North America, there is not a significant difference between the electron flux measurements, while globally, High LEP days have significantly higher levels of electron flux.

This suggests that, within the North America region, the average flux of low pitch angle electrons is significantly reduced relative to the global average on High LEP days.

5.1.3 Comparison to model results

To compare the Van Allen Probe data to our modeled electron precipitation counts, we must first convert measurements of electron flux (cross sectional density of electrons within a
Figure 5.1: Pitch angle distribution for 33 keV, 226 keV, and 1575 keV electrons
duct of magnetic field lines) to a total count of electrons within the L shells affected. To do this, we integrate over the energy and geographic ranges, using a similar method as Thorne et al. (2005) equation 2.

Our integration of the electron flux must take into account both the bounce time of the electrons and the limited angular scope of the detectors. We carry out the integration using the following equation:

\[ N_e = \int_{1.5}^{2.5} \int_{\phi_{\min}}^{\phi_{\max}} \int_{33\text{keV}}^{1064\text{keV}} R_e^2 2\pi \sin\alpha A \tau_b \ d\alpha \ dE \ d\phi \ dL \]

Where \( A \) is the Van Allen probe measurements, in units of \( \text{el cm}^{-2} \text{ster}^{-1} \text{s}^{-1} \), \( \alpha \) is the pitch angle, \( R_e \) is the Earth’s radius, \( \phi \) is the longitude in radians (integrated over North America). \( \tau_b \) is the bounce time, a function of magnetic latitude, pitch angle, and energy. The \( 2\pi \sin\alpha \) term accounts for the circumference of the circular opening of a cone.

We calculate the bounce time \( \tau_b \) using the following formula from Walt, 2005 (eqn 4.28):

\[ \tau_b = 0.117 \times \frac{L}{\beta} (1 - 0.4635 \sin^{3/4}\alpha) \]

Where \( \beta = \frac{v}{c} \), the velocity of the electron divided by the speed of light.
Carrying out this integral, we can estimate the number of electrons in the radiation belts over the range of longitudes spanning North America. For any given L shell and longitude combination, the Van Allen probe takes many measurements. Therefore, we have access to a distribution of total electron count measurements.

Figure 5.3 shows this distribution as a Cumulative Distribution Function (CDF). The CDF indicates that the electron count varies roughly uniformly in the range of $1 \times 10^{20}$ to $1 \times 10^{21}$ electrons.

Given the precipitation estimates in Figure 4.7, the total daily precipitated electrons typically represent 0.01% to 1% of the total electrons in the radiation belts, with some days reaching as high as 10% of electrons. This percentage does not appear to fully account for the gaps between “High LEP” and “Low LEP” days visible in the Van Allen data, which, under some pitch angles and energies, is up to 20% of the electron flux. However, this percentage is larger than that estimated by Burgess and Inan (1993), which suggests that lightning-induced whistlers surpass the plasmaspheric hiss as the dominant loss mechanism in the 1.5 L – 2.5 range.
CHAPTER 6
CONCLUSION

6.1 Summary

Our use of a machine learning classifier for automatic detection of LEP events has allowed us to assemble large database events across North America. This database provides insights into the occurrence and behavior of events. We observe a quantified trend of amplitude and phase components having a negative perturbation. We also observe that a large share, over 30%, of LEP events only see a perturbation in the start phase.

Using computer simulations to model the full process of lightning-induced precipitation, we are able to show a linear relationship between VLF signal perturbation and the total electrons depleted during the corresponding events. We apply this model towards our database to interpolate a distribution of precipitated electron count per event, and an estimate of the total electron precipitation per day.

Finally, our database can be compared with data from the Van Allen probes, giving us an ability to situate our results in the context of direct radiation belt measurements. We see a general correlation between changes in the electron distribution and the occurrence rate of LEP events. We also find that our estimates for total daily electron precipitation range from 0.01% to 1% of the total electrons in the radiation belts, suggesting that LEP events are a predominant mechanism for electron losses in the lower belt.

6.2 Further work

Our research has touched upon many facets of the total body of LEP investigation, ranging from the detection and analysis of perturbed VLF signals, to the use of computer simulations to model LEP processes, to analyzing direct satellite measurements in the magneto-
sphere. As a result, there are many avenues in which this work can be expanded on for future research.

6.2.1 LEP detection

While our use of machine learning methods, specifically Artificial Neural Networks (ANNs) and Convolutional Neural Networks (CNNs), allowed us to detect LEP events much faster and at a larger scale than previous work, our classifier did not have perfect accuracy. A larger training set allows for the network weights and biases to converge at a higher accuracy (Foody et al., 1995). With a sufficiently large amount of training data, it may be possible to combine the multiple narrow band signal samples corresponding to a given lightning stroke to expand the number of features considered. Gross and Cohen (2018) demonstrated that inputs from multiple narrowband signal samples from the same time period can be used as inputs to an ANN classifier to interpolate information about the overlaying ionosphere.

In addition, having a wider range of LF radio receivers would allow the set of transmitter-receiver paths to cover a larger range of geographic conditions. With sufficient coverage in the LF receiver network, in theory it may be possible to observe all of the LEP events occurring within a given day over North America. This would remove the need to interpolate the total amount of electron precipitation using the fraction of lightning visible to the current network of receivers.

6.2.2 LEP modeling

While the relationship between total electron precipitation and VLF signal perturbation, using the SAM and LWPC simulations, was approximately linear, we obtained this relationship by assuming a "two-parameter" model as the initial ionosphere in the D region. However, work by McCormick and Cohen (2021) suggests that the ionosphere may be better modeled using additional parameters. Furthermore, there are significant seasonal variations in the ionosphere’s electron density (Richardson & Cohen, 2021). Future work
should investigate the use of these ionospheric models with the SAM simulations in order to better interpolate the electron precipitation quantities associated with VLF signal perturbations.
REFERENCES


75


Parrot, M., Berthelier, J., Lebreton, J., Sauvaud, J., Santolik, O., & Blecki, J. (2006). Examples of unusual ionospheric observations made by the DEMETER satellite over


