

STATISTICAL ANALYSIS AND MODELING OF LIGHTNING  
USING RADIO REMOTE SENSING

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

Fadi G. Zoghzoghy

June 2015

© Copyright by Fadi G. Zoghzy 2015  
All Rights Reserved



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

---

(Umran Inan) Principal Adviser

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

---

(Morris Cohen)

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

---

(John Gill)

Approved for the Stanford University Committee on Graduate Studies

# Abstract

Lightning is one of the most fascinating, naturally abundant, and powerful electrical processes on Earth. On average,  $\sim 50$  lightning flashes occur around the globe every second, each releasing billions of Joules of energy. Lightning leads annually to billions of dollars in damages and to thousands of deaths, injuries, and fires, and is of great practical interest to those concerned with the safety of aircraft, spacecraft, ground-based electronic systems and urban environments.

Interestingly, to date, the physical nature of lightning is not fully understood, in part due the difficulty of making direct ground-truth measurements inside thunderclouds or inside the lightning channel, and in part due to the wide range of timescales (from nanoseconds to hours) that are involved and the seemingly random nature of these processes. However, using radio emissions from individual discharges and leveraging the abundance of naturally occurring lightning enables a wide range of experimental techniques. Recent studies also suggest that lightning flashes have larger peak currents over the ocean than over land, which has important implications on the safety and design of aviation and navigation systems. These observations remain poorly understood.

In this dissertation, we use remote sensing data in tandem with statistical techniques and electromagnetic modeling to investigate the lightning discharge process, quantify the lightning occurrence probability and recovery time as a function of distance and time, and introduce a proxy method for monitoring the charging processes in the cloud. To better understand oceanic lightning, we conduct observations with a sensitive Very Low Frequency/Low Frequency (1–450 kHz) radio receiver system deployed aboard the NOAA Ronald W. Brown research vessel to collect magnetic

field radio emissions from deep-ocean lightning. We construct electromagnetic models to simulate lightning waveforms and compare them with the data acquired to infer properties of causative lightning currents. We analyze thousands of land and oceanic waveforms and compare lightning properties of land and oceanic thunderstorms.

# Acknowledgements

The work presented in this dissertation is the result of a challenging yet amazing journey. Many helped make this experience possible, interesting, and fun.

First I would like to thank my dissertation advisor Professor Umran Inan who provided me with the incredible opportunity to be part of the Stanford Very Low Frequency (VLF) group and who believed in my capabilities from day one. I am extremely grateful for his constant support, help, and for our numerous discussions that shaped my work.

I also owe a great debt of gratitude to Professor Morris Cohen for being an exemplary academic mentor. Had it not been for his generous guidance and scientific insights, I would not have achieved the personal and professional maturity to complete this dissertation. I am also deeply thankful to Dr. Ryan Said and Dr. Robert Marshall who have been tremendously supportive.

I thank Professors John Gill, Joseph Kahn, and Tony Fraser-Smith, who provided profound feedback as members of my oral examination and reading committees.

My time at Stanford would not have been as much fun and productive without my colleagues in the VLF group, especially Patrick Blaes, Rasoul Kabirzadeh, Justin Li, Chris Young, Dr. Vijay Harid, Dr. Can Liang, Dr. Kevin Graf, and Dr. George Jin. A big thank you also goes to Shaolan Min and Helen Niu for taking care of the administrative work and to Jeff Chang for his efforts developing the Low Frequency receiver system. I would also like to thank Dr. Steve Goodman, of NOAA, for arranging access to the Ronald Brown research vessel.

My five years at Stanford have been blessed with good friends who tirelessly encouraged me. Thank you Karen Khalaf, Nicolas Kseib, Fadi Wakim, Nader Haddad,

Sarah Fakherddine, Tania Bakhos, Nathan Volman, Charles Feng, Anwar Ghauche, and Jad Nehme.

My childhood friends from Lebanon helped me become the ambitious person I am today. For that I am thankful to Karam Barnaba, Roudy Farraj, Rodrigue Akl, Charbel Moubarak, Joe Kerbage, and Ziad Chebib.

To my parents, George and May, and to my siblings, Joe and Lamia: my eternal gratitude will always go to you. Despite the geographical distance and spread of our family, being constantly blessed with your love and generosity pushed me to pursue achievements that are beyond my limits. Your emotional support and sacrifice made obstacles appear seamless. Thank you.

Fadi Zoghzoghy

Palo Alto, CA

June 2, 2015

*This work has been supported by the Defense Advanced Research Project Agency under the ‘NIMUBS’ program, via grant HR0011-10-1-0058-P00001 to Stanford University with subcontract to the Georgia Institute of Technology. NLDN and GLD360 data are provided by Vaisala, Inc. under a cooperative agreement. NALMA data are provided by the University of Alabama Huntsville and New Mexico Tech.*

# Contents

<b>Abstract</b>	<b>iv</b>
<b>Acknowledgements</b>	<b>vi</b>
<b>1 Introduction to Lightning</b>	<b>1</b>
1.1 Lightning . . . . .	1
1.2 Satellite Observations . . . . .	6
1.3 Ground-based Remote Sensing . . . . .	8
1.3.1 North Alabama Lightning Mapping Array (NALMA) . . . . .	9
1.3.2 National Lightning Detection Network (NLDN) . . . . .	10
1.3.3 Global Lightning Detection Network (GLD360) . . . . .	11
1.4 Oceanic lightning . . . . .	13
1.5 Contributions of this Work . . . . .	15
<b>2 Lightning Detection Instrumentation</b>	<b>17</b>
2.1 Electromagnetic Remote Sensing . . . . .	17
2.1.1 Earth and the ionosphere . . . . .	18
2.1.2 Ideal Parallel Plate Waveguide . . . . .	19
2.1.3 Earth-ionosphere Waveguide . . . . .	22
2.1.4 Ray Hop Components . . . . .	22
2.2 Data Acquisition . . . . .	26
2.2.1 Low Frequency Receiver . . . . .	26
2.2.2 Instrumentation Errors and Correction . . . . .	29
2.3 Oceanic LF Observations . . . . .	30

<b>3</b>	<b>Ground Wave Modeling and Analysis</b>	<b>34</b>
3.1	Return Stroke Modeling . . . . .	34
3.2	Analytical Model . . . . .	37
3.3	Ground Wave Full-Wave Method (gwFWM) . . . . .	40
3.3.1	Model Description . . . . .	40
3.3.2	Validation and Comparison . . . . .	44
3.3.3	Earth Curvature . . . . .	45
3.4	Sensitivity Analyses . . . . .	47
3.4.1	Current Rise-Time and Ground Conductivity . . . . .	47
3.4.2	Current Fall-Time . . . . .	51
3.4.3	Return Stroke Speed . . . . .	52
<b>4</b>	<b>Waveform to Current Parameters Modeling</b>	<b>56</b>
4.1	Introduction to Inverse Modeling . . . . .	56
4.2	Artificial Neural Network . . . . .	58
4.2.1	Model Structure . . . . .	59
4.2.2	Training Process . . . . .	62
4.3	gwFWM Inverse Modeling . . . . .	65
4.4	Application to gwFWM Simulation Waveforms . . . . .	72
4.4.1	Training: Waveform Features to Lightning Parameters . . . . .	72
4.4.2	Inverse Model Accuracy Testing . . . . .	76
4.5	Application to Experimental Waveforms . . . . .	79
4.5.1	Current Rise-Time . . . . .	80
4.5.2	Return Stroke Speed . . . . .	81
4.5.3	Peak Current . . . . .	83
<b>5</b>	<b>Statistical Patterns in Lightning Location</b>	<b>85</b>
5.1	Remote Sensing of Electric Field Recovery . . . . .	86
5.1.1	Statistical Approach . . . . .	87
5.1.2	Application to Negative CG Stroke Data . . . . .	88
5.2	Application to Lightning Flash Data . . . . .	92
5.2.1	Negative CG Flash Data . . . . .	92

5.2.2	Negative CG Multiplicity . . . . .	93
5.2.3	Negative CG Peak Current . . . . .	95
5.2.4	Ground and Cloud Flash Type . . . . .	97
5.2.5	Land and Ocean Lightning . . . . .	100
5.3	Patterns between Negative CG Strokes . . . . .	101
5.3.1	Previous LMA Studies . . . . .	102
5.3.2	Application to NLDN Negative CG Stroke Data . . . . .	103
5.3.3	Proposed Mechanism . . . . .	106
5.3.4	LMA Case Studies . . . . .	108
5.3.5	LMA Data around NLDN Negative CG Data . . . . .	112
<b>6</b>	<b>Analysis of Oceanic Peak Currents</b>	<b>116</b>
6.1	Large-scale Peak Current Statistics . . . . .	116
6.1.1	Application to NLDN CG Stroke Data . . . . .	118
6.1.2	Application to NLDN IC Stroke Data . . . . .	124
6.2	Systematic Network Errors . . . . .	126
6.2.1	Range Filtering and Correction . . . . .	127
6.2.2	Peak Current Estimation . . . . .	131
6.3	Discussion of Ocean Lightning Findings . . . . .	133
<b>7</b>	<b>Summary and Suggestions for Future Work</b>	<b>136</b>
7.1	Future Work . . . . .	138
7.1.1	Coastal VHF Array . . . . .	138
7.1.2	Inverse Modeling of LF and VHF Data . . . . .	139
7.1.3	Coastal Thunderstorms . . . . .	139



# List of Tables

3.1	Attenuation function $A(z)$ with altitude for MTLL and MTLE. . . .	36
3.2	Comparison of the gwFWM, FDTD EMP, and the analytical model. .	44
6.1	Summary of geometric mean statistics from Figure 6.1–6.4. The last column shows the percentage increase between the 0–15 km oceanic region and the 0–15 km land region. . . . .	126
6.2	Original and post-correction increase in geometric mean in GLD360 peak currents in the 500–515 km oceanic region, compared to land. .	133

# List of Figures

1.1	Cloud-to-ground and intra-cloud lightning in a lightning-producing thundercloud. . . . .	2
1.2	Cartoon of the different process in a typical cloud-to-ground negative discharge. The downward negative leaders are shown in red, the upward positive leaders in blue, and the return strokes in green. . . . .	4
1.3	Thematic map of lightning occurrence computed using OTD optical data. Adapted from <i>Christian et al.</i> [2003]. . . . .	6
1.4	Radiation spectrum corresponding to initial negative cloud-to-ground flashes. Adapted from <i>Uman</i> [1987, p.118]. . . . .	8
1.5	Thematic map of lightning occurrence computed using one year of GLD360 data. Adapted from <i>Said et al.</i> [2013]. . . . .	12
1.6	Thematic map of the geometric mean of peak current estimates computed using one year of GLD360 data. Adapted from [ <i>Said et al.</i> , 2013].	12
2.1	Setup of the ideal parallel plate waveguide. . . . .	20
2.2	Cartoon of the ground wave (in red), 1 <sup>st</sup> sky wave (in green), and 2 <sup>nd</sup> sky wave propagation to the receiver. . . . .	23
2.3	Sample experimental LF lightning waveform. . . . .	24
2.4	Block diagram of the LF receiver system. . . . .	26
2.5	Cartoon of the spheric rotation process. (Left panel) The incoming wave $B(t)$ is captured by the two orthogonal antennas. (Right panel) The axes are digitally rotated to find the azimuthal magnetic field. . . . .	28

2.6	(Left) Picture of the Ronald W. Brown research vessel and the LF receiver magnetic loop antenna. (Right) Map of the voyage starting from San Juan, Puerto Rico in February 2013 and ending in San Francisco, California, U.S. in May 2014. . . . .	30
2.7	Pictures of the outdoor (left) and indoor (right) receiver system aboard the Ronald W. Brown research vessel. . . . .	31
2.8	Feature extraction tool applied to a waveform corresponding to a GLD360-reported stroke that occurred on 17-Jul-2013 08:15:05 UT at (lat = 33.49, lon = -76.99), with -19 kA reported peak current, 285 km from the location of the ship. . . . .	32
3.1	Illustration of return stroke modeling using engineering models. (left) The channel-base current $i(z = 0, t)$ is selected following a linear rise to the peak and an exponential fall. (right) The current pulse decays in magnitude as it travels upward at the return stroke speed. . . . .	36
3.2	Illustration of the geometry of the planar configuration. . . . .	37
3.3	gwFWM model setup. The lightning sources are modeled at the boundaries of the stratified region, the ionosphere is replaced with radiation boundary conditions, and each medium can have an arbitrary conductivity and permittivity. . . . .	41
3.4	gwFWM modeling flow. (left) The channel is divided into segments. (middle) The resulting current moments are convolved with the impulse responses. (right) The resulting ground wave is found. . . . .	42
3.5	Simulated ground waves using the gwFWM (in solid red), the analytical approximations (in solid black), and the FDTD EMP (in dashed blue) for the same input parameters: MTLL model, 2 $\mu$ s rise-time, 30 $\mu$ s fall-time, 100 kA peak current, 3 km source height, $2c/3$ return stroke speed, flat 2 mS/m Earth, and frequency content up to 500 kHz. . . . .	43

3.6	Attenuation due to Earth curvature (shown on a dB scale) as a function of propagation distance (in kilometers) for different frequency components: 20 kHz (in blue), 50 kHz (in red), 100 kHz (in green), 200 kHz (in black), and 400 kHz (in cyan). . . . .	46
3.7	Ground wave for different values of channel-base current rise-time: 1 $\mu$ s (in black), 2 $\mu$ s (in blue), 3 $\mu$ s (in red), and 4 $\mu$ s (in green). . . . .	48
3.8	Measured 20%-90% waveform rise-time, with the corresponding 25 <sup>th</sup> and 75 <sup>th</sup> percentile error bars, as a function of propagation distance to the receiver for three categories: (in blue) 7,027 oceanic discharges with propagation path over the ocean, (in green) 17,945 land discharges with propagation path over land, and (in cyan) 683 land discharges with mixed propagation path. . . . .	49
3.9	Ground wave for different values of channel-base current fall-time: 20 $\mu$ s (in black), 27 $\mu$ s (in blue), 34 $\mu$ s (in red), and 40 $\mu$ s (in green). . . . .	51
3.10	Ground wave for different return stroke speeds: $c/5$ (in black), $c/2$ (in blue), $3c/4$ (in red), and $c$ (in green) . . . . .	52
3.11	(Top panel) Average peak-normalized waveform for four lightning groups between 480 km and 520 km distance to the ship: (in blue) 1,162 land -CGs with GLD360 peak currents between 15 and 50 kA, (in red) 23 land -CGs >100 kA, (in green) 336 oceanic -CGs 15-50 kA, and (in black) 61 ocean -CGs >100 kA. (Bottom panels) Probability mass function of the waveform 20%-90% rise-times (left) and 90%-20% fall-times (right) of the four groups, with vertical lines corresponding to the medians. . . . .	53
4.1	Forward (left to right) and inverse (right to left) modeling flows. . . . .	57
4.2	Structure of the single hidden layer artificial neural network. . . . .	59
4.3	Expressions and shapes of two popular activation functions: the sigmoid (left) and the hyperbolic tangent sigmoid (right). . . . .	61

4.4	Estimated ground conductivity in the United States. Adapted from <i>Fine</i> [1953]. Lightning strikes over land occurred in the region highlighted by the red hexagon, propagating to the ship which was docked in Charleston, South Carolina. . . . .	66
4.5	Diagram of the flow of the statistical inverse model. . . . .	69
4.6	Illustration of the inverse modeling of simulation waveforms (top) and experimental waveforms (bottom). . . . .	71
4.7	Scatter matrix that captures correlations between three current parameters (top-down: current rise-time, return stroke speed, and peak current) and four waveform features (left-right: 20%–90% rise-time, ground waveform peak, 90%–20% fall-time, and 20% to 20% integral). . . . .	74
4.8	Median testing error with 25 <sup>th</sup> and 75 <sup>th</sup> percentile error bars for predictions of current rise-time (top panel), peak current (middle panel), and return stroke speed (bottom panel). . . . .	78
4.9	Estimated distribution of current rise-time for 1,098 oceanic (in blue) and 1,649 land (in red) discharges. . . . .	80
4.10	Estimated distribution of return stroke speed for 1,098 oceanic (in blue) and 1,649 land (in red) discharges. . . . .	81
4.11	Estimated distribution of peak current for 1,098 oceanic (in blue) and 1,649 land (in red) discharges. . . . .	83
5.1	The number of strokes as a function of distance from the source stroke, plotted separately for the 1 <sup>st</sup> second (blue), second 2 <sup>nd</sup> (green), 3 <sup>rd</sup> second (red), and 15 <sup>th</sup> second (black). NLDN Data from the largest lightning cluster of 23–August–2007 are used. . . . .	88
5.2	The flash occurrence density as a function of time, separately parameterized by concentric rings from the source. The recoveries in the 0–2 km region (blue), 2–4 km region (red), 4–6 km region (green), and 10–12 km region (black) are shown. All –CG flash data from August 23, 2007 are used. . . . .	91

5.3	(Top) The flash occurrence density in the 0–5 km region around the source discharge as a function of time, parameterized by flash multiplicity. The flash occurrence densities are plotted separately for flashes with multiplicity 1 (blue), multiplicity 2 (red), and higher multiplicities (green). (Bottom) Similar analysis but for the 5–10 km region. –CG flash data from August 22–24, 2007 are used. . . . .	94
5.4	The flash suppression effect as a function of time and parameterized by distance, separately for –CG single-stroke flashes with peak currents in the 5–20 kA range (blue) and for –CG single-stroke flashes with peak currents greater than 30 kA. Single-stroke –CG flash data from August 22–24, 2007 are used. . . . .	96
5.5	The flash suppression effect as a function of time parameterized by flash type and distance from the discharge: (top–left) recovery of CG lightning occurrence following a CG discharge, (top–right) recovery of IC lightning occurrence following an IC discharge, (bottom–left) recovery of IC lightning occurrence following a CG discharge, (bottom–right) recovery of CG lightning occurrence following an IC discharge. The recoveries in the 0–2 km region (blue), 2–4 km region (red), 6–8 km region (green), and 10–12 km region (black) are shown. NLDN CG and IC flash data from July 02–25, 2011 are used. . . . .	98
5.6	The flash suppression effect as a function of time for land and for oceanic lightning parameterized by distance. The solid lines correspond to the recovery following land lightning and the dashes lines for the oceanic lightning recovery. All GLD360 flashes from July 02–25, 2011 are used. . . . .	100

5.7	Spatial and temporal distribution of subsequent strokes around initial strokes, using 2,033,935 NLDN –CG strokes from July 2010. The histogram is scaled by the differential area of each radius bin, the summation over the histogram entries is normalized to unity, and the resulting values are displayed (color coded) on a logarithmic scale. The white dashed line corresponds to a radial distance of 750 meters, marking an approximate separation between the NLDN-reported location of subsequent strokes that recur along the same channel and the reported location of strokes that form a new one [Stall <i>et al.</i> , 2009]. The solid black line corresponds to the least squares linear fit through the peaks of lightning occurrence as a function of radial distance. . . . .	104
5.8	Cartoon of the suggested physical process that governs the formation of distant subsequent channels in –CG flashes. (Panel I) A stepped leader propagates downward at a speed of 100–300 km/s to form an initial channel. (Panel II) The return stroke propagates upward along the ionized channel at $\sim c/3$ . (Panel III) A leader follows the return stroke propagating outward in the cloud. (Panel IV) The cloud leader turns into a downward stepped leader and forms a new lightning channel 3–7 kilometers from the location of the initial one. . . . .	107
5.9	Case studies of the occurrence of North Alabama LMA sources coincident with NLDN –CG return strokes. The panels correspond to three –CG flashes, showing the altitude of LMA sources as a function of time relative to the initial NLDN stroke. The times of the initial and subsequent strokes are displayed in blue and red, respectively. . . . .	110
5.10	(Top) Top-down view of an NLDN –CG flash with overlaid LMA sources, color coded in time relative to the initial stroke (blue). (Bottom) Side-view of the same –CG flash, showing the altitude of sources as a function of time color-coded in time consistent with the top panel. The event occurred on July 12, 2011 at 19:44 UT. . . . .	111

5.11	Temporal distribution of LMA sources around NLDN strokes, parameterized by distance, using NLDN and LMA data from 2004 to 2010. The curves are normalized to have unit area and are displayed on a logarithmic time scale. . . . .	113
6.1	Distribution of NLDN peak currents of initial $-$ CGs from July 2013 in the following regions: 30–45 km ocean (in blue), 15–30 km ocean (in red), 0–15 km ocean (in green), 0–15 km land (in cyan), 15–30 km land (in magenta), and 30–45 km land (in black) . . . . .	118
6.2	Same as Figure 6.1 but for same-channel subsequent $-$ CGs. . . . .	120
6.3	Same as Figure 6.1 but for initial $+$ CGs. . . . .	122
6.4	Same as Figure 6.1 but for $-$ ICs (top) and $+$ ICs (bottom). . . . .	125
6.5	Occurrence and average peak current of GLD360 initial $-$ CGs in oceanic grid cells (2013). . . . .	127
6.6	Original (blue) and corrected (dashed green, dashed red) peak current distributions of initial $-$ CGs reported by GLD360 in the 500–515 km oceanic region (2013). . . . .	130
6.7	Distribution of the ratio of GLD360 to NLDN peak currents over ocean (in blue) and over land (in red). . . . .	132



# Chapter 1

## Introduction to Lightning

### 1.1 Lightning

Lightning is formed inside cumulonimbus clouds, in particular those that feature strong convective updrafts that cause frozen water particles of different sizes and temperatures to rub and exchange charge as they updraft, separating electrical charge vertically. The charge separation is believed to result from the graupel-ice mechanism, in which light ice particles, carried by the updraft, pick up a net positive charge while heavier graupel particles pick up a net negative charge [[Rakov and Uman, 2007](#), p.86]. Figure 1.1 shows the classic picture of the charge structure inside a thundercloud which involves three layers of charge, a positive layer at the top, a negative layer in the middle, and a (sometimes ignored) smaller positive layer at the bottom [[MacGorman and Rust, 1998](#), ch.3]. Lightning flashes neutralize the charge separation and are typically classified as either intra-cloud (IC) or cloud-to-ground (CG) discharges. Other types of lightning, such as cloud-to-cloud and cloud-to-air discharges, exist but are rare and often overlooked.

Roughly 75% of lightning activity corresponds to IC discharges [[Rakov and Uman, 2007](#), p.4], which occur within the cloud and involve a breakdown connecting a positive and a negative layer. ICs tend to dominate the early stages of thunderstorm development [[Rakov and Uman, 2007](#), p.49] but are difficult to study due to the inability to capture high resolution optical data and to measure currents and charge transfers

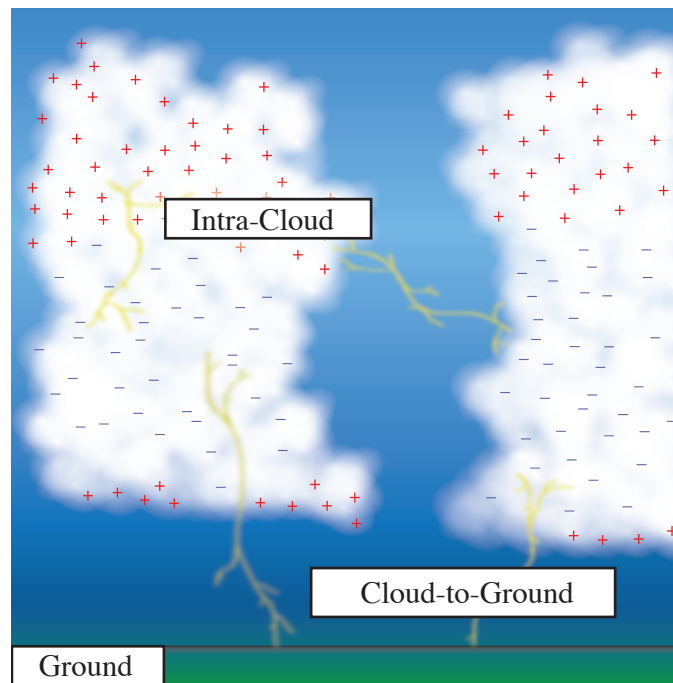


Figure 1.1: Cloud-to-ground and intra-cloud lightning in a lightning-producing thundercloud.

inside the cloud. Thus IC studies are mostly confined to ground-based remote sensing techniques, using electric field, magnetic field, or acoustic systems [[Rakov and Uman, 2007](#), p.321].

CG flashes account for the remaining  $\sim 25\%$  of lightning activity [[Prentice and Mackerras, 1977](#)] and involve a breakdown that connects the cloud to ground, with currents flowing along the channel that are typically stronger than their IC counterparts. The majority of lightning research has focused on studying CGs, in part due to their direct impact on our environment, and in part due to the ability to obtain high speed camera recordings of the CG channel and to directly measure currents at the bottom of the channel in rocket-triggered and tower lightning. CG flashes could either initiate from the ground (upward propagation) or from the cloud (downward propagation) and could either have a negative or a positive polarity. Natural CGs are typically downward (initiate in the cloud), upward flashes are extremely rare ( $< 1\%$ ) and usually follow a breakdown that initiates from tall ground objects. Approximately 90% of downward CGs have a negative polarity ( $-CG$ ) and the remaining  $\sim 10\%$  are positive ( $+CG$ ).

Figure 1.2 depicts the various stages of a typical downward  $-CG$  discharge, initiating in the middle negative charge layer inside the thundercloud [[Jacobson and Krider, 1976](#)]. Due to updrafts, charge separation builds up, leading to preliminary breakdown. The background conditions that precede the breakdown remain poorly understood but it is generally agreed that an intensification in the local electric field causes the dielectric breakdown of air providing conditions for the stepped leader.

The stepped leader consists of a negatively charged plasma that travels to ground with an average speed of 200 km/s, forming a conductive path in virgin air [[Rakov and Uman, 1990a](#)]. The stepped leader propagates in intermittent steps, with an average step length of 50 m and an overall duration of tens of ms [[Rakov and Uman, 1990a](#)], leading to a tree-like structure. As the leader approaches ground, positive leaders are launched from ground upward, due to the intensification of the electric potential between the tip of the downward leader and ground, at least one of which connects to the negative stepped leader tens of meters above ground. This attachment process marks the beginning of the return stroke.

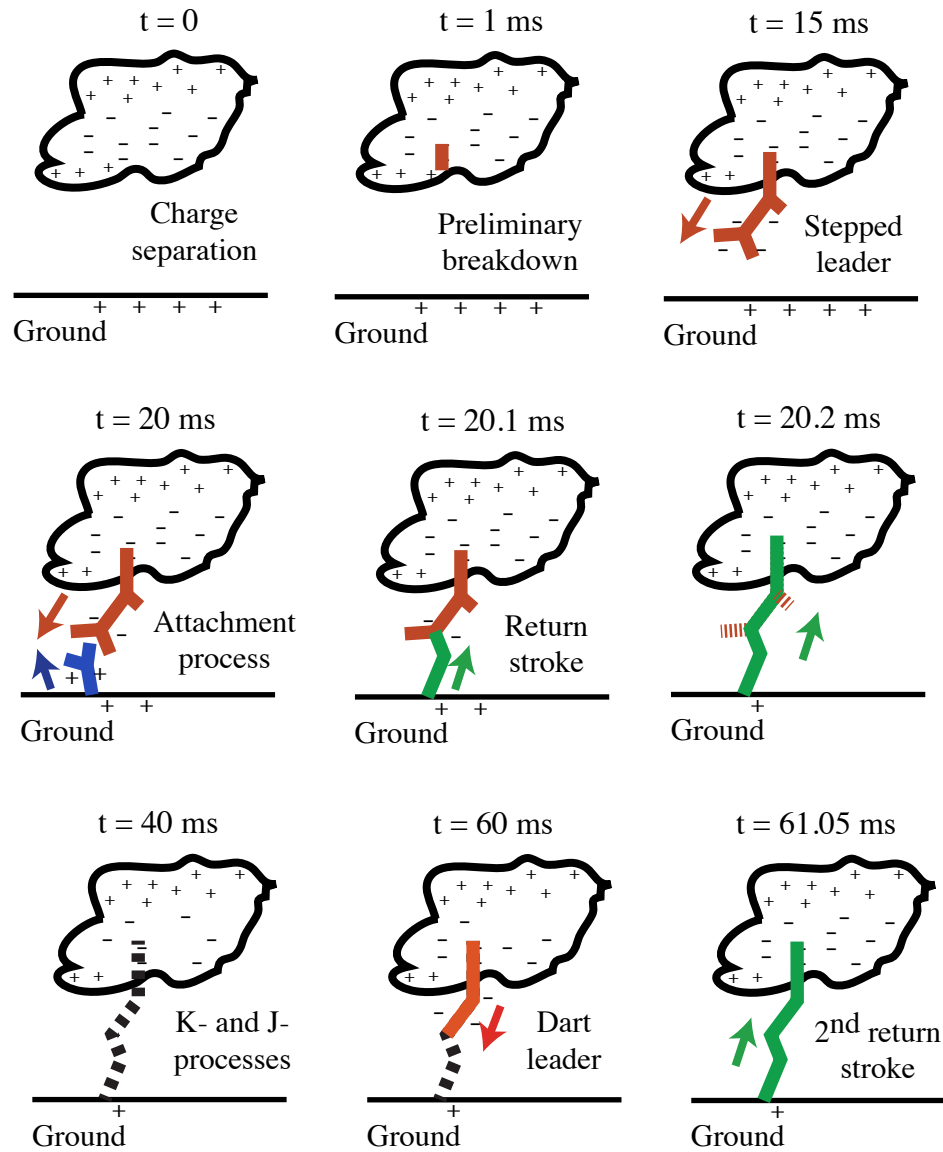


Figure 1.2: Cartoon of the different process in a typical cloud-to-ground negative discharge. The downward negative leaders are shown in red, the upward positive leaders in blue, and the return strokes in green.

The return stroke illuminates one of the drawn paths and travels upward at speeds which have been optically observed to vary between one-third to one-half the speed of light, with typical peak currents ranging from a few to many hundreds of kA. Peak currents of the return stroke (which only lasts several microseconds) do not correlate well with the total charge transferred in a CG flash [Cummer *et al.*, 2013]. The bulk of the charge is transferred to ground by long-lasting (and weaker) continuing currents, which last hundreds of milliseconds [Rakov and Uman, 2007, p.222]. Nevertheless peak current measurements quantify the intensity of the impulsive phase of the return stroke and are of great practical interest.

–CG flashes last for  $\sim 1$  second and may consist of multiple (3–4 on average) return strokes separated by tens of milliseconds [Berger *et al.*, 1975; Rakov and Uman, 2007, p.4]. The first return stroke occurs in the ionized channel that was initially formed by the stepped leader. The following subsequent return strokes either recur along the same existing channel (which has an elevated temperature and higher conductivity) via a dart leader, or in a newly-formed channel up to several kilometers away from the first ground contact point with another stepped leader or a dart-stepped leader [Rakov and Uman, 2007, p.164-165]. Unlike the stepped leader, the dart leader travels in a more continuous fashion and roughly two orders of magnitude faster (total duration of 1–2 ms) due to the higher conductivity of the lingering channel.

J- (for ‘Junction’) and K- processes occur between successive return strokes (and can recur after the final return stroke) and transport charge from other regions of the cloud to the top of the ionized channel. The J-process lasts tens of milliseconds and carries charge horizontally toward the previous channel, slowly re-building the local electric field. While the charge movement is toward the existing channel, subsequent strokes can still occur in other channels, redistributing charge in other parts of the storm. The K-process has much shorter timescales and is associated with abrupt jumps in the electric field with a risetime (10–90%) shorter than 3 ms [Thottappillil *et al.*, 1990; Rakov *et al.*, 1992]. K-processes are sometimes interpreted as “attempted” leaders that propagate down the existing (but decaying) ionized path, which has an elevated temperature, but fail to reach ground and do not trigger return strokes

[*Rhodes and Krehbiel, 1989; Mazur et al., 1995*].

+CG flashes initiate from a positive cloud layer, generally the top one, and involve processes similar to the −CG processes. However, +CGs typically consist of a single return stroke, preceded by a positively charged stepped leader. Unlike negative leaders which are always optically stepped when they propagate in virgin air, positive leaders can move in either a stepped or a continuous fashion [*Rakov and Uman, 2007, p.223*]. Positive strokes are often followed by long-enduring continuing currents, which reach tens of kiloamperes, approximately an order of magnitude more intense than −CG continuing currents [*Rakov and Uman, 2007, p.222*]. +CGs transfer a large amount of positive charge to ground, triggering upper atmospheric breakdown processes known as sprites [*Cummer and Inan, 1997*].

## 1.2 Satellite Observations

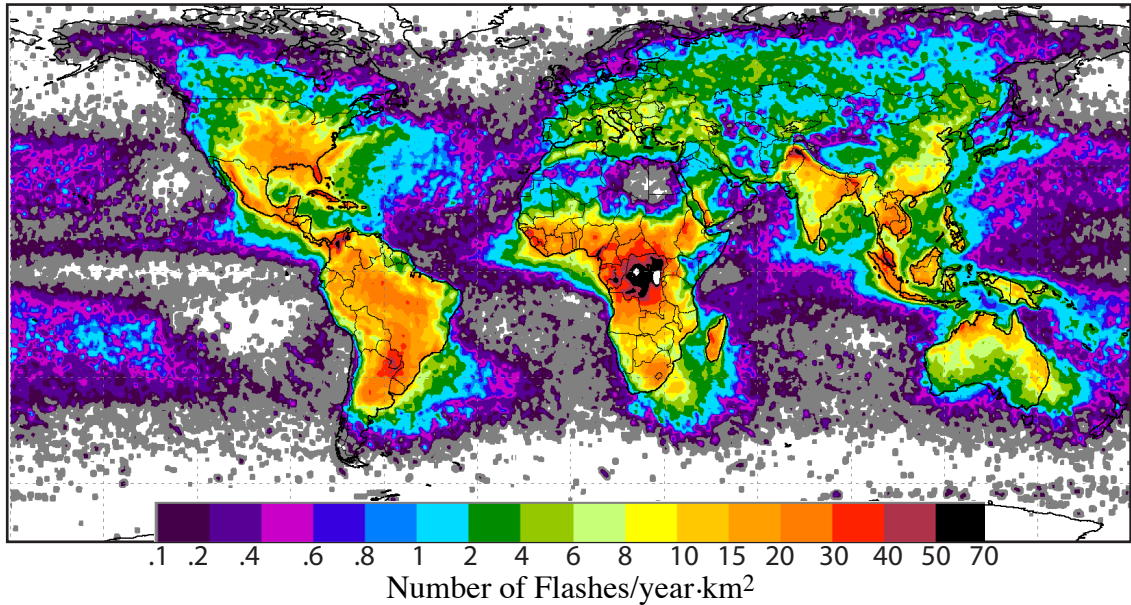


Figure 1.3: Thematic map of lightning occurrence computed using OTD optical data. Adapted from *Christian et al. [2003]*.

Lightning flashes are visible from space, even sometimes to the naked eye, due to the scattering of the optical radiation from the tops of clouds. The Optical Transient Detector (OTD), a space-based charge-coupled device instrument, was deployed in 1995 on the MicroLab-1 satellite. OTD located lightning within its  $1,300 \times 1,300 \text{ km}^2$  field of view as it orbited the globe, with  $\sim 50\%$  detection efficiency and  $\sim 10 \text{ km}$  location accuracy [Christian *et al.*, 2003]. OTD observations helped estimate that nearly 1.4 billion flashes occur annually on our planet and provided lightning climatology maps of the geographical distribution of lightning. Figure 1.3 is a thematic map, adapted from Christian *et al.* [2003], of the global occurrence of lightning, in which red and black mark lightning hot spots as reported by OTD. The observations suggested that the bulk of the activity occurs over land areas with sharp transitions along the coastlines. Physically, the difference in occurrence rate is due to the faster heating of the land surface from daily solar irradiance. The heated surface in turn heats the surrounding air, leading to convection, thunderstorms, and eventually lightning. Land lightning occurrence exhibits large geographical variations, with more frequent flashes in the (warmer) equatorial regions than in polar regions. OTD observations showed that the lightning season in the southern hemisphere lasts from December to February and in the northern hemisphere from June to August. Other space-based lightning sensors have since been deployed such as the Lightning Imaging Sensor, which was launched in 1997 aboard the Tropical Rainfall Measuring Mission satellite (still operational at the time of writing) [Christian *et al.*, 1999; Ushio *et al.*, 2002].

Using optical data from instruments aboard the Vela satellites, Turman [1977] studied extremely intense flashes with optical power  $> 3 \times 10^{12} \text{ W}$ . These flashes were extremely rare, constituting approximately one every 2 million flashes. The detectors located 17 flashes with such optical power between 1972 and 1975. Interestingly, only two out the 17 events ( $\sim 12\%$ ) were over land and the majority ( $\sim 88\%$ ) occurred in oceanic or coastal regions, which is in significant disproportion with the 10:1 land to ocean general lightning occurrence ratio. Up to that point, most of our understanding of lightning stemmed from land observations due to the abundance of land lightning and to the ease of setting up experiments and collecting data over land. The Vela observations however suggested that oceanic lightning, though less frequent, could be

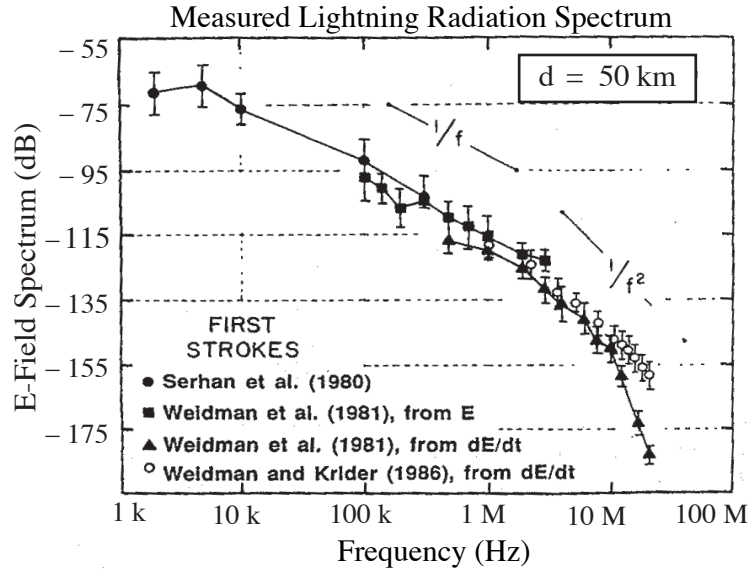


Figure 1.4: Radiation spectrum corresponding to initial negative cloud-to-ground flashes. Adapted from [Uman \[1987, p.118\]](#).

stronger than land lightning. Unfortunately, satellite observations have limited spatial and temporal resolution due to the optical scattering from cloud tops, offering limited information about lightning processes. However, ground-based sensors have been used to collect the electromagnetic radiation from lightning, which contain the radio signatures of the various processes, that could be analyzed to investigate lightning physics.

### 1.3 Ground-based Remote Sensing

Lightning processes radiate impulsive electromagnetic waves from DC to optical frequencies, extending to x-rays and gamma rays. The electromagnetic pulse associated with the return stroke is known as a radio atmospheric, or sferic for short, and its measurement is used as a remote sensing tool to study and to geo-locate lightning. Figure 1.4, adapted from [Uman \[1987, p.118\]](#), shows the typical radiation spectrum of a sferic, which peaks in the Very Low Frequency (VLF; 3–30 kHz) band. VLF



waves propagate efficiently to global distances in the Earth-ionosphere waveguide (as discussed in Section 2.1.1). Various commercial and research-based lightning sensors are used to monitor different frequency bands and are each better at imaging different aspects of lightning. VLF sensors offer global coverage but have limited spatial (several kilometers) and temporal (microseconds) resolution, and are sensitive mostly to the return stroke. On the other hand, Very High Frequency (VHF; 30–300 MHz), Ultra High Frequency (UHF; 0.3–3 GHz), and ground-based optical sensors resolve stepping of the leader channel down to the meter scale, but are limited to line of sight coverage. In this section, we introduce three geo-location networks, each operating in a different frequency range, that we use throughout the thesis in tandem with our VLF/Low Frequency (LF; 30–300 kHz) observations (Section 2.2.1) to study the lightning discharge over both land and oceanic regions.

### 1.3.1 North Alabama Lightning Mapping Array (NALMA)

NALMA consists of 11 VHF receivers deployed across northern Alabama and a base station located at the National Space Science and Technology Center in Huntsville, Alabama [Goodman *et al.*, 2005; White *et al.*, 2013]. The network is operated by the NASA Marshall Space Flight Center, the University of Alabama Huntsville, and New Mexico Tech, and uses a dense array of sensors, providing accurate (10s of meters) three-dimensional maps of the lightning activity [Goodman *et al.*, 2005]. The system remotely senses the sources of impulsive VHF radio signals from lightning by measuring their time of arrival at the different sensors. Typically, hundreds of LMA sources per flash can be reconstructed, producing accurate 3-D lightning channel image maps (latitude, longitude, altitude) with  $\sim 50$  m location error, within 250 km from the center of the LMA network [Goodman *et al.*, 2005; White *et al.*, 2013]. NALMA images intermittent breakdown processes (referred to by the term sources in this work) related to in-cloud or negative stepped leader activity. Positive stepped leaders in +CGs are less likely to exhibit step pulses, as they travel in a more continuous fashion, and do not usually radiate at VHF as strongly as negative leaders. The interest in VHF networks has been growing over the past several years as they provide a

look inside the thundercloud, enabling IC studies and offering a better understanding of breakdown processes. NALMA is limited to line of sight coverage, only collecting lightning data within a few hundred kilometers from the center of the network [Goodman *et al.*, 2005]. Thus, VHF networks are ideal for lightning experiments in localized regions but cannot be used for geographical statistics, and are not available for oceanic studies.

### 1.3.2 National Lightning Detection Network (NLDN)

NLDN, operated by Vaisala, Inc., utilizes LF and VLF sensors and uses time of arrival and magnetic direction finding techniques to provide lightning locations within the continental United States and its vicinity [Cummins and Murphy, 2009; Nag *et al.*, 2011]. The arrival time and azimuth are measured with accuracies of  $\sim 1.5 \mu\text{s}$  and  $\sim 1^\circ$ , respectively. NLDN started in the 1980s as a research network and expanded to cover the contiguous United States by 1989 with the support from the Electric Power Research Institute (EPRI). Later, Global Atmospherics operated NLDN until 2002, when the company was sold to Vaisala, Inc. NLDN is comprised of  $\sim 150$  sensors across the continental USA. The NLDN geo-location accuracy is  $\sim 400$  m with a median error of 308 m [Nag *et al.*, 2011]. The efficient and accurate detection is possible because NLDN sensors are close enough to the source to detect the early part of the ground wave, which is excited by the vertical portion of the return stroke, minimizing polarization errors. The detection efficiency is estimated to be  $\sim 60\text{--}80\%$  for CG strokes and  $10\text{--}20\%$  for ICs [Cummins and Murphy, 2009]. In addition to the lightning type (CG or IC) and polarity (positive or negative), a model-derived measure of peak current  $I_{\text{NLDN}}^{\text{peak}}$  is reported for each event. The peak current estimates are computed using the range-normalized value of the signal strength ( $RNSS$ ) and a simple electromagnetic propagation model given by Cummins *et al.* [1998a], where the  $RNSS$  value relates linearly to  $I_{\text{NLDN}}^{\text{peak}}$ , as shown in Equation 1.1. The model-derived peak current estimates were tested against ground-truth current measurements for  $-CG$  subsequent strokes in triggered lightning at Camp Blanding [Cummins and Murphy, 2009; Nag *et al.*, 2011; Turman *et al.*, 2014]. Generally, NLDN is considered

to be a well established network and is used in various land lightning studies in the continental USA and oceanic lightning studies in neighboring coastal regions (discussed in Section 1.4).

$$I_{\text{NLDN}}^{\text{peak}}(kA) = 0.185 \times RNSS \quad (1.1)$$

### 1.3.3 Global Lightning Detection Network (GLD360)

GLD360 consists of a sparse network of VLF sensors that monitor lightning activity globally [Said *et al.*, 2010]. GLD360, currently operated by Vaisala, Inc., was recently developed by Stanford University and Vaisala, Inc., and has been operating continuously since late 2009. GLD360 geo-locates lightning by processing the efficiently propagating sub-ionospheric VLF component using magnetic direction finding and time-of-arrival methodologies, achieving a 57% flash detection efficiency and 2–5 km accuracy, though the source type (IC versus CG) is not currently reported [Said *et al.*, 2010, 2013]. Using a propagation correction model, the network uses the magnetic field intensity measured at each sensor to estimate the peak current  $I_{\text{GLD360}}^{\text{peak}}$  of each detected stroke with 21% (6%) arithmetic mean (geometric mean) magnitude error [Said *et al.*, 2010, 2013]. The validation was done by comparing GLD360 peak current estimates with those reported by NLDN.

Using one year of GLD360 data with 353 million flashes, Said *et al.* [2013] produced a thematic map of lightning occurrence around the globe, which we present in Figure 1.5, with lightning hot spots color coded in red. The occurrence map is in general agreement with Figure 1.3, which was generated using OTD satellite observations. Unlike satellites, which monitor a narrow region within their field of view, GLD360 is the first network to report the location and peak current measurements  $I_{\text{GLD360}}^{\text{peak}}$  for individual discharges around the world, enabling global lightning statistics including in deep oceanic regions.

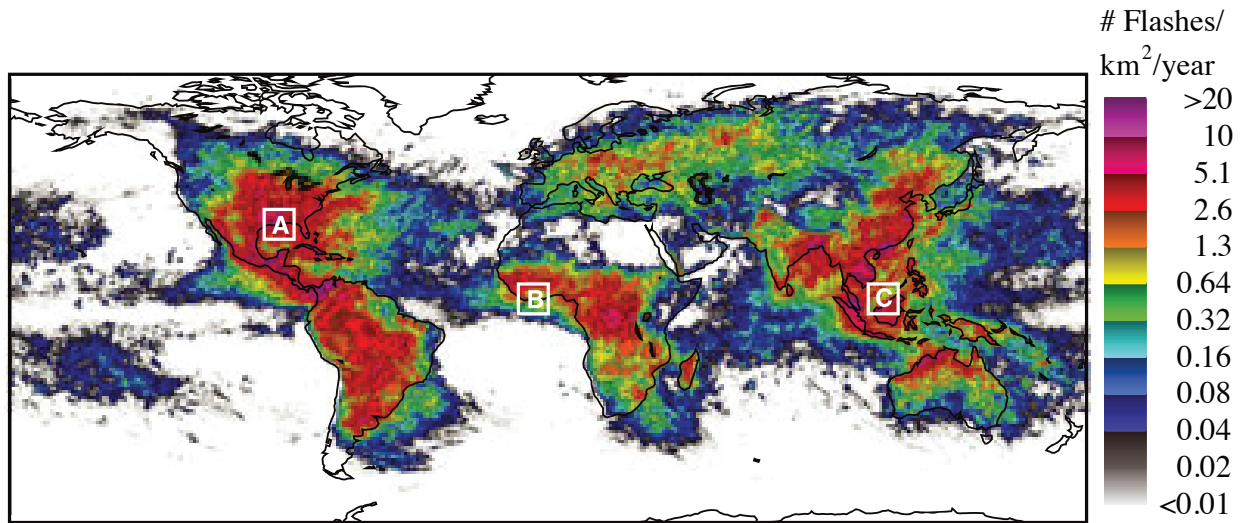


Figure 1.5: Thematic map of lightning occurrence computed using one year of GLD360 data. Adapted from [Said et al. \[2013\]](#).

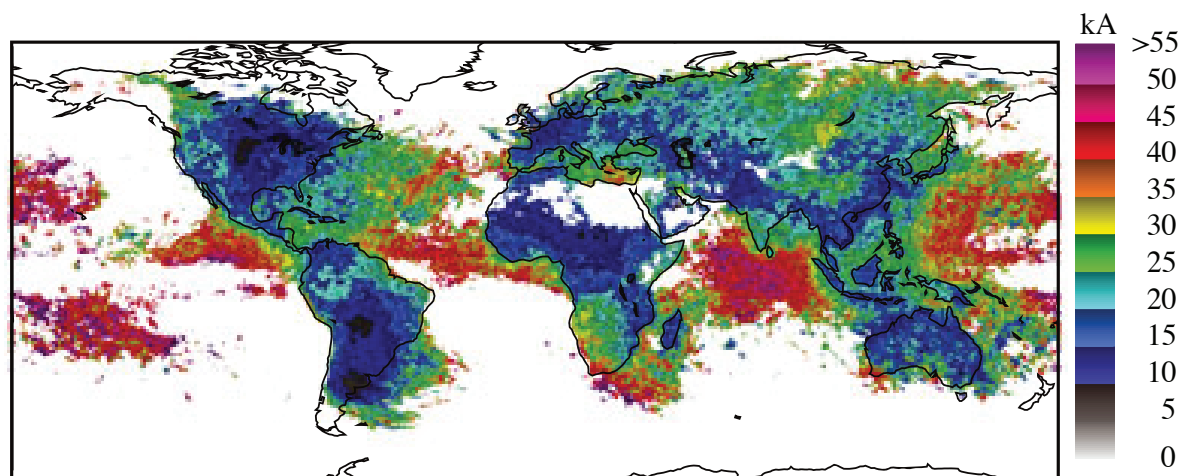


Figure 1.6: Thematic map of the geometric mean of peak current estimates computed using one year of GLD360 data. Adapted from [\[Said et al., 2013\]](#).

## 1.4 Oceanic lightning

The Vela observations from Section 1.2 suggested that the strongest lightning events occurred over the ocean and in coastal regions. Using the same GLD360 dataset from the previous section, *Said et al.* [2013] computed the first thematic map of the distribution of the geometric mean of  $I_{\text{GLD360}}^{\text{peak}}$  around the globe, which we show in Figure 1.6. The results suggested stronger lightning over the ocean than over land. The geometric mean of peak current estimates for ocean –CG flashes were 22%-88% higher compared to land –CGs, with sharp transitions in peak currents along the coastlines. *Said et al.* [2013] analyzed three coastal regions, marked using three boxes, shown on the map of Figure 1.5, in which the peak current enhancement is extremely sharp at the land-ocean boundaries. The authors however admitted that the fairly new GLD360 network could be overestimating peak currents in deep ocean regions [*Said et al.*, 2013]. Nevertheless, the sharp transition at land-ocean boundaries suggested that higher oceanic peak currents may result from physical differences between oceanic and land lightning and cannot be solely attributed to network inaccuracies. To date, GLD360 provides the highest resolution (spatial and temporal) global lightning dataset, especially in deep oceanic regions, which we thoroughly investigate in this thesis.

Multiple other geo-location systems reported similar observations, showing stronger lightning over the ocean with sharp peak current increase at the coastlines. *Lyons et al.* [1998] used NLDN data to show that the median NLDN-reported peak currents for initial –CG strokes were higher over salt water than over land. *Orville and Huffines* [2001] aggregated ten years of NLDN data between 1989 and 1998 that suggested that peak current magnitudes increased from 27 kA to over 30 kA along the eastern coast of the United States and the Gulf coast. *Orville et al.* [2011] aggregated nine years of NLDN data between 2001 and 2009 and presented thematic maps of median peak currents of first and subsequent positive and negative CGs in the continental U.S. and along the eastern coastlines. The study showed that estimated peak currents of initial –CGs were higher over the ocean than over land, with sharp transitions at the land-ocean boundaries. Various other studies [*Cummins et al.*, 2005;

[Cummins and Murphy, 2009](#); [Orville et al., 2011](#); [Hutchins et al., 2013](#)] supported these findings, reporting stronger strokes over the ocean with sharp transitions along much of the coastlines. Interestingly, these studies found that, in contrast to initial  $-CGs$ , the sharp increase in peak currents at the land-ocean boundaries was absent for subsequent  $-CGs$  that follow a dart-leader (occur in the same channel as the previous stroke) and for  $+CGs$ .

The reasons for the differences and similarities between ocean and land lightning remain poorly understood. Many interpretations of the statistical results have been offered. The general agreement in the lightning community is that the observed pattern, in which the sharp transition in estimated peak currents is only observed for initial  $-CGs$ , is due to physical differences in the attachment process or in the initiation of  $-CGs$  over the ocean. In this work, we investigate different competing explanations for the larger lightning radiation from ocean discharges.

The stronger radiation could result from more intense flashes in oceanic thunderstorms due to meteorological or climatological differences over the ocean. For instance, [Cooray et al. \[2013\]](#) suggested that the enhancement is due to differences in the charge structure in ocean clouds as the unfavorable conditions for the formation of the lower positive charge layer lead to lightning initiation at higher cloud potentials (compared to their land counterparts), resulting in larger peak currents in initial  $-CGs$ . The authors argued that the positive charge pocket does not promote  $+CG$  strokes, which is consistent with the lack of enhancement in  $+CGs$ .

Some explanations are based on the fact that radiation fields could be enhanced without increasing the magnitude of the lightning currents inside the channel. For example, a faster return stroke velocity or a shorter channel-base current rise-time, resulting from physical differences in the attachment process or from the high conductivity of salt water, could lead to higher radiation fields at the receiver. [Cummins et al. \[2005\]](#) argued that the primary effects are in the attachment process to salt water in initial  $-CGs$ , following a stepped leader, which agreed with the lack of peak current enhancement in  $-CGs$  that stroke freshwater over the Great Lakes area in the United States [[Cooray et al., 2013](#); [Said et al., 2013](#)].

Other explanations involve possible effects due to more efficient ground wave propagation over highly conductive salt water than over land. [Cooray and Rakov \[2011\]](#) studied the effect of ground conductivity on the return stroke current and concluded that the influence of the ground conductivity on wave propagation is minimal.

In addition, the geographical distribution of the network sensors could lead to systematic biases in peak current estimation due to the non-uniform coverage of deep oceanic areas and to sensor range filtering, which could skew the distribution of detected peak currents toward more powerful (higher peak current) events [[Saïd et al., 2013](#)].

## 1.5 Contributions of this Work

In this dissertation, we use radio remote sensing in tandem with statistical techniques, that leverage the frequent occurrence of lightning, to investigate the physics of the lightning discharge. We analyze millions of VLF, LF, and VHF geo-location data points to study statistical patterns that offer new insight into the physics of lightning. We process thousands of waveforms for ocean and land based lightning to study the poorly-understood differences and similarities between ocean and land lightning. The following scientific advances are reported in this dissertation:

- An electromagnetic lightning Low Frequency radiation and propagation model was developed and the impact of various current parameters on the ground waveform was quantified.
- The inverse of the electromagnetic model was computed using a system of neural networks. The inverse model was applied to thousands of experimental waveforms, collected during a ship-borne Low Frequency receiver experiment. The channel-base current rise-times and return stroke speeds were estimated and found to be similar for ocean and land lightning.
- Source factors that contribute to the observed peak current increase in oceanic lightning were studied, using Low Frequency waveforms and years of geo-location

data points. The increase in peak currents was found to be sudden at the land-ocean boundaries across all stroke types, orders, and polarities.

- A new physical process within lightning discharges was proposed, based on the linkage between initial and subsequent cloud-to-ground return strokes via cloud leaders.
- The flashing probability and recovery as a function of distance and time was quantified following various lightning types and polarities. A proxy method was introduced for monitoring the charging processes in the cloud using lightning locations.



# Chapter 2

## Lightning Detection Instrumentation

### 2.1 Electromagnetic Remote Sensing

In the absence of boundaries, electromagnetic waves radiated by lightning would travel in all directions, propagating in an *unguided* spherical fashion, with the power intensity inversely proportional to the square of the distance from the source. Interestingly, in VLF and Extremely Low Frequency (ELF; 0.3–3 kHz), both the Earth and the ionosphere, which is the ionized layer of the atmosphere, act as good conductors, efficiently guiding radio waves to global distances (few dB attenuation per 1,000 km), in what is known as the Earth-ionosphere waveguide. Without the ionosphere, electromagnetic waves radiated by the lightning discharge would propagate only along the ground with some diffraction, which would significantly limit the range of spheric detection. However, in the Earth-ionosphere waveguide, the electromagnetic energy is *guided* around the globe, using successive reflections off of the two boundaries. The efficient propagation of VLF and ELF waves has lead to many applications varying from long-range communication with submerged submarines (because at ELF/VLF waves penetrate into salt water), to remote sensing studies of the ionosphere, and to the development of global lightning geo-location systems. In this section, we introduce the electric properties of Earth and the ionosphere, we present the ideal parallel

plate waveguide, which could be used as a simple model to approximate the fields that are formed between the ground and the ionosphere at long distances, and we describe the ray hop model, which we use in this dissertation to study sferics at short distances.

### 2.1.1 Earth and the ionosphere

Practically, all dielectric media are lossy, exhibiting finite conductivity  $\sigma$ , which attenuates electromagnetic waves. In a simple medium, the electric field  $\mathbf{E}$  of the wave establishes conduction currents equal to  $\sigma\mathbf{E}$ , dissipating the radio energy in the form of heat, with the quantity  $\sigma|\mathbf{E}|^2$  representing power dissipated per unit volume. When the wave frequency is small enough, charge flows in the lossy medium, efficiently reflecting the incident wave. The power series approximation of the magnitude of the reflection coefficient  $R$ , which is the ratio between the incident and reflected wave, is given by Equation 2.1 [Inan and Inan, 2000, p.204], where  $\omega$  is the angular frequency,  $\epsilon_0 = 8.8542 \times 10^{-12}$  H/m is the electric permittivity of free space, and  $\epsilon_r$  is the relative permittivity of the dielectric. The medium is thus considered a good conductor when  $\sigma \gg \omega\epsilon_0\epsilon_r$ , reflecting the incoming waves with minimal losses.

$$|R|^2 \simeq 1 - 4\sqrt{\frac{\omega\epsilon_0\epsilon_r}{2\sigma}} \quad (2.1)$$

The conductivity of Earth depends on the terrain and soil composition, with sharp variations between dry land and salt water. Over the continents, typical ground conductivities vary between  $10^{-4}$  S/m and  $10^{-2}$  S/m with relative permittivity  $\epsilon_r^{\text{land}} \simeq 3.4$ . Over the oceans, the typical ground conductivity is 5 S/m, approximately three orders of magnitude larger than over land, but still much lower than metallic conductors such as silver with  $\sigma_{\text{silver}} = 6.3 \times 10^7$  S/m, and the relative permittivity of salt water is  $\epsilon_r^{\text{sea}} \simeq 81$ . The good conductor assumption is satisfied over salt water for frequencies up to 100s of MHz, while over land it only holds for VLF frequencies and starts to fail in the LF band, leading to waveform attenuation. The impact of ground conductivity on wave propagation should thus be carefully modeled, especially when comparing waveforms of signals that have propagated over land versus over the ocean

(discussed in Chapter 3).

The ionosphere is the layer of the Earth's atmosphere that is ionized by solar and cosmic radiation, lying 75 to 1,000 km above the surface of the ground. This ionized region contains free electrons and positive ions and is usually characterized by a varying electron density profile with altitude  $N_e(z)$ . The electron densities are maintained by solar radiation during the day and cosmic rays and electron precipitation during the night, leading to high variability [Hargreaves, 1992, p.223]. The ionosphere is important to satellite and long-distance radio communication that operate at frequencies where the ionosphere reflects waves efficiently (typically in the VLF band and below). The reflection height is highly variable and depends on the the density profile, the frequency of the incident wave, and the angle of incidence. At vertical incidence, the waves are reflected at the altitude where the plasma frequency  $\omega_p = \sqrt{N_e(z)q^2/m_e\epsilon_0}$  is equal to  $\sqrt{\omega\nu}$ , where  $q$  and  $m$  are the charge and mass of an electron, respectively, and  $\nu$  is the electron collision frequency [Ratcliffe, 1959, p.110]. At VLF, both the theoretical and experimental reflection heights are consistent and are found to be  $\sim 65$  km during the day [Rasmussen et al., 1980] and  $\sim 85$  km at night [Thomson, 1993].

### 2.1.2 Ideal Parallel Plate Waveguide

The ideal parallel plate waveguide is the simplest example of a waveguide, consisting of two parallel perfectly conducting boundaries, *guiding* radio waves freely in the direction parallel to the two plates. Figure 2.1 illustrates an example setup of the waveguide with the two boundaries located at  $z = 0$  and  $z = h$ , between which a linearly polarized wave is guided in the  $x$ -direction, with wavenumber  $k = \omega/c$ , where  $c$  is the speed of light. The propagation of the waves is governed by Maxwell's equations subject to the following boundary conditions at  $z = 0, h$ :

$$E_{\text{tangential}} = 0, \quad H_{\text{normal}} = 0 \quad (2.2)$$

Generally, the solutions of the wave equations in a waveguide are divided into modes, depending on the components of the magnetic and electric field in the direction

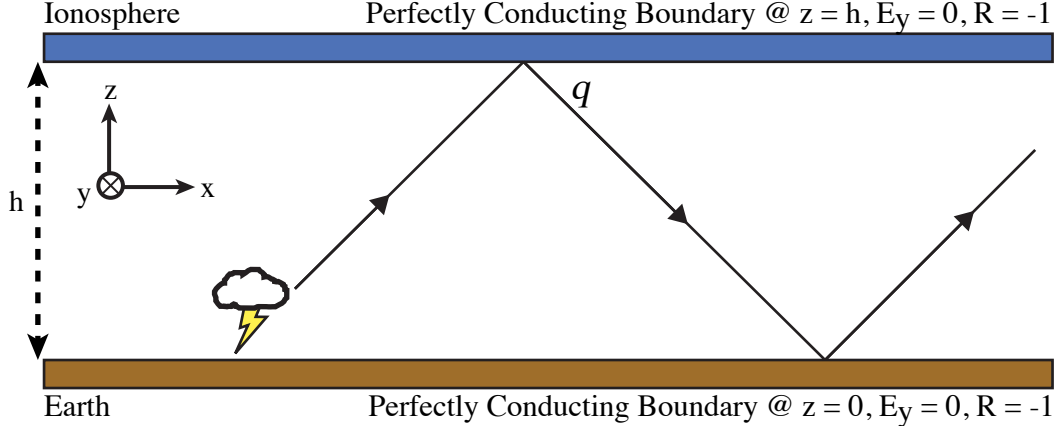


Figure 2.1: Setup of the ideal parallel plate waveguide.

of propagation, and are of three types:

- Transverse Electric (TE):  $E_x = 0, H_x \neq 0$
- Transverse Magnetic (TM):  $E_x \neq 0, H_x = 0$
- Transverse Electromagnetic (TEM):  $E_x = 0, H_x = 0$

For instance, in Figure 2.1, the electric field  $\mathbf{E}$  is in the  $y$ -direction with  $E_y \neq 0, E_x = E_z = 0$ . Given that the reflection coefficient  $R$  equals to  $-1$  at both boundaries at  $z = 0, h$  [Inan and Inan, 2000, p.123], the total electric field  $E_y^{\text{total}}$  is found by adding up the incident wave  $E_y^{\text{incident}}$  and the reflected wave  $E_y^{\text{reflected}}$  as shown in Equation 2.3–2.5.

$$E_y^{\text{total}} = E_y^{\text{incident}} + E_y^{\text{reflected}} \quad (2.3)$$

$$E_y^{\text{total}} = E_0 e^{-jk(x \cos \theta - z \sin \theta)} e^{j\omega t} - E_0 e^{-jk(x \cos \theta + z \sin \theta)} e^{j\omega t} \quad (2.4)$$

$$E_y^{\text{total}} = 2j E_0 \sin(kz \sin \theta) e^{\omega t - kx \cos \theta} \quad (2.5)$$

The boundary conditions are satisfied when  $kz \sin\theta = n\pi$ , and  $n$  is an integer. This condition is known as the *mode equation* and determines the values of  $\theta$  for which a perpendicularly polarized plane wave can exist, for a given frequency  $\omega$ . Each mode has a cut-off frequency  $\omega_n^{\text{cutoff}}$  below which the solution of the *mode equation* is complex, resulting in decaying waves known as evanescent waves. Equation 2.6 shows the cut-off frequencies for the different modes, which solely depends on the height of the waveguide  $h$ , which in our application is the reflection height of the ionosphere. We note that the TEM mode is equivalent to the  $\text{TM}_0$  waves, which have a zero cut-off frequency, indicating that TEM modes could be established for any frequency. However,  $\text{TE}_0$  waves do not exist for the parallel plate waveguide due to the boundary conditions and the geometry which cannot be satisfied for a nontrivial solution.

$$\omega_n^{\text{cutoff}} = \frac{n c \pi}{h} \quad (2.6)$$

An infinitesimal current element, known as a Hertzian dipole [Budden, 1961, p.41], could inject waves in the waveguide, which interact with the two boundaries and establish waveguide modes. The fields  $\mathbf{E}$  and  $\mathbf{H}$  can be derived using the Hertzian dipole, with charge  $q$  and length  $l$ , and are usually expressed using the Hertz vector  $\mathbf{U}$  [Stratton, 1941, p.28]. Using expressions from Budden [1961, p.43], we could study the way different lightning types would establish modes in the parallel plate approximation of the Earth-ionosphere waveguide. We find that, in this ideal setting, vertical sources can only excite  $\text{TM}_n$  and TEM waveguide modes. The relative amplitudes of these modes are given by the ‘excitation factors’, which depend on the height of the source and  $\theta_n$  is the propagation angle for the  $n^{\text{th}}$  mode. Horizontal sources on the other hand can only excite  $\text{TE}_n$  modes. Thus, vertical sources, such as a vertical CGs, excite the propagating TEM mode, while horizontal sources, such as horizontal ICs, cannot excite the TEM mode, limiting the range and efficiency of long-range detection of ICs as they do not launch any propagating waves below the first cut-off frequency. In addition, horizontal sources near the ground have zero height gain functions and do not excite any modes as the radiation from the horizontal source is canceled out by the current image below the perfectly ground.

### 2.1.3 Earth-ionosphere Waveguide

The Earth-ionosphere waveguide is however very different from an ideal parallel plate waveguide. The lossy ground boundary, the frequency-dependent and lossy ionospheric properties, and the anisotropy due to the Earth's tilted magnetic field lead to considerable deviations from this simple model. The Earth and the ionosphere are not perfect conductors, attenuating differently the various frequency content of the propagating waves. The ionospheric reflection height is also frequency dependent which distorts the signal as it reflects at different heights, with higher frequencies traveling longer distances.

The tilted geomagnetic field of Earth renders the ionosphere an anisotropic plasma. The extent of the effects of this anisotropy depend on the relative values of the electron-neutral collision frequency and the electron gyro-frequency. Remarkably, this asymmetry leads to different attenuations for waves that propagate East-ward than waves that propagate West-ward. This difference in attenuation peaks between 1 and 4 kHz, reaching a difference of  $\sim 45$  dB/1000 km [[Barr, 1971](#)]. Additionally, due to the coupling of the wave polarization at the anisotropic ionospheric boundary, pure TM and TE modes cannot exist in the Earth-ionosphere waveguide. Instead, the propagating waves are a superposition of quasi-TM and quasi-TE modes. QTE (QTM) modes are similar to TE (TM) modes but they have a small electric (magnetic) field component in the direction of propagation [[Budden, 1961](#), p.151].

### 2.1.4 Ray Hop Components

At global distances, the modeling of waveguide modes established by lightning is essential for analyzing and studying VLF and ELF sferics. At such large distances, the frequency content above VLF is significantly attenuated, limiting the extracted information and the range of physical processes that could be studied to VLF resolution of  $\sim 10 \mu\text{s}$ . In this dissertation, we are interested in studying physical processes with shorter timescales, such as the speed of propagation of the return stroke along the channel, requiring broadband LF waveforms collected at short distances from the lightning source, within several hundred kilometers. At these distances, we could

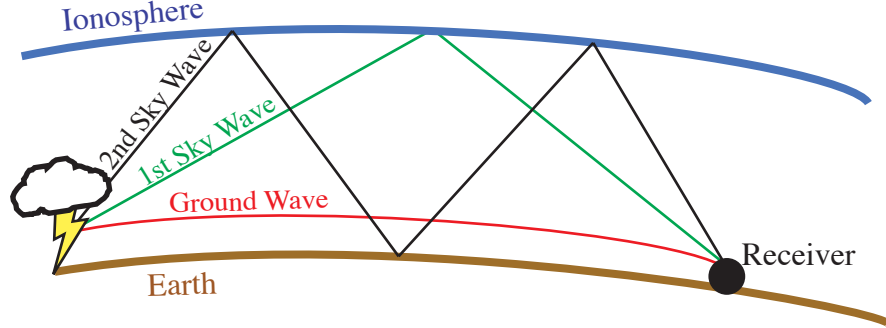


Figure 2.2: Cartoon of the ground wave (in red), 1<sup>st</sup> sky wave (in green), and 2<sup>nd</sup> sky wave propagation to the receiver.

model the lightning waveform as a series of ray hops, consisting of the ground wave, which is the signal that propagates along the ground directly from the source to the receiver, the first sky wave, which reflects off of the ionosphere once before reaching the receiver, the second sky wave, which reflects twice off of the ionosphere, and so on.

Figure 2.2 illustrates the ray hop model with the ground wave, 1<sup>st</sup> sky wave and 2<sup>nd</sup> sky wave components, and Figure 2.3 shows a sample waveform received at 300 km from the source lightning, consisting of these three components. The ground wave diffracts over the Earth and has the shortest distance, arriving first to the receiver, followed by the 1<sup>st</sup> sky wave, and the 2<sup>nd</sup> sky wave. The extra distance traveled by the  $m^{\text{th}}$  sky wave with respect to the distance  $d_g$  traveled by the ground wave is denoted by  $\Delta d_m$  and is given by Equation 2.7–2.8, where  $h$  is the reflection height of the ionosphere and  $R_E$  is the radius of Earth. This extra distance corresponds to a time delay  $\Delta t_m = \Delta d_m/c$ , which decreases with distance due to geometry. For instance, during the day (assuming  $h = 65$  km), the time delay between the onset of the ground wave and the 1<sup>st</sup> sky wave is  $\sim 179 \mu\text{s}$  at 300 km, consistent with the experimental time delays in Figure 2.2, and decreases to  $\sim 104 \mu\text{s}$  at 500 km, to  $\sim 80 \mu\text{s}$  at 750 km, and to  $\sim 70 \mu\text{s}$  at 1,000 km, as the differential distance decreases.

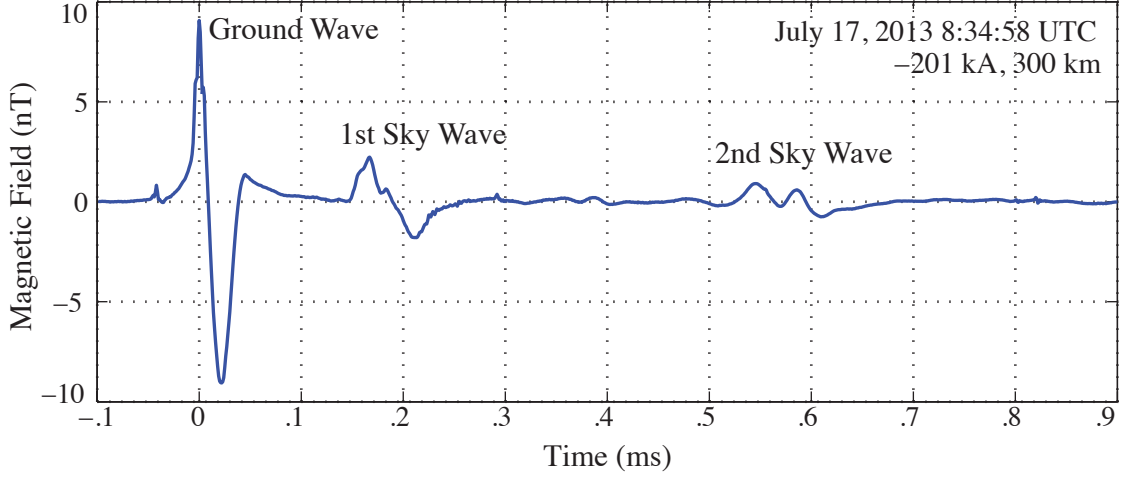


Figure 2.3: Sample experimental LF lightning waveform.

$$\Delta d_m = 2n\sqrt{2R_E^2 + h^2 + 2hR_E - 2R_E(R_E + h)\cos(\frac{d_g}{2nR_E})} - d_g \quad (2.7)$$

$$\Delta d_m \simeq \sqrt{4n^2h^2 + d_g^2 + hd_g^2/R_E} - d_g \quad (2.8)$$

The sferic thus could be expressed as a summation of the electromagnetic components from the ground wave and sky waves as shown in Equation 2.9, where  $B_g(t)$  corresponds to the magnetic field of the ground wave and  $B_m(t)$  is the magnetic field of the  $m^{\text{th}}$  sky wave.

$$s(t) = B_g(t) + \sum_{m=1}^{\text{inf}} B_m(t) \quad (2.9)$$

Assuming a vertical dipole and that both the transmitter and the receiver are on the ground, the fields of the  $m^{\text{th}}$  sky wave could be approximated using Equation 2.10 [Watt, 1967, p.204].

$$B_{m,\gamma} = B_0 \cos\alpha C_t C_r f_m R_m(\gamma, \alpha) \frac{e^{-jk\Delta d}}{1 + \Delta d/d} \quad (2.10)$$



The magnetic field  $B_{\gamma,m}$  depends on the angle of incidence  $\gamma$  of the wave onto the ionosphere, the launching angle  $\alpha$  at the ground, the antenna factor ( $\cos\alpha$ ) resulting from the orientation of the dipole, correction factors  $C_t$  and  $C_r$  that account for finite conductivity at the transmitter and receiver, respectively, the focusing term  $f_m$  which transforms the energy spread from  $1/d$  (spherical) to  $1/R_E \sin(d/R_E)$  [Saïd, 2009, Appendix D], the impact of reflecting from the ground and the ionosphere denoted by  $R_m$ , and the extra phase and spread due to the extra propagation distance  $\Delta d_m$  compared to direct propagation. However, in this dissertation, we focus on studying lightning using only the ground wave portion of the lightning waveform, which contains the direct signature of the nearby stroke. The ground wave could be used to infer properties of the current profile inside the source lightning channel without modeling ionospheric interactions that vary with frequency, time of day, which would require more assumptions (such as the ionospheric profile with altitude) and modeling. As we visually inspect Figure 2.3, we find that the ground wave is larger in amplitude and contains sharper features than the sky waves, indicating that at short distances the ground wave contains higher frequency content and have larger signal to noise ratio than the sky waves. In Chapter 3, we present a computationally efficient physical model, which we use to simulate the ground wave portion of the lightning waveform, resulting from an arbitrary current profile along the lightning channel while accounting for propagation physics, and allowing us to analyze the impact of various lightning parameters and propagation path properties on the received lightning waveform.

In the next section, we introduce our high-sensitivity LF instrument which we use to collect LF ground waves close to deep oceanic lightning. Ideally, the system should be positioned within  $\sim 750$  km of the lightning source, close enough to pick up the LF content and to ensure a large enough time separation between the ground wave and 1<sup>st</sup> sky wave.

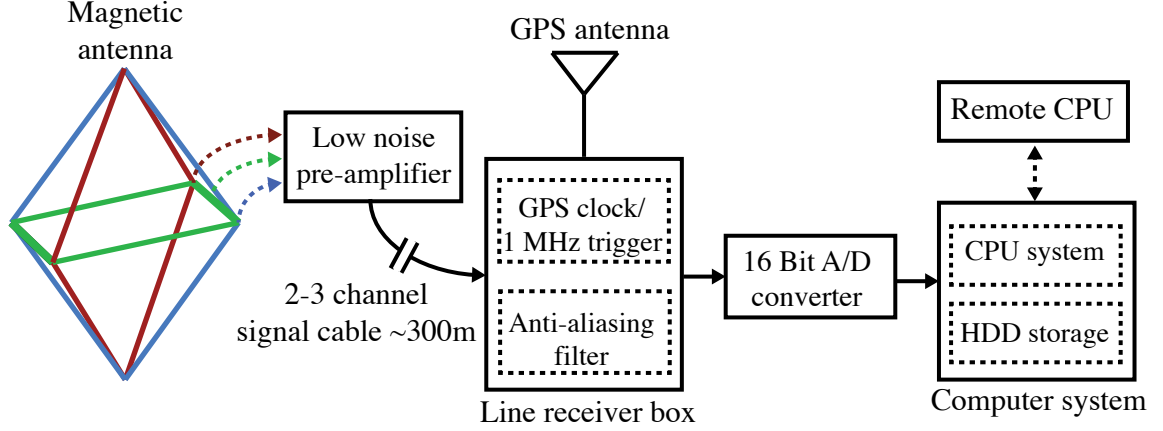


Figure 2.4: Block diagram of the LF receiver system.

## 2.2 Data Acquisition

### 2.2.1 Low Frequency Receiver

We use the Atmospheric Weather Electromagnetic System for Observation Modeling and Education (AWESOME) instrument to collect lightning waveform data. Our highly sensitive broadband LF instrument is similar to the 100 kHz sampling rate VLF version described by [Cohen *et al.*, 2010] but operates at 1 MHz, collecting waveforms with higher frequency content and temporal resolution. Figure 2.4 shows a system diagram of the LF receiver. The magnetic field sensor consists of three wire-loop antennas, orthogonal to each other, and each sensitive to the component of the magnetic field through the plane of the loop. Since the vertical magnetic fields are typically much smaller near the ground, the horizontal loop (shown in green) is not required in many applications. Most receiver sites have two orthogonal antennas designated as Channel 1 and Channel 2, each consisting of  $n$  turns of wire with an air core [Ramo *et al.*, 1994, p.116-117]. The induced voltage  $V$  across the terminals of the antenna is given by Faraday's law, shown in Equation 2.11, where  $\mathbf{A}$  is the area vector of the loop antenna pointing in the direction normal to its surface.

$$V = -n \frac{\partial}{\partial t} \int \mathbf{B} \cdot d\mathbf{A} \quad (2.11)$$

The magnetic antenna is preferably installed in a ‘quiet’ location, away from sources of electric noise, such as power lines. The induced voltage goes through the low noise amplifier and through a long signal cable to reach the indoor line receiver box, which is typically set up a few hundred meters away from the antenna to reduce electronic coupling and to avoid sources of noise in buildings. The signals from the two channels pass through the analog anti-aliasing filter cards and are then sampled using a multi-channel Analog to Digital (A/D) converter from National Instruments, with 16 bit precision providing a 500 kHz bandwidth. The timing of the 1 MHz trigger of the A/D converter is synchronized for all sites using a Global Positioning System (GPS) unit. The impedance of the antenna and the input impedance of the line receiver have to be matched to make sure that the output voltage is frequency independent for the desired range of operation [Harriman *et al.*, 2009] and the gains of the pre-amplifier and line receiver are adjusted such that the output voltage of the line receiver falls between  $\pm 5$  V to prevent clipping. The receiver sends one second of LF data daily back to a central processor at Stanford University to help monitor data quality. The system produces  $\sim 63$  TB of data/channel/year, ten times more than the VLF version, and only stores permanent broadband LF waveforms for return strokes reported by NALMA, NLDN, and GLD360. The design details of the LF instrument, which was mainly developed to study oceanic lightning in this dissertation, are yet to be published, but are similar to the designs of the VLF version [Cohen *et al.*, 2010].

Figure 2.5 illustrates the process by which we extract the angle of arrival and the magnetic waveforms corresponding to lightning. The incoming sferic excites the two vertical antennas, inducing signal  $B_1(t)$  and  $B_2(t)$  in Channel 1 and 2, respectively. The sampled waveform signals  $\mathbf{B}_1$  and  $\mathbf{B}_2$  ( $T_s = 1 \mu\text{s}$ ) are used to compute the azimuthal magnetic field  $\mathbf{B}_{az}$  and the angle of arrival of the waveform  $\theta_{\text{calc}}$ . First, the calibrated signals are plotted parametrically and the angle of arrival  $\theta_{\text{calc}}$  is computed using a linear least squares approach given by Equation 2.12–2.17, similar to the one

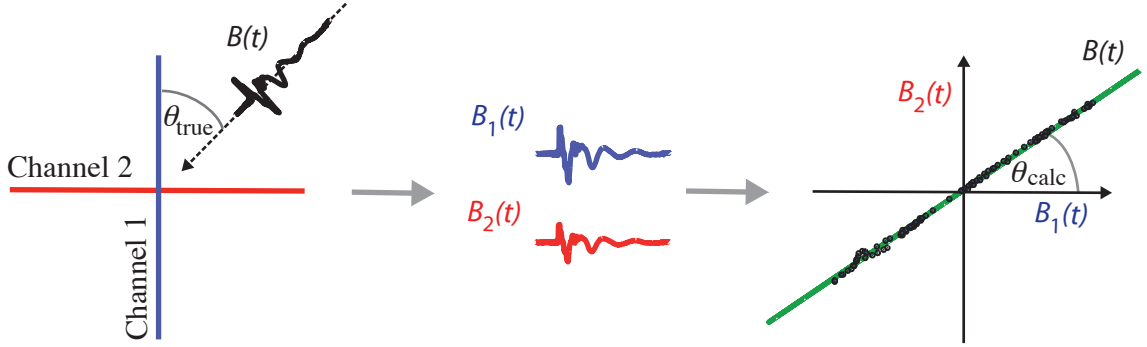


Figure 2.5: Cartoon of the spheric rotation process. (Left panel) The incoming wave  $B(t)$  is captured by the two orthogonal antennas. (Right panel) The axes are digitally rotated to find the azimuthal magnetic field.

used by [Saïd \[2009, p.77\]](#). The computed arrival angle  $\theta_{\text{calc}}$  is extracted using the best-fit line (in green). The axes are then digitally rotated to find the azimuthal magnetic field  $\mathbf{B}_{\text{az}}$ , containing most of the energy, and the (negligible and often overlooked) radial component  $\mathbf{B}_{\text{rad}}$ .

$$u = \frac{\mathbf{B}_2^t \mathbf{B}_2 - \mathbf{B}_1^t \mathbf{B}_1}{2 \mathbf{B}_2^t \mathbf{B}_1} \quad (2.12)$$

$$a_{1,2} = u \pm \sqrt{1 + u^2} \quad (2.13)$$

$$a_{1,2}^{\text{error}} = \frac{\|\mathbf{B}_2 - a_{1,2} \mathbf{B}_1\|_2^2}{1 + a_{1,2}^2} \quad (2.14)$$

$$\theta_{\text{calc}} = \begin{cases} \arctan(a_1) & : a_1^{\text{error}} \leq a_2^{\text{error}} \\ \arctan(a_2) & : a_1^{\text{error}} > a_2^{\text{error}} \end{cases} \quad (2.15)$$

$$\mathbf{B}_{\text{az}} = \mathbf{B}_1 \cos \theta_{\text{calc}} + \mathbf{B}_2 \sin \theta_{\text{calc}} \quad (2.16)$$

$$\mathbf{B}_{\text{rad}} = \mathbf{B}_2 \cos \theta_{\text{calc}} - \mathbf{B}_1 \sin \theta_{\text{calc}} \quad (2.17)$$

### 2.2.2 Instrumentation Errors and Correction

Various sources of instrumentation errors could impact the LF measurements. Ideally the two antenna loops are in an orthogonal position with identical post-calibration gains. However, coupling between the electronics and the loops, electronic mis-calibrations, and differences in the size and shape of the two loops could lead to a mismatch in the gain of the two channels. We apply a statistical technique, similar to the one used by [Wood \[2004\]](#), to mitigate the effects of these errors. We denote by  $\alpha$  the ratio of the magnitude response of the first channel to the one of the second channel and by  $\xi$  the angle from which the ideally orthogonal antenna loops skew from orthogonality, due to slight inaccuracies in the setup. These errors, which are functions of frequency, lead to a sinusoidal variation in the calculated angle of arrival  $\theta_{\text{calc}}$  given by Equation 2.18, affecting the amplitude and shape of the collected waveform.  $\rho$  is a constant angle that specifies the orientation of the plane of the first antenna loop (Channel 1) with respect to geographical North.

$$\theta_{\text{true}} = \tan^{-1} \left[ \alpha \frac{\theta_{\text{calc}}}{\cos \xi} - \tan \xi \right] + \rho \quad (2.18)$$

We compute the optimal values for  $\alpha$ ,  $\xi$ , and  $\rho$  by solving for the ‘best’ nonlinear least-square fit, as shown in Equation 2.19. To do so, we collect  $N$  (usually  $N$  is in the thousands) lightning waveforms around the receiver, each corresponding to an NLDN event. We compute  $\theta_{\text{true}}$  for each event using the NLDN-reported location and  $\theta_{\text{calc}}$  using the collected waveform, following the approach introduced in the previous section. We then use the Levenberg-Marquardt algorithm (discussed in Section 4.2.2) to solve for  $\alpha$ ,  $\xi$ , and  $\rho$ . Once we have  $\alpha$  and  $\xi$ , we reprocess our LF data using the corrected  $\mathbf{B}_1^c$  and  $\mathbf{B}_2^c$  given in Equation 2.20 and 2.21, accounting for any gain offset and a skewed antenna geometry.

$$\arg \min_{\alpha, \xi, \rho} \chi^2(\alpha, \xi, \rho) = \sum_{i=1}^N \|\theta_{\text{true}} - f(\theta_{\text{calc}}; \alpha, \xi, \rho)\|^2 \quad (2.19)$$

$$\mathbf{B}_1^c = (1/\alpha) \mathbf{B}_1 \quad (2.20)$$

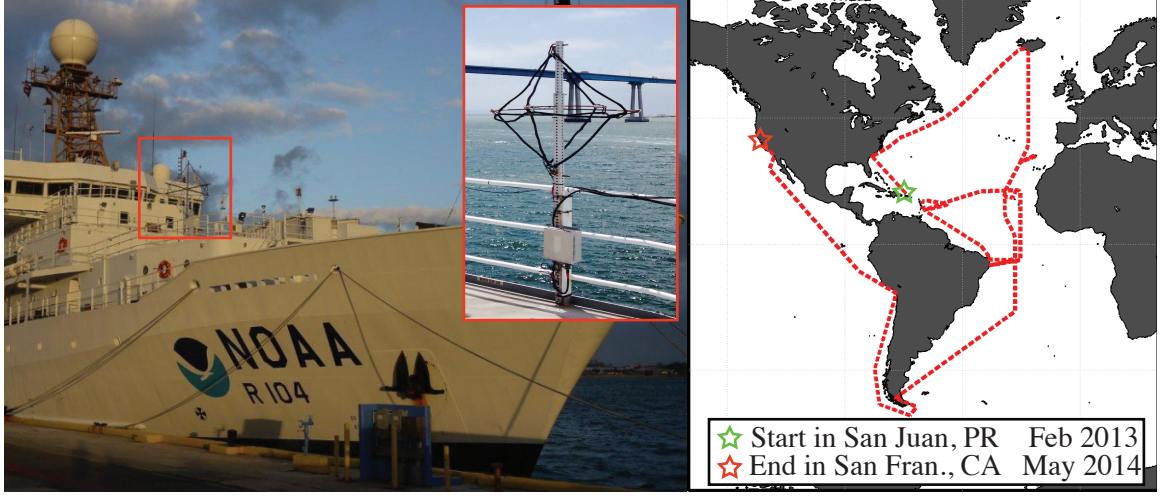


Figure 2.6: (Left) Picture of the Ronald W. Brown research vessel and the LF receiver magnetic loop antenna. (Right) Map of the voyage starting from San Juan, Puerto Rico in February 2013 and ending in San Francisco, California, U.S. in May 2014.

$$\mathbf{B}_2^c = (-\tan \xi / \alpha) \mathbf{B}_1 + (1 / \cos \xi) \mathbf{B}_2 \quad (2.21)$$

## 2.3 Oceanic LF Observations

Our aim is to investigate the physics of deep-oceanic lightning and the observed enhancement in GLD360-reported peak currents over the ocean (refer to Section 1.4). We are interested in the ground wave portion of the lightning sferic, which, unlike the sky waves, propagates directly from the source lightning to the receiver. The ground wave carries the direct signature of the radiated fields and could be used to infer properties of the lightning current profiles, without the complications of modeling ionospheric interactions. Thus when reasonably close, our LF receiver could capture radio emissions from the lightning source and provides a higher resolution waveform than the one captured using VLF components traveling in the Earth-ionosphere waveguide (in terms of frequency, time, and direct lightning-to-receiver propagation), which is used by long range VLF networks such as GLD360.

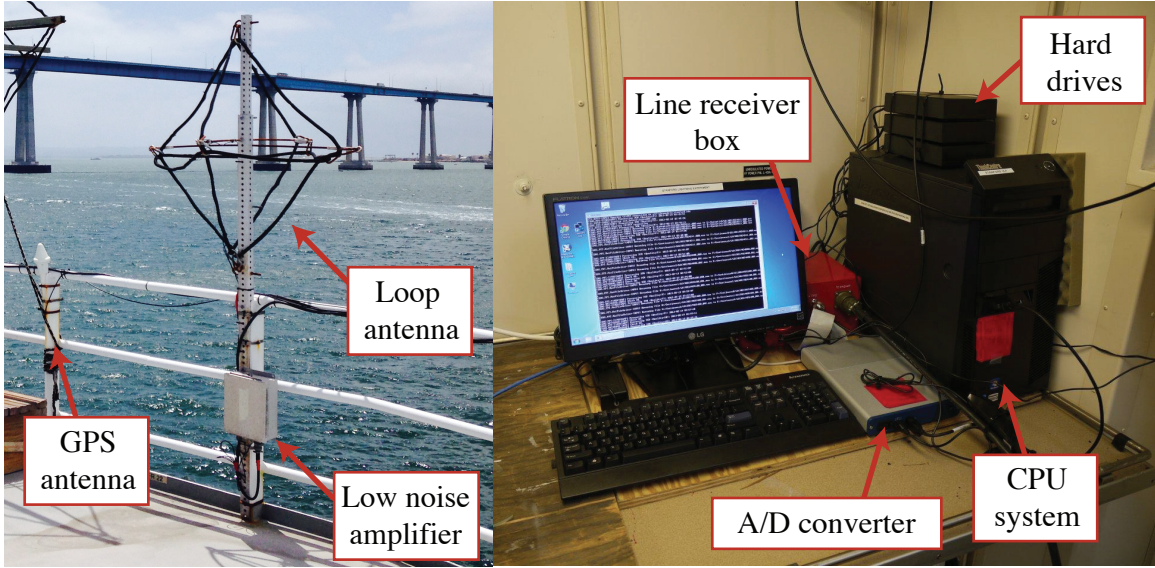


Figure 2.7: Pictures of the outdoor (left) and indoor (right) receiver system aboard the Ronald W. Brown research vessel.

We conduct a novel lightning experiment in which we install the AWESOME LF system, discussed in Section 2.2.1, aboard the National Oceanic Atmospheric Agency (NOAA) Ronald W. Brown research vessel to detect impulsive radio emissions from deep-oceanic discharges at short distances. Figure 2.7 shows pictures of the outdoor setup, which consists of a three wire-loop antennas (57 cm x 57 cm), including the horizontal loop which is useful to account for the sway of the ship, and the indoor computer system, line receiver box, and A/D converter. The experiment started in February, 2013 in San Juan, Puerto Rico, and ended in May, 2014 in San Francisco, California, as shown in the map of the voyage in Figure 2.6. One second of LF data was transmitted daily back to Stanford University to help verify data quality. Throughout the experiment, the system locally stored permanent broadband LF data around tens of thousands of GLD-reported return strokes within a 750 kilometer radius from the ship, successfully collecting deep-oceanic waveforms corresponding to CGs with large GLD-reported peak currents.

We develop two statistical tools that we use in this dissertation to analyze the collected LF waveforms. The first is an automated feature extraction technique that goes



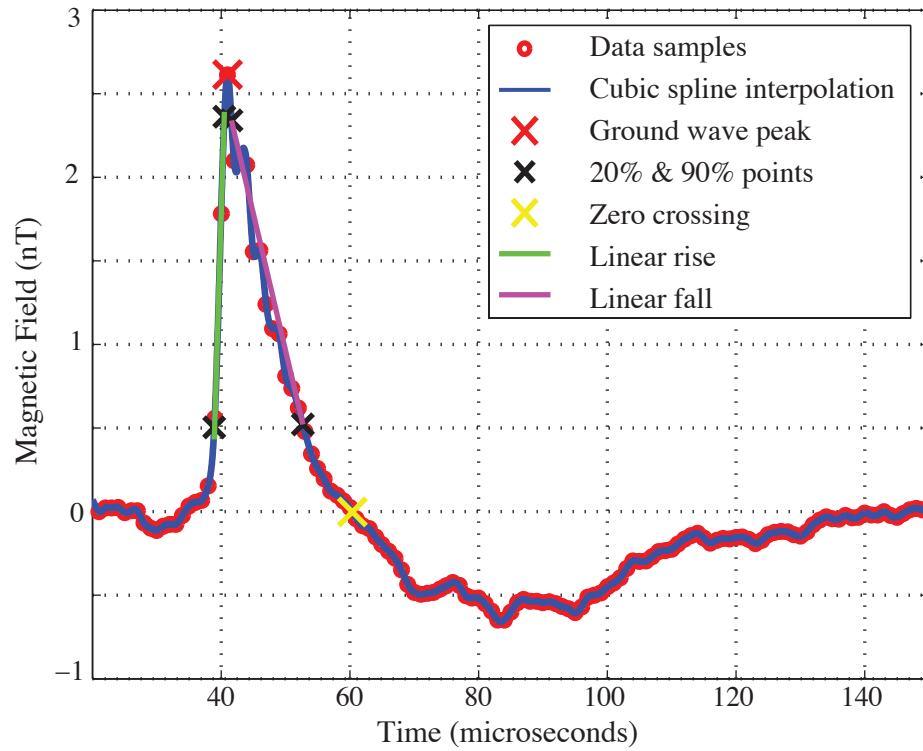


Figure 2.8: Feature extraction tool applied to a waveform corresponding to a GLD360-reported stroke that occurred on 17-Jul-2013 08:15:05 UT at (lat = 33.49, lon = -76.99), with -19 kA reported peak current, 285 km from the location of the ship.



through each collected LF time series, uses cubic spline interpolation to up-sample the data to 30 MHz, and extracts features such as rise-time, linear rise slope, waveform peak, fall-time, slope, and peak to zero crossing time. We apply the feature extraction technique to an oceanic  $-CG$ , 284.5 km from the ship, on July 17, 2013 at 08:15:05 UTC, and display the waveform in Figure 2.8. This technique enables us to study features in the individual sferics and to develop a better understanding of the impact of various physical lightning properties on the ground wave features. The second tool is a waveform aggregation technique which is useful to capture average effects and to compare the average ground waves of different groups of lightning events. The technique goes through a group of lightning sferics, aligns them in time, computes their average waveform, and compares it to the average waveform of other groups. In the next chapter, we use both techniques to analyze thousands of waveforms, to compare oceanic and land waveforms, and to study the impact of ground conductivity, return stroke speed, and peak currents on the experimental LF waveforms.

## Chapter 3

# Ground Wave Modeling and Analysis

In this dissertation, we aim to investigate physical differences and similarities between ocean and land lightning, using remote sensing of the ground wave portion of the lightning waveform. As presented in Chapter 2, we conducted a novel ship-borne experiment and collected tens of thousands of Low Frequency waveforms radiated by CG return strokes. In this chapter, we construct a computationally efficient electromagnetic radiation model to simulate lightning ground waves, propagating over lossy and curved ground. We validate our model against two similar models. The first consists of analytical approximations of ground wave propagation from [Cooray \[1987\]](#). The second is a finite difference time domain (FDTD) numerical model introduced by [Marshall \[2012\]](#). We then use the model to develop a better understanding of the effects of various current properties on the amplitude and shape of LF waveforms and to compare land and ocean lightning.

### 3.1 Return Stroke Modeling

In order to model the electromagnetic waves that are radiated by lightning and observed at a known distance we first have to model the currents in a typically upward CG return stroke. This step involves specifying the spatial and temporal current

profile  $i(t, z)$  along the channel, where  $t$  is time and  $z$  is altitude. Generally return stroke models fall under one of four categories: gas dynamic models, electromagnetic models, distributed-circuit models, and engineering models, which are presented in detail in [Rakov and Uman \[2007, p.394\]](#). Gas dynamic models solve the gas dynamic equations, involving the conservation of energy, momentum, and mass and are mainly concerned with the evolution of a short segment of the lightning channel and its corresponding shock wave. Electromagnetic models involve solving Maxwell's equations to find the current along the lightning channel and the surrounding electromagnetic fields, based on a thin-wire antenna approximation to the lightning channel. Distributed circuit models model the lightning discharge as a transient process along a transmission line with a characteristic resistance, inductance, and capacitance. We focus on the fourth category of models, namely on engineering models, which have a small number of adjustable parameters that are enough to specify  $i(z, t)$ , such as peak current, return stroke speed, and channel-base current rise-time.

Engineering models relate the temporal current profile at altitude  $z$  to the current profile at the base of the lightning channel  $i(z = 0, t)$  as shown in Equation 3.1, where  $u(t)$  is the unit step function,  $A(z)$  is the attenuation of the amplitude of the current with altitude, and  $v_{rs}$  is the return stroke speed.

$$i(z, t) = u(t - z/v_{rs}) A(z) i(z = 0, t - z/v_{rs}) \quad (3.1)$$

Figure 3.1 illustrates the way engineering models specify the current profile  $i(z, t)$ . The shape of the channel base current  $i(z = 0, t)$  is restricted to a family of curves with linear rise and exponential fall, which are specified using three parameters, a peak current value  $I_{\text{peak}}$ , at time  $t_{\text{rise}}$ , and exponential fall time constant  $\tau_{\text{fall}}$ , as shown in Equation 3.2.

$$i(z = 0, t) = \begin{cases} 0 & , t < 0 \\ (I_{\text{peak}}/t_{\text{rise}}) t & , 0 \leq t \leq t_{\text{rise}} \\ I_{\text{peak}} e^{-\left(\frac{t - t_{\text{rise}}}{\tau_{\text{fall}}}\right)^2} & , t > t_{\text{rise}} \end{cases} \quad (3.2)$$

In this dissertation, we use two engineering models, the modified transmission line

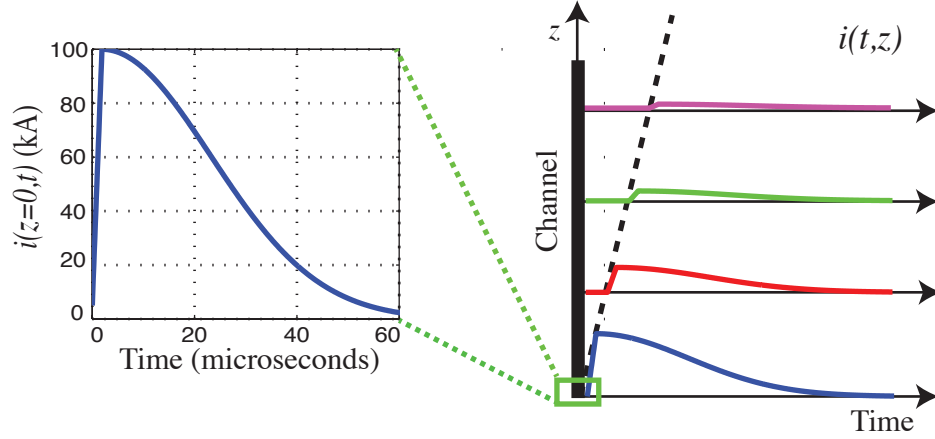


Figure 3.1: Illustration of return stroke modeling using engineering models. (left) The channel-base current  $i(z = 0, t)$  is selected following a linear rise to the peak and an exponential fall. (right) The current pulse decays in magnitude as it travels upward at the return stroke speed.

model with linear current decay with height (MTLL) [Rakov and Dulzon, 1987] and the modified transmission line model with exponential decay with height (MTLE) [Nucci et al., 1988], which only differ in the way they model the current attenuation with altitude,  $A(z)$ . The two attenuation functions are shown in Table 3.1, where  $L$  is the MTLL length of the lightning channel and  $\lambda$  is the MTLE exponential decay constant with altitude. The decay with altitude of lightning currents is in part due to the neutralization of charge that is deposited along the channel by the stepped leader and in part due to energy losses, mostly in the form of heat. The two models thus require a total of five parameters to fully specify  $i(z, t)$ :  $I_{\text{peak}}$ ,  $t_{\text{rise}}$ ,  $\tau_{\text{fall}}$ ,  $v_{\text{rs}}$ , and either  $L$  (MTLL) or  $\lambda$  (MTLE).

Model	$A(z)$
MTLL	$1 - z/L$
MTLE	$e^{-z/\lambda}$

Table 3.1: Attenuation function  $A(z)$  with altitude for MTLL and MTLE.

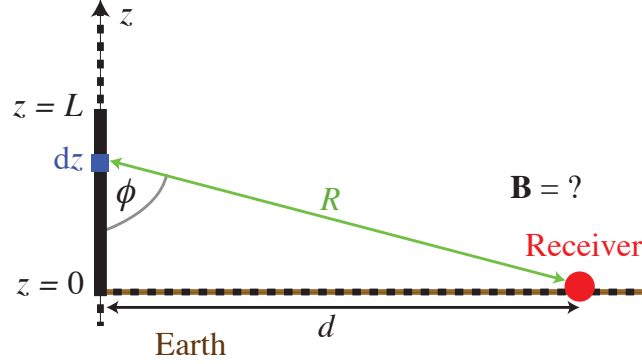


Figure 3.2: Illustration of the geometry of the planar configuration.

The MTLL and MTLE model have been validated on the ground using field experiments where the channel-base current  $i(z = 0, t)$  could be measured directly, which is the case for triggered lightning and lightning that strikes tall towers equipped with current sensors. Above the ground, validation has come from optical data to measure the corresponding return stroke speed  $v_{rs}$ , and comparing the measured fields with the model-derived fields. Other studies use a typical channel-base current profile and a typical return stroke speed to compare the model-derived fields with typical measured waveforms. The various validation studies found that the relationship between peak current  $I_{\text{peak}}$  and the peak fields is well predicted by both of the two models [Rakov and Uman, 2007, p.407-410].

## 3.2 Analytical Model

The bulk of the sferic energy is radiated by the bottom part the lightning channel, where the return stroke currents are most intense. The bottom portion of the channel is typically modeled as a vertical line due to the boundary conditions imposed by the conductive ground. Figure 3.2 illustrates the geometric factors that are used in computing the electromagnetic fields resulting from a vertical channel of length  $L$  over a flat conducting Earth represented by a perfectly conducting plane. The derived fields assume a planar configuration (flat Earth) and a perfectly conducting plane,

which are good approximations at short distances ( $< 500$  km) and up to frequencies for which the ground could be considered as a good conductor. The method of image currents is applied to replace the conducting ground with another conducting channel segment between  $z = 0$  and  $z = -L$ , with current magnitudes symmetric to the real currents above the ground, but traveling in the same upward direction. At a distance  $d$ , we compute the magnetic flux density  $\mathbf{B}$  on the ground using the magnetic potential vector  $\mathbf{A}$  as given by Equation 3.3, after solving for the components of the vector potential in a cylindrical coordinate system shown in Equation 3.4–3.5.

$$\mathbf{B} = \nabla \times \mathbf{A} \quad (3.3)$$

$$A_d = A_\theta = 0 \quad (3.4)$$

$$A_z(R, t) = 2 \frac{\mu_0 \mu_r}{4\pi} \int_{z=0}^{z=L} \frac{i(z, t')}{R} dz \quad (3.5)$$

where  $R = \sqrt{d^2 + z^2}$  is the propagation distance, the factor of 2 comes from the image current,  $\mu_0 = 4\pi \times 10^{-7}$  H/m is the magnetic permeability of free space,  $\mu_r$  is the relative permeability, and  $t'$  is the retarded time given by  $t' = t - R/c$ . After the appropriate mathematical computations using Equation 3.4–3.5, we find that the corresponding magnetic flux density on the ground is

$$B_\theta(R, t) = \frac{\mu_0 \mu_r}{2\pi} \int_{z=0}^{z=L} \frac{\sin \phi}{R^2} i(z, t') + \frac{\sin \phi}{cR} \frac{\partial i(z, t')}{\partial t'} dz \quad (3.6)$$

where  $c = 2.998 \times 10^8$  m/s is the speed of light in free space. For  $d \gg L$  the resulting azimuthal component of the magnetic field radiation is approximated by

$$B_\theta(R, t) = \frac{\mu_0 \mu_r S(t)}{2\pi R^2} + \frac{\mu_0 \mu_r}{2\pi cR} \frac{\partial S(t)}{\partial t} \quad (3.7)$$

The first term is the induction term and the second term is the radiation term. The induction term, which is proportional to  $1/R^2$ , attenuates faster with distance than the radiation term and tends to be negligible for distances  $R \gg c/2\pi f$ , which

is certainly the case for our application.  $S(t)$  is the current moment, which is the effective source of VLF/LF radiation, found by integrating the current along the channel as shown in Equation 3.8.

$$S(t) = \int_0^L i(z, t') dz \quad (3.8)$$

The finite ground conductivity attenuates and disperses the ground waveform. We model this effect using correction factors derived by [Norton \[1936\]](#) and validated by [Cooray \[1987\]](#) against collected lightning waveforms. The analytical model assumes a planar configuration (i.e., does not account for the effects of Earth curvature), limiting its accuracy to propagation distances shorter than 500 km [[Cooray, 1987](#)]. The correction technique is applied in the Fourier space due to the frequency dependence of the electrical property of the ground. The corrected radiation field  $dB_\theta^c(R, j\omega)$  resulting from a channel segment  $dz$ , at an elevation  $z$ , propagating to a distance  $d$  over a perfect conductor is given by Equation 3.9, where  $\omega$  is the angular frequency,  $j^2 = -1$ , and  $I(j\omega, z)$  is the Fourier transform of  $i(t, z)$  given by Equation 3.10.

$$dB_\theta^c(R, j\omega) = \frac{j\omega\mu_0\mu_r}{2\pi c\sqrt{d^2 + z^2}} I(j\omega, z) e^{(j\omega t - j\omega \frac{\sqrt{d^2 + z^2}}{c})} dz \quad (3.9)$$

$$I(j\omega, z) = \int_{-\infty}^{+\infty} i(t, z) e^{-j\omega t} dt \quad (3.10)$$

The frequency domain radiation field  $B_\theta^c(R, j\omega)$  is computed by integrating over all the radiation elements along the length of the channel  $L$  after accounting for attenuation due to finite ground conductivity  $\sigma$ , as shown in Equation 3.11, where  $G(j\omega, z, d, \sigma)$  is the attenuation function derived by [Norton \[1936\]](#) given by Equation 3.12. The time domain radiation fields  $B_\theta^c(R, t)$  are computed using the inverse Fourier transformation given by Equation 3.13 and expressions for the components of  $G(j\omega, z, d, \sigma)$  are given by Equations 3.14–3.16, where  $\text{erfc}$  stands for the complementary error function.

$$B_\theta^c(R, j\omega) = \int_{z=0}^L G(j\omega, z, d, \sigma) dB_\theta^c(R, j\omega) \quad (3.11)$$

$$G(j\omega, z, d, \sigma) = 1 + j\sqrt{\pi\alpha}e^{-\alpha} \operatorname{erfc}(-j\sqrt{\alpha}) \quad (3.12)$$

$$B_{\theta}^c(R, t) = \int_{-\infty}^{+\infty} B_{\theta}^c(R, j\omega) e^{j\omega t} d\omega \quad (3.13)$$

$$\alpha = \frac{2j\omega\sqrt{d^2 + z^2}\beta^2}{c(1 - \gamma)^2} \quad (3.14)$$

$$\beta^2 = \frac{j\omega}{j\omega\epsilon_r\epsilon_0 - \mu_r\mu_0\sigma c^2} \quad (3.15)$$

$$\gamma = \frac{z/d - \beta}{z/d + \beta} \quad (3.16)$$

The analytical model provides an efficient method to approximate the ground wave fields at close distances from the lightning source but does not model ground curvature which could have a large impact on LF components of the ground wave (discussed in Section 3.3.3). In the following section, we present the ground wave Full-Wave Method (gwFWM) which we develop to efficiently simulate ground waves and to visualize the impact of source parameters (such as a faster return stroke speed or a shorter current rise-time) on the shape of the ground wave while accounting for radiation and propagation physics including curvature.

## 3.3 Ground Wave Full-Wave Method (gwFWM)

### 3.3.1 Model Description

We build a ground wave model on top of the finite element frequency domain Full-Wave Method (FWM) to radiate and propagate the ground wave fields (gwFWM) over a lossy and curved Earth. FWM was developed by *Lehtinen and Inan* [2008, 2009] and has been validated in many VLF ionospheric applications [*Lehtinen and Inan*, 2008, 2009; *Cohen et al.*, 2011, 2012; *Graf et al.*, 2013].

Figure 3.3 illustrates the stratified medium setup in which we find the fields created



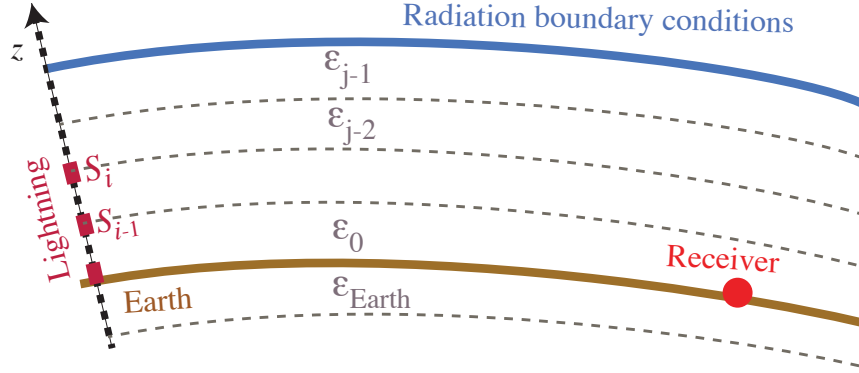


Figure 3.3: gwFWM model setup. The lightning sources are modeled at the boundaries of the stratified region, the ionosphere is replaced with radiation boundary conditions, and each medium can have an arbitrary conductivity and permittivity.

by an arbitrary combination of harmonically varying sources. The current sources are modeled at the boundaries between layers, using boundary conditions. The medium of the  $j^{th}$  layer has arbitrary electromagnetic properties, specified by a complex permittivity  $\epsilon_j = \epsilon' - j\sigma_j/\omega$ , where  $\epsilon'$  is the permittivity and  $\sigma_j$  the conductivity (for a lossy medium). The ionosphere is ignored and replaced with radiation boundary conditions, as we are only interested in modeling the ground wave portion of the spheric (i.e., does not include ionospheric reflections). Snell's law is applied in the plane-stratified medium to calculate the fields for each horizontal wave vector component in the Fourier space over horizontal coordinates. The technique recursively computes the reflection coefficients and mode amplitudes and is stable against the numerical “swamping,” which is inherent in other techniques [Budden, 1985, p.574-576]. Lehtinen et al. [2012] described a general curvilinear stratified system to include the effects of curvature into the planar method and validated the results in a cylindrically coordinate system.

Figure 3.4 illustrates a sample gwFWM setup. First, the current profile  $i(t, z)$  is specified using return stroke engineering models. The channel is then divided into  $N$  segments, four in this example, and the current moment  $S_i(t)$  is computed in each segment ( $i = 0, 1, 2, 3$ ). FWM is used to calculate the radiation and propagation

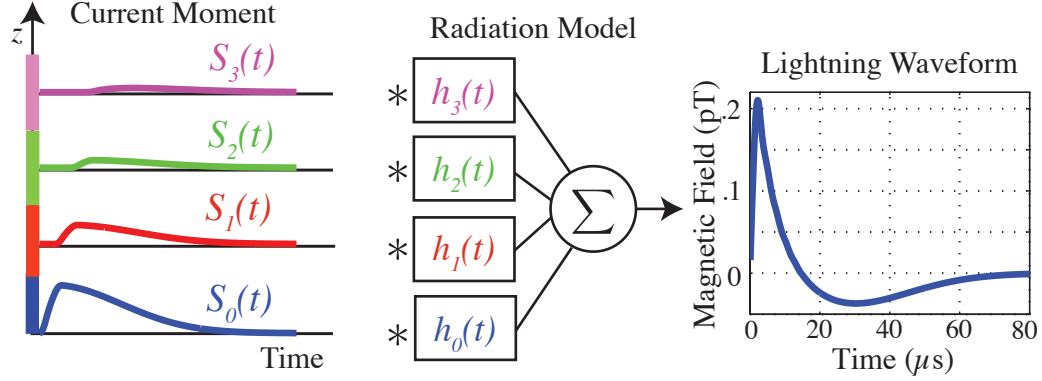


Figure 3.4: gwFWM modeling flow. (left) The channel is divided into segments. (middle) The resulting current moments are convolved with the impulse responses. (right) The resulting ground wave is found.

frequency response  $H_i(j\omega, z, d, \sigma, R_{\text{curv}})$  of the  $i^{\text{th}}$  segment.  $H_i(j\omega, z, d, \sigma, R_{\text{curv}})$  models the radiation of electromagnetic waves launched by currents in the  $i^{\text{th}}$  segment and their propagation over lossy and curved ground to the receiver. The frequency response is computed up to  $f_{\text{max}} = 500$  kHz bandwidth, which is consistent with the bandwidth of the LF receiver system. The frequency spacing is  $\Delta f = 10$  kHz which leads to a time domain spacing of  $1/\Delta f = 100 \mu\text{s}$ , larger than the typical duration of the ground wave and preventing time domain aliasing. The impulse response is found using the inverse Fourier transform operation  $h_i(t, z_i, d, \sigma, R_{\text{curv}}) = \text{F}^{-1} H_i(j\omega, z_i, d, \sigma, R_{\text{curv}})$ , where  $\text{F}^{-1}$  is the inverse Fourier transform operator. The input current moments are convolved with the corresponding impulse response, summed up, and the resulting waveform is generated. We compute all the impulse responses  $h_i(t, z_i, d, \sigma, R_{\text{curv}})$  only once and construct the forward model. The model could then be used efficiently (in parallel) to evaluate ground waves for different current profiles (but same propagation properties  $d, \sigma$ , and  $R_{\text{curv}}$ ), without recomputing the channel response.

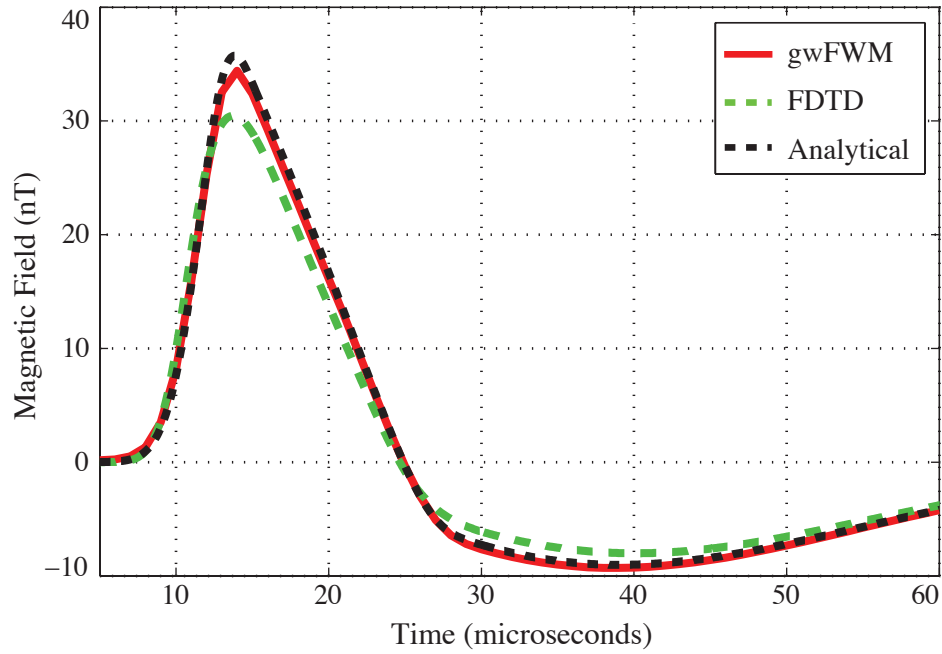


Figure 3.5: Simulated ground waves using the gwFWM (in solid red), the analytical approximations (in solid black), and the FDTD EMP (in dashed blue) for the same input parameters: MTLL model,  $2\ \mu\text{s}$  rise-time,  $30\ \mu\text{s}$  fall-time, 100 kA peak current, 3 km source height,  $2c/3$  return stroke speed, flat 2 mS/m Earth, and frequency content up to 500 kHz.

	gwFWM	FDTD	Analytical
Ground conductivity	Yes	Yes	Yes
Curvature	Yes	Yes	No
Horizontal segmentation	No	Yes	No
Runtime	5–10 sec	2–3 hr	5–10 sec

Table 3.2: Comparison of the gwFWM, FDTD EMP, and the analytical model.

### 3.3.2 Validation and Comparison

Figure 3.5 shows three ground waves simulated at 300 km distances, the first (in solid red) using the gwFWM, the second (in dashed blue) using the FDTD EMP from *Marshall* [2012], and the third (in dashed black) using the analytical model. All are computed using the same input parameters: (i) MTLL return stroke model with 2  $\mu$ s current rise-time, 30  $\mu$ s fall-time, 100 kA peak current, 3 km source height, and  $2c/3$  return stroke speed; (ii) wave propagation over flat Earth, with finite ground conductivity of 2 mS/m, and (iii) frequency content up to 500 kHz. We limit the comparison among the three models to a flat Earth due to the planar limitations of the analytical model. The gwFWM ground wave is consistent with the output of the two other models. We repeat the comparison for different current profiles  $i(t, z)$  and find similar results that validate our gwFWM model.

Table 3.2 compares the advantages and shortcomings of the lightning models. The three models account for a finite ground conductivity, though only gwFWM and FDTD EMP account for ground curvature. The FDTD model (time domain) is the only model among the three that could simulate non-uniform horizontal conductivity profiles. Thus FDTD EMP could model the ocean-land boundary which is useful to investigate electromagnetic reflections off of the ocean-land transition and to compare the efficiency of source excitation over the ocean and over land (due to the stronger image current over higher conductivity ground). We use the FDTD EMP to simulate the same lightning stroke twice, once over the ocean with 300 km of ocean propagation ( $\sigma(d) = 5$  S/m for all  $d$ ) and a second time over land with only 0.25 km of land propagation and the remaining 299.75 km of propagation over the ocean ( $\sigma(d) =$

2 mS/m for  $d \leq 0.25$  km and  $\sigma(d) = 5$  S/m for  $d > 0.25$  km). We compare the simulated ground waves and find them to be visually identical, suggesting that source excitation over land and ocean does not contribute to the observed electromagnetic enhancement. We also probe the fields in the second run at the ocean-land boundary and do not find any reflections from the abrupt transition in ground conductivity at  $d = 0.25$  km. The results suggest that the sharp GLD360-reported enhancement along the coastlines is not due to the more efficient electromagnetic excitation or to electromagnetic reflections, and is either due to differences at the source lightning and/or due to network inaccuracies.

However, FDTD is computationally expensive, taking several hours to evaluate each run. Unlike FDTD EMP, gwFWM (frequency domain) is limited to uniform horizontal conductivity profiles but is computational fast, taking 5 to 20 seconds to evaluate each run. This efficiency offers many practical advantages and enables us to conduct sensitivity analyses (Section 3.4). We use the gwFWM to study the impact of Earth curvature on ground wave propagation (Section 3.3.3), the effects of different lightning parameters on the simulated waveform and to compare our LF data to the simulated waveforms.

### 3.3.3 Earth Curvature

We use gwFWM to study the impact of Earth curvature on VLF and LF ground wave propagation. Figure 3.6 shows the attenuation due to Earth curvature ( $R_E = 6,371$  km) as a function of propagation distance, parameterized by frequency. For each distance  $d_i$  and frequency  $f_j$ , we compute the ratio of the magnitude response of a curved Earth to the response of a flat Earth, and plot them on a dB scale. Ground curvature acts as a low-pass filter with increasing attenuation for larger frequency components and for longer propagation paths. The attenuation of the 20 kHz component (in blue) goes from  $<1$  dB at 200 km to  $\sim 6$  dB at 1,000 km. Higher frequency components such as 200 kHz (in black) and 400 kHz (in cyan) attenuate faster, reaching 6 dB attenuation at 400 km and 500 km, respectively. We note that these results are specific to the ground wave portion of the sferic (they do not account for ionospheric

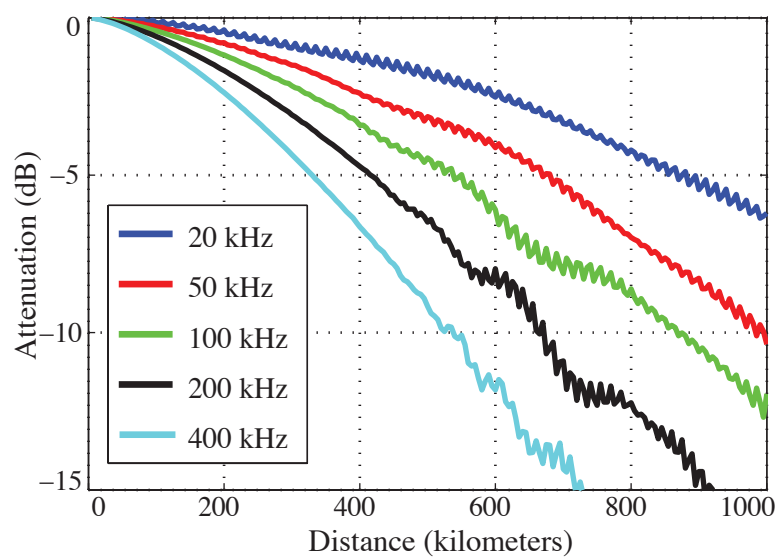


Figure 3.6: Attenuation due to Earth curvature (shown on a dB scale) as a function of propagation distance (in kilometers) for different frequency components: 20 kHz (in blue), 50 kHz (in red), 100 kHz (in green), 200 kHz (in black), and 400 kHz (in cyan).

reflections) and are important when comparing theoretical ground waves to data, as the attenuation due to curvature varies for different frequency components. We note that the oscillations in the results of Figure 3.6 are due to model approximations in the conversion from the wave vector domain to the spatial domain. The attenuation curves could also be used to find an empirical attenuation formula  $C(j\omega, d)$  that could be added to the analytical model to account for the impact of curvature, by adding  $C(j\omega, d)$  to Equation 3.11.

In the next section, we conduct a one-factor at a time sensitivity analysis where we hold all gwFWM inputs constant and vary one at a time to study its impact on the simulated waveform.

### 3.4 Sensitivity Analyses

In this section, we are interested in studying the relationship between features of  $i(t, z)$  and features of the radiated waveforms. For instance, it is clear from Equation 3.2 that the channel base MTLL and MTLE peak current  $I_{\text{peak}}$  linearly scales  $i(t, z)$  leading to a (linear) change in the magnitude of the output waveform. If all else is constant, two simulations with  $I_{\text{peak}}^1 = a I_{\text{peak}}^2$  will result in  $i^1(t, z) = a i^2(t, z)$  and  $B^1(t) = a B^2(t)$ . The two output waveforms are similar in shape but scaled in magnitude. Next, we explore the (nonlinear) effects of current rise-time, current fall-time, and return stroke speed on the simulated waveform.

#### 3.4.1 Current Rise-Time and Ground Conductivity

Figure 3.7 shows four simulated ground waves at 500 km from four gwFWM runs with the current rise-time on the ground varying between 1  $\mu\text{s}$  (in black) and 4  $\mu\text{s}$  (in green). The other model parameters are held constant (MTLE,  $\beta_{\text{decay}} = 1.08$  km, 30  $\mu\text{s}$  fall-time,  $c/2$  return stroke speed, 100 kA peak current,  $R_E = 6,371$  km,  $\sigma = 5$  S/m). The simulated ground waves have 20%-90% rise-time and peak values of 1.43  $\mu\text{s}$  and 26.66 nT (1  $\mu\text{s}$  current rise-time; in black), 1.77  $\mu\text{s}$  and 25.14 nT (2  $\mu\text{s}$  current rise-time; in blue), 2.29  $\mu\text{s}$  and 23.63 nT (3  $\mu\text{s}$  current rise-time; in red), 2.88

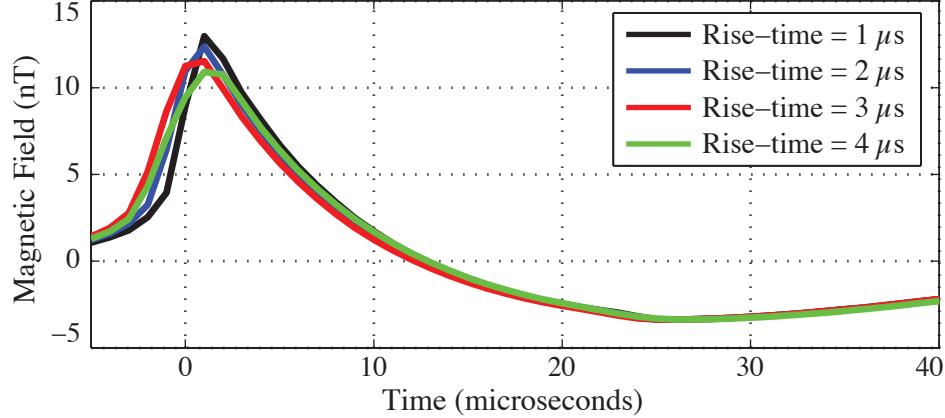


Figure 3.7: Ground wave for different values of channel-base current rise-time:  $1\ \mu\text{s}$  (in black),  $2\ \mu\text{s}$  (in blue),  $3\ \mu\text{s}$  (in red), and  $4\ \mu\text{s}$  (in green).

$\mu\text{s}$  and 22.29 nT ( $4\ \mu\text{s}$  current rise-time; in green). We measure a  $\sim 17\%$  drop in the LF waveform peak and a  $\sim 100\%$  increase in the 20%-90% rise-time as we vary the current rise-time from  $1\ \mu\text{s}$  to  $4\ \mu\text{s}$ . These results suggest that different channel-base current rise-times affect the shape of the corresponding waveform and that a change in current rise-time at the lightning source could be inferred by analyzing the 20%-90% rise-time, which should be observable in our LF data and could potentially provide more insight into the underlying attachment and the difference in current rise-time between oceanic and land CGs.

NLDN and various other LF geo-location networks use the LF peak of the ground wave to estimate peak currents [Cummins et al., 1998a; Nag et al., 2012] and calibrate their algorithms using subsequent return strokes from triggered lightning experiments [Nag et al., 2011; Turman et al., 2014]. We note that the  $\sim 17\%$  change in the waveform peak from shorter rise-times could potentially lead to errors in the estimation of peak currents for initial negative and positive CGs, which tend to have current rise-times much different (five times longer) than current rise-times of subsequent negative strokes [Berger et al., 1975].

We apply the feature extraction tool to 25,655  $-$ CGs collected throughout the voyage and use all the (20%-90% rise-time, lightning-to-receiver distance) pairs to



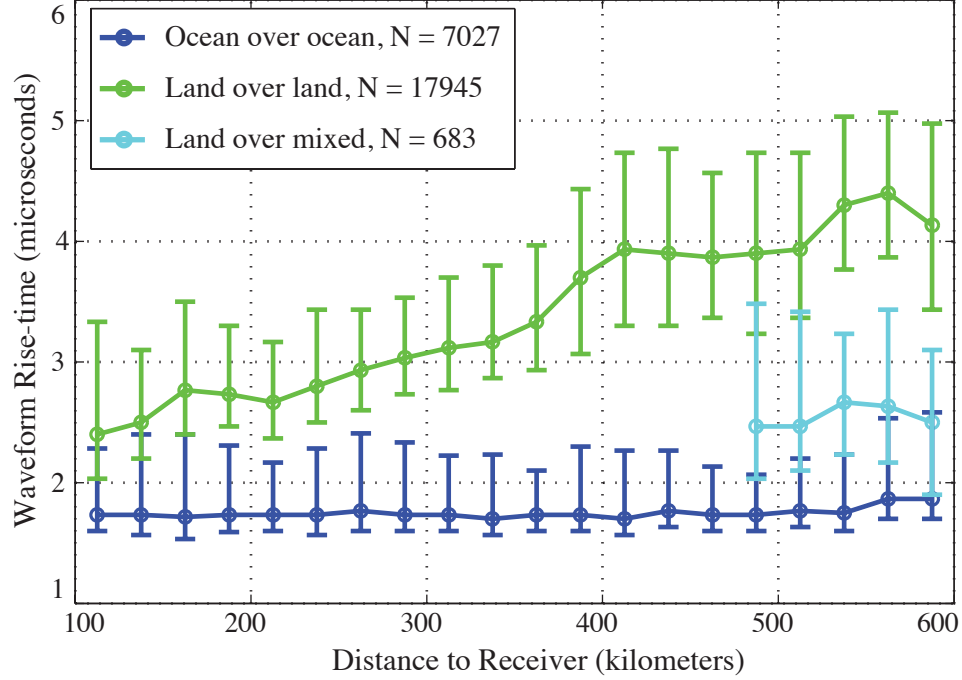


Figure 3.8: Measured 20%-90% waveform rise-time, with the corresponding 25<sup>th</sup> and 75<sup>th</sup> percentile error bars, as a function of propagation distance to the receiver for three categories: (in blue) 7,027 oceanic discharges with propagation path over the ocean, (in green) 17,945 land discharges with propagation path over land, and (in cyan) 683 land discharges with mixed propagation path.

generate Figure 3.8. We divide the strokes into three categories, the first (in blue) consists of 7,027 ocean –CGs that propagate over the ocean (>90% of the propagation path is over the ocean), the second (in green) consists of 17,945 land –CGs that propagate over land (>90% of the propagation path is over land; this occurs when the ship is docked), and the third (in cyan) consists of 683 land –CGs that propagate to the ship over a mixed path, >50% of which consists of ocean propagation (occurs when the ship is sailing close to land). For each category, we group the events in 25 km distance bins and compute their median 20%-90% rise-time. We plot the results and include the 25<sup>th</sup> and 75<sup>th</sup> percentile error bars.

Figure 3.8 shows that ground waves from ocean strokes with ocean propagation

(first category; in blue) have lower rise-times than waveforms from land strokes with land propagation (second category; in green). The median rise-times of the first category (in blue) are consistently lower than the 25<sup>th</sup> percentile of the second category (in green). The third category (in cyan), which consists of land strokes with mixed land-ocean paths, has 20%-90% rise-times that fall between the other two categories, with rise-times closer to ocean strokes with ocean propagation, and much lower than those of land strokes with land propagation. This shows that ground waves from land strokes have significantly lower rise-times when they partially propagate over the ocean (in cyan) than when they only propagate over land (in blue). Our results suggest that poor ground conductivities attenuate the higher frequency components, smoothing the rising slope of the waveform. We further investigate this trend using the gwFWM model and simulate two ground waves at 400 km for the same lightning stroke (MTLE,  $\beta_{\text{decay}} = 1.08$  km,  $1 \mu\text{s}$  rise-time,  $30 \mu\text{s}$  fall-time,  $c/3$  return stroke speed, 100 kA peak current,  $R_E = 6,371$  km), one propagation over ocean propagation ( $\sigma_{\text{ocean}} = 5$  S/m) and the other over land ( $\sigma_{\text{land}} = 2$  mS/m). We measure  $1.65 \mu\text{s}$  and  $3.64 \mu\text{s}$  rise-times at 400 km for ocean and land propagation, respectively, matching closely with our LF data.

Our experimental and theoretical observations suggest that lower ground conductivity attenuates the LF components and leads to a measurable change in the ground wave 20%-90% rise-time. Our results are in agreement with previous studies that investigate the impact of ground conductivity on LF radiation fields [*Bardo et al., 2004*; *Cummins et al., 2005*]. Lower ground conductivities attenuate the highest frequency components in the radiated fields and smoothen the shape of the rise of the waveform, leading to an increase in the measured ground wave rise-time.

We compare oceanic to land lightning by analyzing the median waveform rise-time at 500 km of the first category (ocean over ocean; in blue) to the rise-time of the third category (land over mixed; in cyan). We measure the waveform rise-time for oceanic (in blue) and land (in cyan) CGs to be  $1.8 \mu\text{s}$  and  $2.4 \mu\text{s}$ , respectively. As discussed earlier, longer rise-times in land CGs are in part due to the mixed propagation path (consisting of lower conductivity land) while ocean CGs only propagate over the ocean to the ship. However, the difference between the two waveform rise-times provides an

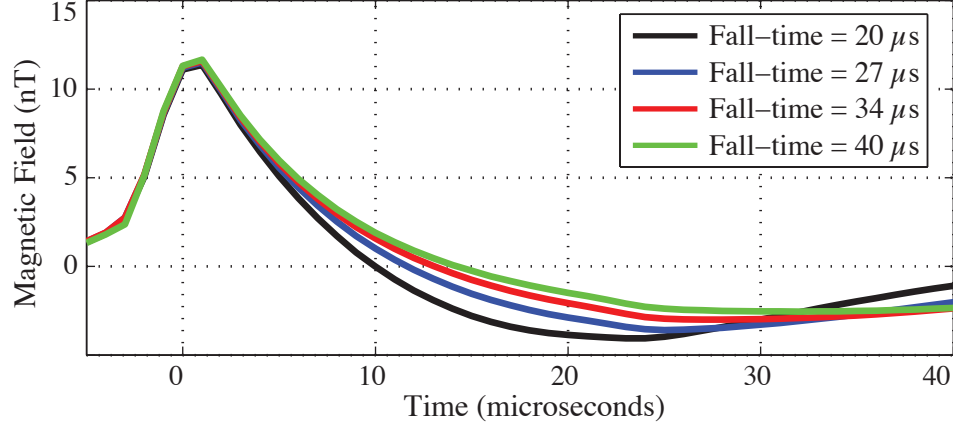


Figure 3.9: Ground wave for different values of channel-base current fall-time:  $20 \mu\text{s}$  (in black),  $27 \mu\text{s}$  (in blue),  $34 \mu\text{s}$  (in red), and  $40 \mu\text{s}$  (in green).

upper bound on the (potential) difference in oceanic and land channel-base current rise-times. We compare the 20%-90% waveform rise-times at 500 km to the simulated waveforms from Figure 3.7 (also at 500 km) and find that current rise-times of  $2 \mu\text{s}$  and  $3 \mu\text{s}$  lead to 20%-90% rise-times of  $1.8 \mu\text{s}$  and  $2.3 \mu\text{s}$  and to ground wave peaks of 25.1 nT and 23.6 nT, respectively. Our results suggest that land and ocean flashes have current rise-times within  $1 \mu\text{s}$  (this difference represents our upper bound that assumes all of the increase is from a difference in current rise-time and ignoring the partial propagation over land), resulting in  $<7\%$  change in the LF peak. We note that this change does not necessarily explain the stronger GLD360 peak current estimates over the ocean, which are measured in VLF.

### 3.4.2 Current Fall-Time

Figure 3.9 shows simulation results in which we vary the channel base current fall-time from  $20 \mu\text{s}$  (in black) to  $40 \mu\text{s}$  (in green) while holding other parameters constant (MTLE,  $\beta_{\text{decay}} = 1.08 \text{ km}$ ,  $3 \mu\text{s}$  rise-time,  $c/2$  return stroke speed, 100 kA peak current,  $R_E = 6,371 \text{ km}$ ,  $\sigma = 5 \text{ S/m}$ ), and plot the simulated ground waves at 500 km. The results suggest that current fall-time has no effect on the rise and on the

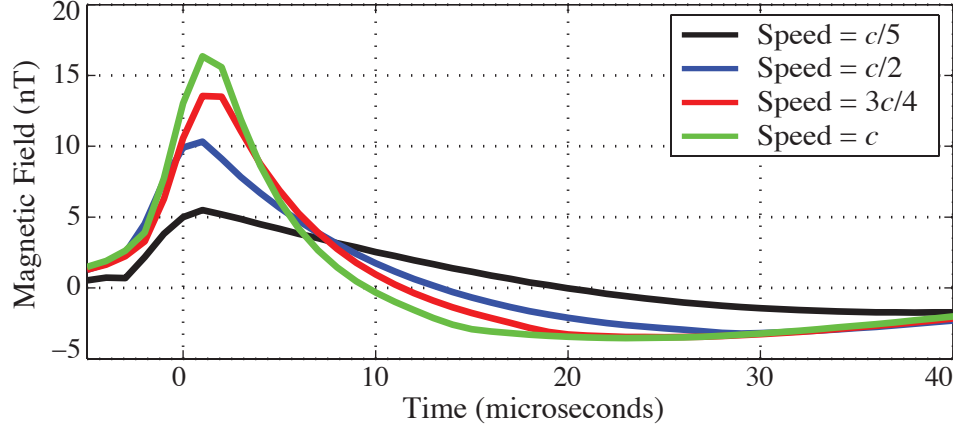


Figure 3.10: Ground wave for different return stroke speeds:  $c/5$  (in black),  $c/2$  (in blue),  $3c/4$  (in red), and  $c$  (in green)

peak of the ground wave, and mostly leads to a change in the shape of the tail of the waveform; longer current fall-times produce longer waveform fall-times. Thus a change in the fall-time of the channel base current (due to a difference in the attachment process or the ground conductivity) is unlikely to explain the observed GLD360 oceanic enhancement.

### 3.4.3 Return Stroke Speed

Figure 3.10 shows four runs where we hold all parameters constant ( $MTLE$ ,  $\beta_{\text{decay}} = 1.08$  km,  $3 \mu\text{s}$  rise-time,  $30 \mu\text{s}$  fall-time,  $100$  kA peak current,  $R_E = 6,371$  km,  $\sigma = 5$  S/m) and vary the return stroke speed along the lightning channel between  $c/5$  (in black) and  $c$  (in green), and plot the simulated waveforms at  $500$  km. Our results show that a change in the return stroke speed leads to narrower pulses with larger peaks. The theoretical model shows that a change in the lightning return stroke speed is measurable in the waveform and should be captured by our LF data.

Figure 3.11 consists of three panels that use both the feature extraction and waveform aggregation tools to compare four groups of lightning strokes, all between  $480$  km and  $520$  km distance to the ship. The first group (in blue) consists of  $1,162$

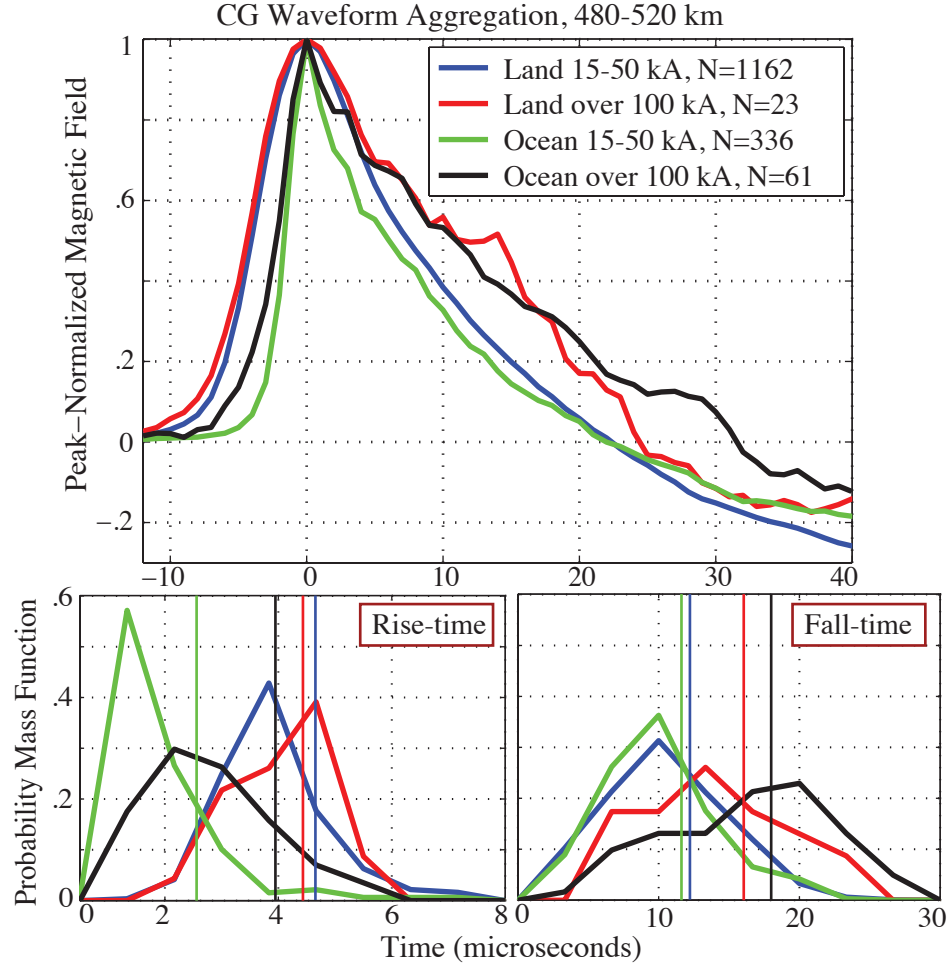


Figure 3.11: (Top panel) Average peak-normalized waveform for four lightning groups between 480 km and 520 km distance to the ship: (in blue) 1,162 land –CGs with GLD360 peak currents between 15 and 50 kA, (in red) 23 land –CGs >100 kA, (in green) 336 oceanic –CGs 15-50 kA, and (in black) 61 ocean –CGs >100 kA. (Bottom panels) Probability mass function of the waveform 20%-90% rise-times (left) and 90%-20% fall-times (right) of the four groups, with vertical lines corresponding to the medians.

land –CGs which have GLD360-reported peak currents between 15 and 50 kA, the second (in red) consists of 23 land –CGs with peak currents higher than 100 kA, the third (in green) consists of 336 ocean –CGs with peak currents between 15 and 50 kA, and the fourth group (in black) consists of 61 ocean –CGs stronger than 100 kA. The average waveforms of the four groups are displayed in the top panel, after normalizing the peaks to unity and setting them to time zero (as we are interested in the shape of the ground waves). The bottom panels show the probability mass function of the waveform 20%-90% rise-times (left) and 90%-20% fall-times (right) of the four groups. The bottom panels also include color-coded vertical lines which correspond to the medians of each distribution.

Both the top and the bottom left panels show that waveforms of land lightning tend to have longer rise-times and a smoother rising edge, in agreement with our results from Section 3.4.1, due to the lower land conductivity. The top and the bottom right panels show that weak land (in blue) and weak ocean (in green) events have a similar falling edge and 90%-20% fall-time distributions, which is also the case for strong land (in red) and strong ocean (in black) strokes. These results show that land and ocean events (of similar GLD360-reported intensity, discussed later) tend to have a similar ground wave fall shape, suggesting that, other than the rising edge that is affected by ground conductivity, ocean and land ground waves have similar shapes. As discussed in the return stroke sensitivity analysis (from Section 3.4), different return stroke speeds in the lightning channel lead to a measurable change in the shape of the ground wave, which is not the case in the collected waveforms, suggesting that it is unlikely for ocean and land discharges to have different return stroke speeds. In conclusion, our modeling and experimental analyses suggest that stronger oceanic lightning observations are unlikely to result from faster return stroke speed in oceanic CGs (due to a difference in the attachment process or the ground conductivity).

Interestingly, Figure 3.11 shows that weak ocean strokes (in green) have (on average) shorter rise-times and fall-times than strong ocean strokes (in black), and weak land strokes (in blue) have shorter rise-times and fall-times than strong land strokes (in red). This suggests that weaker GLD360-reported strokes, whether over land or

over the ocean, tend to have (on average) narrower ground waves than stronger ones. The consistent difference in the shapes of weak and strong lightning CGs suggests a difference in source parameters of weak and strong discharges, potentially resulting from different return stroke speeds.

# Chapter 4

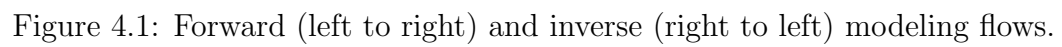
## Waveform to Current Parameters Modeling

In the previous chapter, we presented a physical model that simulates ground waveforms for a set of lightning inputs. The problem of predicting the output given a complete description of the physical system is known as the *forward* problem. We have used the *forward* model to quantify the sensitivity of the ground waveform to properties of the return stroke, such as the speed, and to properties of the propagation path, such as curvature and conductivity. In this chapter, we attempt to solve for the *inverse* problem, in which we use the lightning waveform to infer current parameters of the causative lightning source. Unlike the *forward* model, the *inverse* model is typically ill-posed and does not have a unique solution. Throughout this chapter, we present a novel statistical inverse model of LF lightning ground wave radiation and propagation. The *inverse* model estimates properties of the current inside the lightning channel using either simulation waveforms or collected experimental waveforms.

### 4.1 Introduction to Inverse Modeling

Figure 4.1 illustrates the flows of the *forward* and *inverse* models. In the *forward* direction, electromagnetic radiation and propagation models, such as gwFWM, start with a current profile along the channel to generate the lightning waveform at the





receiver. In our application, the *forward* gwFWM model was used to study the impact of various current parameters on the lightning waveform and to compare features in simulation and experimental waveforms (as discussed in Section 3.4). The aim of this chapter is to develop the *inverse* modeling flow in which the model starts with the ground waveform to estimate properties of causative lightning currents. In practice, this is important as it allows us to infer lightning currents inside the channel using radio remote sensing. The *inverse* model could then be applied to thousands of experimental waveforms collected throughout the experiment to compare source parameters of land and ocean lightning (Section 4.5).

In this dissertation, we develop the first statistical inverse model that estimates the current  $i(z, t)$  along the lightning channel using the LF waveform  $B(t)$ . The inverse relationship between  $B(t)$  and  $i(z, t)$  is nonlinear, ill-posed, and difficult to obtain analytically (discussed in Section 4.3). We use artificial neural networks, a class of nonlinear statistical models, to capture the complex dynamics of the physical *forward* model and to approximate the inverse mapping. In the next section, we present an overview of the structure, configuration, and training process of neural networks.

## 4.2 Artificial Neural Network

In this section, we describe Artificial Neural Networks (ANNs), a class of statistical learning methods which were developed separately in the fields of computer science, artificial intelligence, and statistics [Hastie et al., 2009, p.389]. The name *neural network* derives from the fact that these techniques were initially used in an attempt to model the activity of the human brain, which consists of layers of connected neurons [Hastie et al., 2009, p.394]. Since then, the term has evolved to include a slew of learning methods and nonlinear statistical techniques whose purpose is to estimate unknown functions that depend on a large number of inputs. ANNs could be used in both regression problems, in which the aim is to model how the value of one or more outputs vary when the inputs change, and in classification problems, in which the model is trained to predict the class (category) of the input observation. In this dissertation, we apply the widely used single hidden layer neural network, often

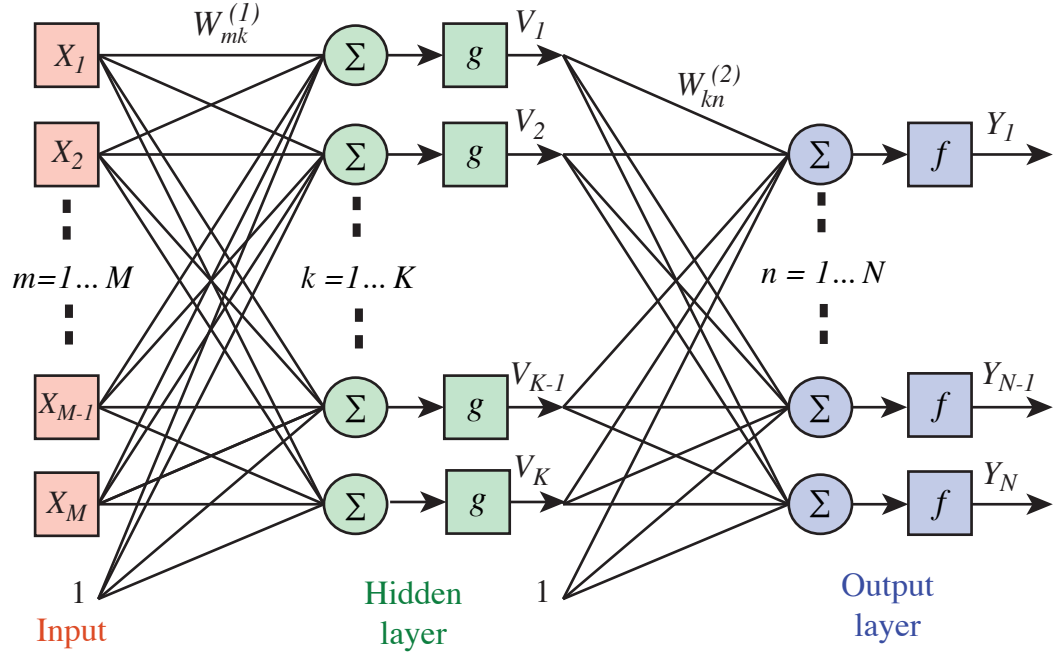


Figure 4.2: Structure of the single hidden layer artificial neural network.

known as the single layer perceptron, to calculate the best fit solution of the inverse of the gwFWM model. The resulting ANN could then be applied to collected lightning waveforms to infer values of the current along the channel (regression problem).

### 4.2.1 Model Structure

The hype surrounding neural networks makes them seem magical and useful in any application. As we show here, ANNs are simply nonlinear statistical methods that model target outputs as nonlinear functions of linear combinations of the inputs. In some ways they can be thought of as nonlinear interpolation schemes. Figure 4.2 illustrates the typical structure of the single hidden layer feed-forward network, consisting of a cascade of two regressions. The first is in the hidden layer (in green) and the second is in the output layer (in blue). The hidden layer operates on  $\mathbf{X} \in \mathbb{R}^M$ , the input vector to the neural network, and outputs the vector  $\mathbf{V} \in \mathbb{R}^K$ . The output layer then operates on  $\mathbf{V}$  to produce the output vector  $\mathbf{Y} \in \mathbb{R}^N$ . In the hidden layer,

each element  $V_k$  of  $\mathbf{V}$  is a function  $g$  of a linear combination of the elements of  $\mathbf{X}$ , as shown in Equation 4.1.

$$V_k = g\left(W_{0k}^{(1)} + \sum_{m=1}^M W_{mk}^{(1)} X_m\right), \quad k = 1, \dots, K \quad (4.1)$$

Similarly, the elements of  $\mathbf{Y}$  are a function  $f$  of linear combinations of the elements of  $\mathbf{V}$ , as shown in Equation 4.2. The additional weights  $W_{0k}^{(1)}$  and  $W_{0n}^{(2)}$  come from the bias units connected to the nodes (also known as neurons) of the two layers, adding the intercept terms which are represented using the constant ‘1’ in the diagram.

$$Y_n = f\left(W_{0n}^{(2)} + \sum_{k=1}^K W_{kn}^{(2)} V_k\right), \quad n = 1, \dots, N \quad (4.2)$$

The choice of the so-called activation functions  $f$  and  $g$  vary for different applications and between classification and regression problems. Typically, the activation function is the same across the nodes of the same layer. For instance, if all the activation functions in both layers are linear, the network collapses to a standard linear fitting (regression) model. The use of activation functions is inspired by biological systems whose neurons fire at a rate proportional to the action potential. The Heaviside unit step function could thus be used as a binary representation of whether the node is firing or not. However, it is computationally efficient to have a gradual transition between the on and off states. The smooth transition could be thought of as the average behavior of many binary nodes, allowing the network to reproduce the same performance with many fewer nodes. Also, the smoothness and differentiability of activation functions lead to a faster training process (discussed in Section 4.2.2).

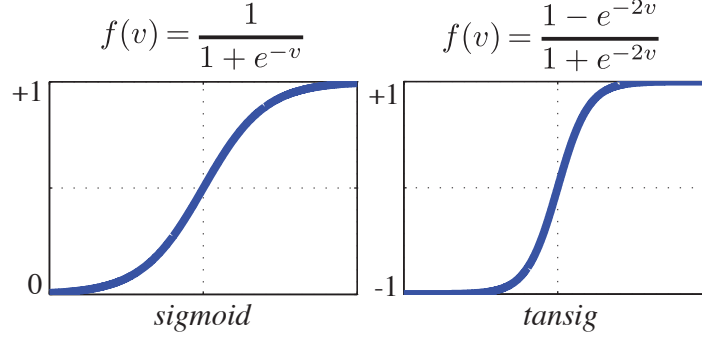


Figure 4.3: Expressions and shapes of two popular activation functions: the sigmoid (left) and the hyperbolic tangent sigmoid (right).

Typically in regression problems, similar to the one at hand, the nodes of the output layer have linear activation functions and hidden nodes have nonlinear activation functions. The neural network could thus be considered a nonlinear generalization of the standard linear model. Figure 4.3 shows the expressions and shapes of two commonly used activation functions: the sigmoid (left) and the hyperbolic tangent sigmoid (*tansig*, right) functions. The two functions start off flat and gradually approach an asymptote, modeling a 100% neuron firing rate. The sigmoid and the *tansig* have derivatives with analytical expressions which is advantageous in the training process. The main difference between them is that the sigmoid function normalizes to  $(0, 1)$  range and *tansig* normalizes to  $(-1, 1)$  range. In this dissertation, we apply the *tansig* function in the hidden layer and the linear function in the output layer ( $g = \textit{tansig}$  and  $f = \textit{linear}$ ). In this configuration, the ANN can accurately approximate any nonlinear function with a finite number of discontinuities, given enough nodes in the hidden layer.

The number of nodes in the hidden layer  $K$  defines the functionality and purpose of the hidden layer. In cases where  $K$  is smaller than the number of inputs  $M$ , the hidden layer acts like a Principal Component Analysis (PCA) decomposition, where  $\mathbf{V}$  is a lower-dimensional representation of  $\mathbf{X}$ . In our application,  $K$  is larger than  $M$  and  $\mathbf{V}$  could be thought of as a basis expansion of the original  $\mathbf{X}$ . The parameters of the basis expansion are learned from the data, which is typically more powerful than

standard techniques which use a pre-determined basis expansion approach. The basis expansion allows the system to capture nonlinear and complex relationships between the various elements of the original input  $\mathbf{X}$  and the output predictions  $\mathbf{Y}$ .

### 4.2.2 Training Process

Initially, the total dataset, consisting of  $N_{\text{total}}$  pairs of inputs and outputs  $(\mathbf{x}, \mathbf{y})$ , is divided into a training set ( $N_{\text{train}} = 80\% N_{\text{total}}$ ) and a testing (or validation) set ( $N_{\text{test}} = 20\% N_{\text{total}}$ ). The training set is used to train the network and the testing set to assess its performance. To prevent from biasing, once the testing set is used, the weights and properties of the network cannot be re-adjusted. These changes would lead to biased weights as the neural network tries to better fit the testing set. In that case, new samples (new testing set) would be required to re-assess the strength of the network.

The performance of the network depends on the number of layers, the size of each layer, the configuration of activation functions, and the strength of the connections or weights. In regression problems, the strength of the neural network is determined by the misfit between predictions and target outputs. System parameters are calculated using the training set in a way that minimizes the error function, which is computed using the mean squared error between predictions and target outputs as shown in Equation 4.3, where  $P(\cdot)$  represents the neural network operator,  $\mathbf{x}_i^{\text{train}}$  and  $\mathbf{y}_i^{\text{train}}$  are the input and output vectors of the  $i^{\text{th}}$  training sample, respectively.

$$e = \frac{1}{T} \sum_{i=1}^{N_{\text{train}}} \|\mathbf{y}_i^{\text{train}} - P(\mathbf{x}_i^{\text{train}})\|^2 \quad (4.3)$$

The number of layers, size of each layer, and activation functions are determined using the training set, following  $K$ -fold cross validation and sometimes using trial and error. The weights and biases are adjusted following an iterative algorithm that minimizes the mean squared error. Most optimization algorithms use the gradient of the error function with respect to the network weights and biases to find the optimal solution. The gradient is computed using back-propagation, which consists of

running the training data down the network then going backwards through the layers to calculate the gradient using chain rule derivations [[Hastie et al., 2009](#), p.295-397].

We present an overview of gradient descent, conjugate gradient, Newton/quasi-Newton, and Levenberg-Marquardt. The four techniques are based on back-propagation but differ in the way the weights and biases are updated following each iteration. Generally, the update is divided into two stages. In the first stage, the direction of the update is calculated and, in the second stage, the step size of the update is computed. We note that many versions of each technique have been developed however the purpose of this section is to familiarize the reader with the training process of ANNs.

In gradient descent, the gradient  $\mathbf{g}_k$  of the error function with respect to the weights is found using back-propagation. At the  $l^{th}$  iteration, the weights and biases are found using the update function given in Equation 4.4, where  $\mathbf{u}_l$  is the vector the weights and biases and  $\beta_l$  is the learning rate [[Cauchy, 1847](#)].

$$\mathbf{u}_{l+1} = \mathbf{u}_l - \beta_l \mathbf{g}_l \quad (4.4)$$

The direction of the update is in the steepest descent direction, which is opposite to the direction of the gradient. The magnitude of the step is determined by the learning rate  $\beta_l$ . If  $\beta_l$  is large, the step size is large, which could lead to instabilities and, if  $\beta_l$  is small, the step size is small, which leads to slow convergence. In order to work around this trade-off, certain gradient descent algorithms adjust the magnitude of the learning rate in a way that depends on the stability of the algorithm at the  $l^{th}$  iteration [[Hagan et al., 1996](#), p.12-18]. The training process resumes until the error function is smaller than a certain threshold, or the magnitude of the gradient is too small, or the total training time or total number of iteration exceeds pre-determined limit. The main drawback of gradient descent is that it leads to a zig-zag path to the optimal solution with slow convergence. Although the error function decreases the most along the steepest descent direction, it does not necessarily lead to the most efficient convergence. The other three techniques, namely conjugate gradient, Newton/quasi-Newton, and Levenberg-Marquardt, use numerical approaches that lead to 10 to 100

times faster convergence [*Hagan et al.*, 1996, ch.9] (refer to Chapter 9 of *Hagan et al.* [1996] for a review of basic numerical optimization).

In conjugate gradient, at the  $l^{th}$  iteration, the search direction  $\mathbf{q}_l$  is selected in a way that the component of the gradient parallel to the previous step, which has been made zero, remains zero. The direction in which this condition is satisfied is the so-called conjugate direction. The algorithm starts out by searching in the steepest descent direction  $\mathbf{q}_0 = -\mathbf{g}_0$ . Next, a line search is performed to determine the step size  $\beta_0$  along that direction and used in the update function given by Equation 4.5 [*Fletcher and Reeves*, 1964]. In following iterations, the conjugate direction  $\mathbf{q}_l$  is determined by combining the steepest descent direction  $-\mathbf{g}_l$  with the previous search direction  $\mathbf{q}_{l-1}$ , as shown in Equation 4.6. The scalar  $\alpha_l$  varies for different conjugate gradient algorithms. For instance, in the Fletcher-Reeves algorithm  $\alpha_l$  is given by Equation 4.7 and in the Polak-Ribire algorithm it is given by Equation 4.8 [*Fletcher and Reeves*, 1964; *Hagan et al.*, 1996].

$$\mathbf{u}_{l+1} = \mathbf{u}_l + \beta_l \mathbf{q}_l \quad (4.5)$$

$$\mathbf{q}_l = -\mathbf{g}_l + \alpha_l \mathbf{q}_{l-1} \quad (4.6)$$

$$\alpha_l = \frac{\mathbf{g}_l^T \mathbf{g}_l}{\mathbf{g}_{l-1}^T \mathbf{g}_{l-1}} \quad (4.7)$$

$$\alpha_l = \frac{\Delta \mathbf{g}_{l-1}^T \mathbf{g}_l}{\mathbf{g}_{l-1}^T \mathbf{g}_{l-1}} \quad (4.8)$$

The quasi-Newton method is similar to the Newton method. The two techniques try to achieve second order convergence, which leads to larger step sizes and faster convergence than gradient descent. The update function of the Newton method is shown in Equation 4.9, where  $H_l$  is the Hessian of the error function after the  $l^{th}$  iteration, as given component-wise by Equation 4.10. The computation of the Hessian is complex and computationally expensive. In the quasi-Newton technique, approximations of the Hessian are applied following the Broyden, Fletcher, Goldfarb, and



Shanno (BFGS) algorithm [[Broyden, 1970](#); [Fletcher, 1970](#); [Goldfarb, 1970](#); [Shanno, 1970](#)] (refer to [Dennis and Schnabel \[1983\]](#) for a description of the BFGS algorithm).

$$\mathbf{u}_{l+1} = \mathbf{u}_l - H_l^{-1} \mathbf{g}_l \quad (4.9)$$

$$(\mathbf{H}_l)_{i,j} = \frac{\partial^2 e}{\partial u_i \partial u_j} \quad (4.10)$$

The Levenberg-Marquardt (LM) method is a commonly used minimizer of the mean squared error function. In Levenberg-Marquardt, the Hessian  $H$  is approximated by  $J^T J$ , where  $J$  is the Jacobian computed using first derivatives of the errors with respect to the weights and biases. The gradient is calculated using  $\mathbf{g} = J^T \mathbf{e}$ , where  $\mathbf{e}$  is the vector of errors, with the update function given in Equation 4.11, where  $I$  is the identity matrix and  $\mu_l$  is a scalar. If the scalar  $\mu_l$  is large, LM behaves like gradient descent with small step sizes (slow) and, if  $\mu_l$  is small, the update collapses to Newton's method (fast and accurate close to minimum). In the training process, the value of  $\mu_l$  is decreased after each iteration that lowers the error function (LM shifts closer to Newton's method) and the value of  $\mu_l$  is increased if the step would increase the error function (LM shifts closer to gradient descent), which would insure that the error function is reduced after every iteration.

$$\mathbf{u}_{l+1} = \mathbf{u}_l - \left[ J^T J + \mu_l I \right]^{-1} J^T \mathbf{e} \quad (4.11)$$

In our implementation of ANNs, we use the mean squared error to measure the misfit between the predictions and the target outputs, and apply the Levenberg-Marquardt algorithm to train the weights and biases of the network.

### 4.3 gwFWM Inverse Modeling

We develop a statistical inverse model of gwFWM using neural networks. The inverse model is trained to estimate the current profile  $i(z, t)$  along the channel using

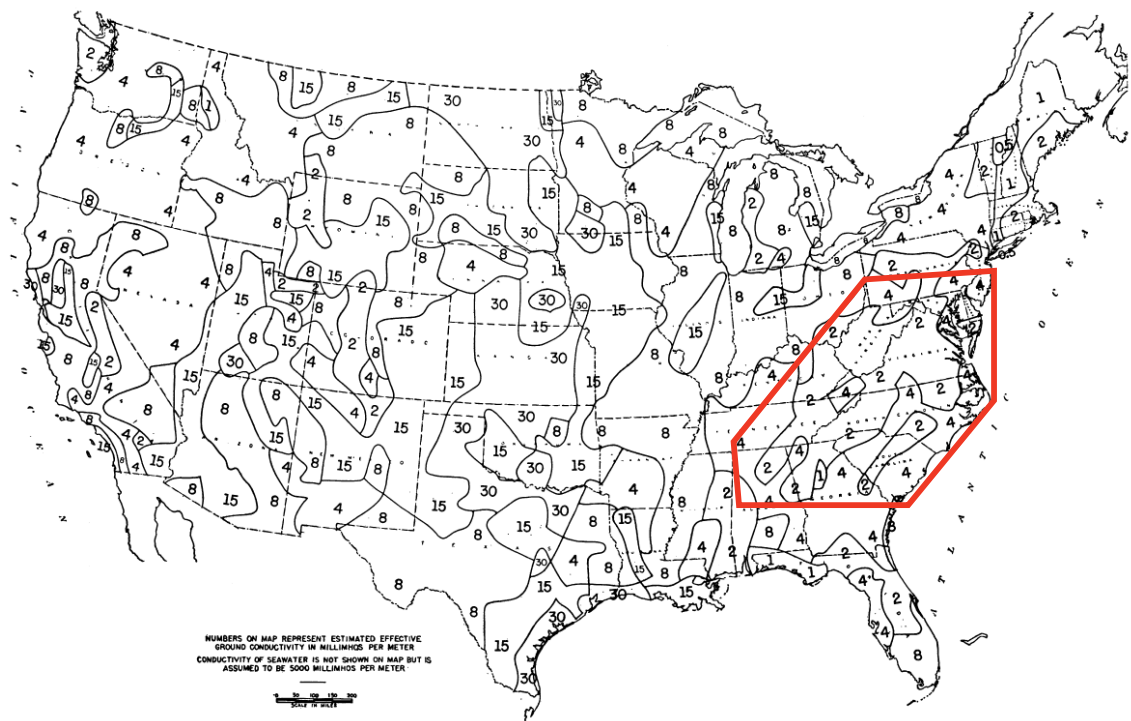


Figure 4.4: Estimated ground conductivity in the United States. Adapted from *Fine* [1953]. Lightning strikes over land occurred in the region highlighted by the red hexagon, propagating to the ship which was docked in Charleston, South Carolina.

the lightning waveform  $B(t)$ . The inverse relationship between  $B(t)$  and  $i(z, t)$  is complicated, nonlinear, nonconvex, and is not unique as many current profiles lead to the same waveform. Thus several simplifications are required to reduce the complexity of the inverse problem without affecting the underlying physics. We assume that (i) the distance  $d$  between the lightning source and the receiver is known, (ii) the ground conductivity along the propagation path is known and uniform, (iii) return stroke currents follow the MTLE model (discussed in Section 3.1) with a known decay constant with altitude. This limits the possible solutions of  $i(z, t)$  to a family of profiles that are fully specified using five parameters: the peak channel-base current  $I_{\text{peak}}$ , the channel-base current rise-time  $t_{\text{rise}}$ , the channel-base current fall time constant  $\tau_{\text{fall}}$ , the return stroke speed  $v_{\text{rs}}$ , and the exponential decay constant with altitude  $\lambda$ . In summary, we are looking to infer the MTLE current profile that lead to the observed lightning waveform, while assuming knowledge of the propagation distance  $d$ , of the ground conductivity  $\sigma$ , and of the MTLE decay constant  $\lambda$ .

In practice, the distance  $d$  could be accurately computed using the GPS location of the receiver system and the GLD360 location of lightning. The GLD360 geolocation accuracy is 2–5 km, much smaller than the range of propagation distances of interest (200 km and 600 km). This is equivalent to  $\sim 1\%$  error in  $d$ , which has little impact on the shape and amplitude of the ground wave. As for the second assumption, the conductivity over ocean and land is known and could be modeled with a uniform  $\sigma$ . Over the ocean, the conductivity of saltwater is  $\sim 5$  S/m. Over land, the conductivities are determined using conductivity maps similar to the one shown in Figure 4.4. These maps typically have poor spatial resolution of a few hundred kilometers. Our collected experimental data were captured along the eastern coast of the continental U.S., in the region that is highlighted with a red hexagon, where the ground conductivity is mostly uniform and around 2–4 mS/m. Therefore, uniform and known ground conductivities could be assumed without loss of information.

We train and test the inverse model using simulations of the forward gwFWM model (discussed in Section 3.3). Each forward run requires seven input parameters: propagation distance  $d$  (known), ground conductivity  $\sigma$  (known), and five MTLE return stroke parameters ( $\lambda$  is known, remaining inputs are selected at random, as

discussed later). As shown in Equation 4.12, the inputs are denoted by the vector  $\mathbf{C}$  and the resulting magnetic field time series by  $\mathbf{B}$ , with  $S$  corresponding to the gwFWM model operator. The elements of the time series are given by  $B_n = B(t + d/c + nT_s)$ , where the  $d/c$  term accounts for the propagation time delay,  $T_s$  is the sampling period, and  $n$  is an integer. For instance,  $B_{-10}$  corresponds to the tenth sample prior to the onset of the waveform.

$$\mathbf{C} = \begin{pmatrix} I_{\text{peak}} \\ t_{\text{rise}} \\ \tau_{\text{fall}} \\ v_{\text{rs}} \\ \lambda \\ d \\ \sigma \end{pmatrix} \quad \mathbf{B} = S(\mathbf{C}) = \begin{pmatrix} B_{-20} \\ B_{-19} \\ \dots \\ B_0 \\ B_1 \\ \dots \\ B_{98} \\ B_{99} \end{pmatrix} \quad (4.12)$$

We emphasize that the aim is to apply the inversion to gwFWM simulation waveforms and, more importantly, to experimental LF waveforms to infer current properties corresponding to simulated and to actual return strokes. This introduces additional challenges that should be accounted for in the design of the statistical inverse model. Unlike simulations, collected LF waveforms are (i) noisy, (ii) band-limited to 500 kHz (sampling period  $T_s^{\text{LF}} = 1\mu\text{s}$ ), and (iii) are not perfectly aligned in time. The time misalignment is due to the small error  $\delta d \simeq \pm 2\text{--}5$  km in estimating  $d$ , which translates into an arrival time error of  $\delta d/c \simeq \pm 6\text{--}17$   $\mu\text{s}$ , equivalent to several sampling periods. For instance, in simulations  $B_{n=0}^{\text{simulation}}$  always corresponds to the value of the magnetic field at the onset time. However, this is not necessarily the case for experimental waveforms, for which  $B_{n=0}^{\text{actual}}$  could be shifted by several samples. Equation 4.13 shows the expression that models experimental waveforms, with a time shift error term and background noise. The term  $N(0, \tau^2)$  represents the normal zero-mean white noise that is inherently present in data.

$$B_n^{\text{actual}} = B(t + d/c + nT_s + \delta d/c) + N(0, \tau^2) \quad (4.13)$$

We address the first challenge by (i) adding noise to the simulation waveforms before training or testing the network. The noise follows a normal distribution with zero mean and variance  $\tau^2$ . The standard deviation  $\tau$  is determined experimentally by analyzing collected LF data and is found to be  $\sim 0.2$  fT. As for the second challenge, (ii) we limit the bandwidth of simulations to 500 kHz, consistent with specifications of the LF receiver specifications, with  $T_s = T_s^{\text{LF}} = 1\mu\text{s}$ . As for the third challenge, ideally in the absence of time misalignment, the time series vector  $\mathbf{B}$  would be used as input to the neural network. This could work in solving for the inverse of simulation waveforms but would perform poorly with experimental waveforms, where the indexes  $n$  are not consistently aligned. (iii) To work around this, we develop a feature extraction tool (introduced in Section 2.3) to extract waveform features that are not affected by a shift in the onset time. The ground wave features capture changes in the shape and amplitude of the signal and are used as inputs to the neural network, for both simulation and experimental waveforms, instead of the time series.

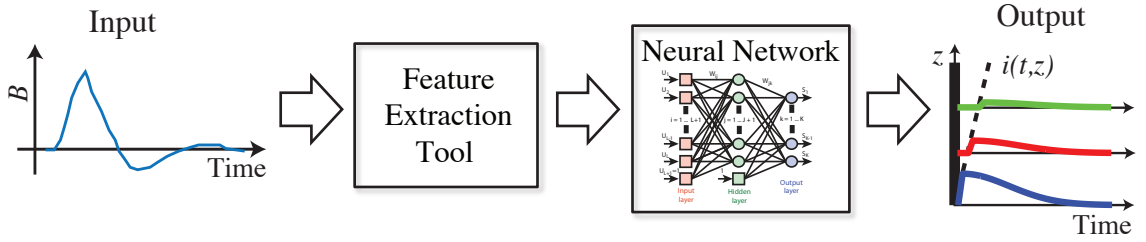


Figure 4.5: Diagram of the flow of the statistical inverse model.

Figure 4.5 illustrates the flow of the proposed inverse model. The waveform  $\mathbf{B}$  goes through the feature extraction tool and the waveform features are then fed into the neural network to infer the corresponding current parameters. As introduced in Section 2.3, the feature extraction tool uses cubic spline interpolation to extract ground wave features from  $\mathbf{B}$ , namely the peak of ground wave, the 20%-90% rise-time, the 90%-20% fall-time, and the integral under the ground wave between the 20% of the rising edge and the 20% of the falling edge, as shown in Figure 2.8. The four extracted features are independent of time misalignment and capture intuitive

variations in shape and amplitude, as discussed in the sensitivity analyses of Section 3.4. Additionally, the mapping from a time series vector  $\mathbf{B} \in \mathbb{R}^{120}$  to a vector of four waveform features reduces the dimensions of the input vector by more than 95%, leading to smaller network size, fewer weights, and more efficient training and testing.

The proposed inversion flow is valid for one distance  $d$  and one ground conductivity  $\sigma$ . In fact, each combination  $(d, \sigma)$  requires one neural network that solves for the inverse relationship between the ground wave and the current profile for that distance and ground conductivity. We are interested in two ground conductivities, 2 mS/m for land and 5 S/m for ocean, and in distances between 250 and 550 km, with 25 km spacing. The spacing is chosen to in a way that is small enough that it does not lower the accuracy of the system, which is limited by the resolution of geo-location (2–5 km), the accuracy of the experimental LF data, the accuracy of the forward model. The complete system thus consists of 14 neural networks, each denoted by  $P^{\sigma,d}$ , where  $\sigma \in [2 \text{ mS/m}, 5 \text{ S/m}]$  and  $d \in [250 \text{ km}, 300 \text{ km}, \dots, 550 \text{ km}]$ . Each ANN is trained and validated using 10,000 gwFWM simulations generated at that  $(d, \sigma)$ , with current parameters selected at random. The peak current  $I_{\text{peak}}$  is chosen from the uniform distribution  $U(10 \text{ kA}, 300 \text{ kA})$ , the current rise-time  $t_{\text{rise}}$  from  $U(1 \mu\text{s}, 10 \mu\text{s})$ , the current fall time constant  $\tau_{\text{fall}}$  from  $U(20 \mu\text{s}, 40 \mu\text{s})$ , the return stroke speed  $v_{\text{rs}}$  from  $U(c/5, c)$ , where  $c$  is the speed of light in free space, while the exponential decay constant with altitude  $\lambda$  is held constant at 1.08 km. In total,  $14 \times 10,000 = 140,000$  simulations are required to build the system, which is only reasonable to achieve with a fast and parallelizable forward model such as gwFWM. Each gwFWM run lasts 3–5 seconds on a 3–core processor. The 140,000 gwFWM simulations take less than two days to complete on the Stanford VLF Group cluster computer, which consists of ten 12–core processors.

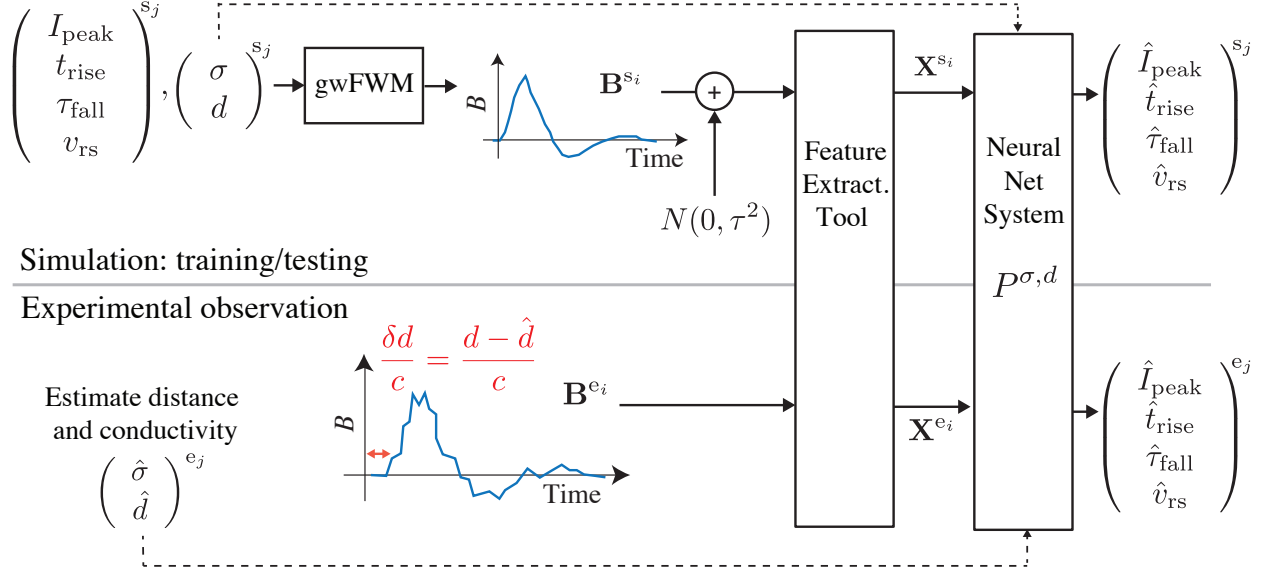


Figure 4.6: Illustration of the inverse modeling of simulation waveforms (top) and experimental waveforms (bottom).

Figure 4.6 summarizes the flow of the inversion for both simulation (top) and experimental (bottom) waveforms. For the  $i^{\text{th}}$  simulation  $s_i$ , we start off with a known distance and conductivity  $(\sigma, d)$ , a decay constant  $\lambda = 1.08$  km, and randomly selected current profile parameters  $(I_{\text{peak}}, t_{\text{rise}}, \tau_{\text{fall}}, v_{\text{rs}})^{s_i}$ . The gwFWM model uses these parameters to simulate the corresponding waveform  $B^{s_i}$ . Then white noise  $N(0, \tau^2)$  is added to  $B^{s_i}$ . At this point, the simulation waveform is ready for inversion. The inverse statistical model, which tries to estimate the original current profile parameters, consists of the feature extraction tool and of the system of ANNs. In the extraction tool, waveform features  $X^{s_i}$  are automatically extracted from  $B^{s_i}$ . These features are used as inputs to the neural network system, which uses the appropriate neural network  $P^{\sigma, d}$ , selected using the known  $(\sigma, d)$ , to predict  $(\hat{I}_{\text{peak}}, \hat{t}_{\text{rise}}, \hat{\tau}_{\text{fall}}, \hat{v}_{\text{rs}})^{s_i}$ . The mismatch between the predictions and the original current profile parameters  $(I_{\text{peak}}, t_{\text{rise}}, \tau_{\text{fall}}, v_{\text{rs}})^{s_i}$  is used to train and to assess the performance of the inverse system.

Once the feature extraction tool and the neural network system are developed and

optimized using the simulation waveforms, they could be applied to an experimental waveform. First, the propagation distance and the ground conductivity are estimated using the position of the lightning source, the receiver, and ground conductivity maps. The estimates are denoted by  $(\hat{d}, \hat{\sigma})$ . As shown in the bottom half of Figure 4.6, the  $i^{th}$  experimental waveform  $\mathbf{B}^{ei}$  is inherently noisy and is often shifted in time by  $\delta d/c$ , where  $\delta d = d - \hat{d}$  is the differential between the true and estimated distance. The experimental waveform  $\mathbf{B}^{ei}$  goes through the feature extraction tool, resulting in a vector of features  $\mathbf{X}^{ei}$ . The waveform features are fed into the neural network system to predict  $(\hat{I}_{\text{peak}}, \hat{t}_{\text{rise}}, \hat{t}_{\text{fall}}, \hat{v}_{\text{rs}})^{ei}$ , using the appropriate network determined by  $(\hat{d}, \hat{\sigma})$ . The predicted source parameters provide important insight into the physics of the return stroke, enabling remote sensing of lightning source parameters, as we thoroughly discuss in Section 4.5. In the next section, we present the training and testing of the inverse model using gwFWM simulations.

## 4.4 Application to gwFWM Simulation Waveforms

In this section, we present the training and testing of neural networks using gwFWM simulations. Each neural network  $P^{\sigma,d}$  is trained using 8,000 gwFWM runs selected at random out of 10,000 simulations at that distance  $d$  and conductivity  $\sigma$  ( $N_{\text{train}} = 80\% N_{\text{total}}$ ). The ANN is then tested with the remaining 2,000 runs ( $N_{\text{test}} = 20\% N_{\text{total}}$ ). In Section 4.4.1, we use the training dataset to study statistical patterns between input and output features and to choose network weights, configuration, and size. In Section 4.4.2, we use the testing dataset to analyze the performance of the inverse model.

### 4.4.1 Training: Waveform Features to Lightning Parameters

The training process of a single hidden layer neural network consists of several steps. Initially, the input and output variables are normalized to standard z-scores, as shown in Equation 4.14, where  $\mu$  and  $\tau$  are the mean and standard deviation of variable  $x$ , respectively. The mean and standard deviation of each input and output variable are



determined using the training data. The scaling of inputs is important as it affects the scaling of the weights of the network, potentially leading to improved performance, and standardizes the analysis of input-output correlations [Hastie *et al.*, 2009, p.398].

$$z = \frac{x - \mu}{\tau} \quad (4.14)$$

Figure 4.7 consists of  $3 \times 4$  scatter plots, each corresponding to the correlation between one return stroke parameter (target output of ANN) and one waveform feature (input of ANN) that are used in the training of  $P_{\sigma=2\text{mS/m}, d=250\text{km}}$ , the neural network that corresponds to the inverse model at 250 km over 2 mS/m. The relationship between input and output variables are similar for different distances  $d$  and conductivities  $\sigma$ , with slight changes in the shape of the correlation. The three return stroke parameters, channel base current rise-time  $t_{\text{rise}}$ , return stroke speed  $v_{\text{rs}}$ , and channel base peak current  $I_{\text{peak}}$ , are specified along the vertical axis. The four ground waveform features, 20%–90% rise-time, waveform peak, 90%–20% fall-time, and 20% to 20% integral, are along the horizontal axis. Each scatter plot consists of 8,000 data points each corresponding to one training simulation. The values of all the variables are normalized to z-scores. The scatter matrix provides insight into statistical patterns between observed waveform features and causative return stroke parameters. Our aim is to build a neural network, that takes advantage of these nonlinear and multivariate relationships, to predict return stroke parameters from waveform features.

The top left scatter plot (column 1, row 1) shows the relationship between current rise-time  $t_{\text{rise}}$  and 20%–90% ground wave rise-time, with a clear correlation between the two variables. The positive correlation suggests that longer  $t_{\text{rise}}$  lead to wider ground wave pulses (longer rise-times). The waveform rise-time could thus be used as a surrogate measure to infer  $t_{\text{rise}}$ . On the contrary, the scatter plot of (column 1, row 2) shows that waveform rise-time is not correlated with return stroke speed  $v_{\text{rs}}$ . The scatter points are spread in a random fashion, indicating that waveform rise-time presents no information about  $v_{\text{rs}}$ . The bottom left scatter plot (column 1, row 3) also shows little correlation between return stroke peak current  $I_{\text{peak}}$  and waveform

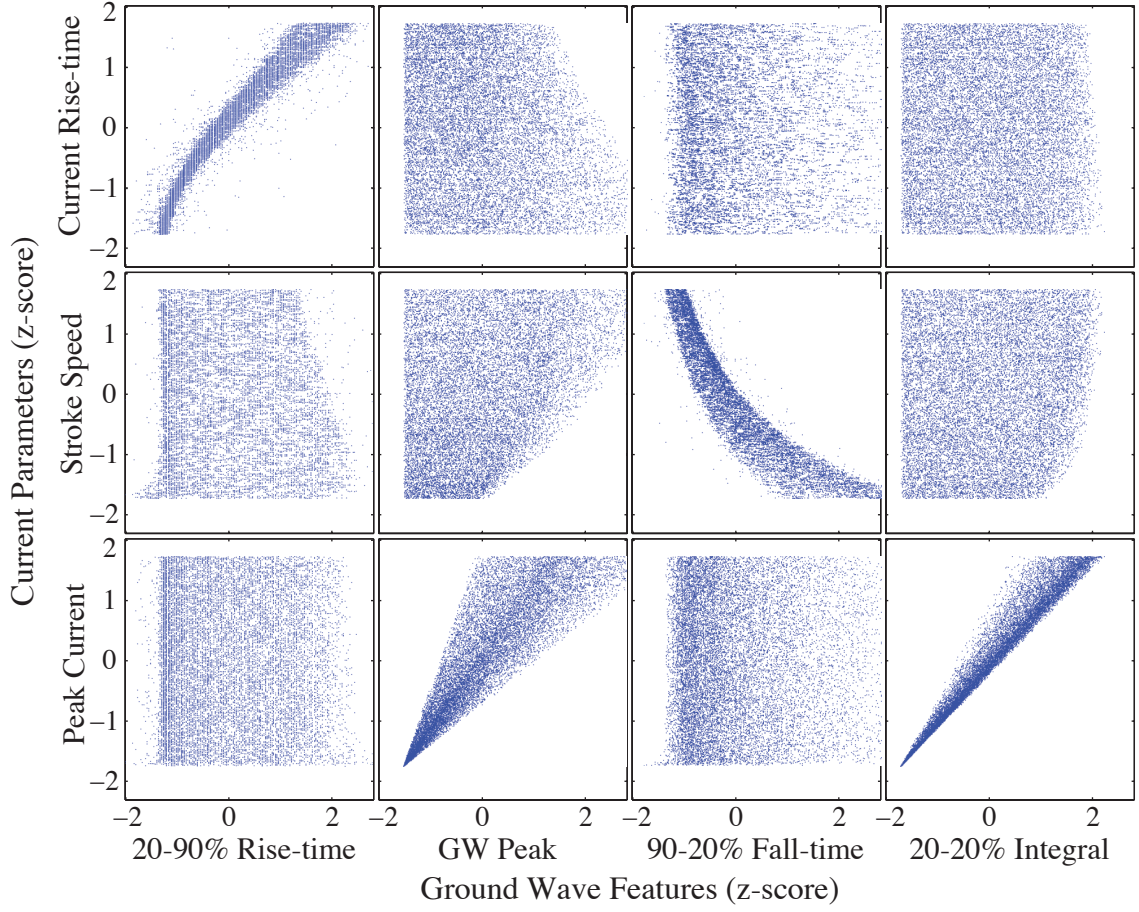


Figure 4.7: Scatter matrix that captures correlations between three current parameters (top-down: current rise-time, return stroke speed, and peak current) and four waveform features (left-right: 20%–90% rise-time, ground waveform peak, 90%–20% fall-time, and 20% to 20% integral).

rise-time.

In all, the analysis of the first column of Figure 4.7 shows that waveform rise-time mostly depends on  $t_{\text{rise}}$ . In the context of inverse modeling, the results suggest that ground wave rise-times provide information about the causative channel base current rise-time but not much about return stroke speed and peak current. Next, we analyze the scatter matrix one row at a time.

As we inspect the first row, we find that the best predictor of  $t_{\text{rise}}$  is the 20%–90% waveform rise-time (column 1). The other three features show little correlation, providing minimal prediction value. The second row shows that the only waveform feature that is useful in predicting  $v_{\text{rs}}$  is the 90%–20% waveform fall-time (column 3). The correlation is negative, with a nonlinear shape, suggesting that faster return strokes lead to shorter waveform fall-times. These results are consistent with our sensitivity analysis from Section 3.4.1. The other three waveform features correlate poorly with  $v_{\text{rs}}$ . Finally, the third row suggests that the best predictor of  $I_{\text{peak}}$  is the 20%–20% integral (column 4) followed by the LF ground wave peak (column 2). In fact, the results are consistent with NLDN (LF geo-location introduced in Section 1.3.2), which solely uses the LF peak to estimate  $I_{\text{peak}}$ , as shown in Equation 1.1. However, our results show that the 20%–20% integral feature could be used as a more accurate predictor of  $I_{\text{peak}}$  than the LF peak. Visually, the correlation in plot (row 3, column 4) is cleaner (narrower and smaller spread) than the one in plot (row 3, column 2). The smoothing effect (low pass filter) of the integration makes the 20%–20% integral feature less sensitive to changes in  $t_{\text{rise}}$  and  $v_{\text{rs}}$  and more immune to noise than the LF peak. This finding is useful in practice as it could lead to improvements in the accuracy of peak current estimates of various geo-location networks, such as NLDN. In addition, more accurate peak current estimates are of great practical interest to those studying other natural phenomena that are associated with lightning. For instance, Elves, which are optical emissions resulting from the heating and ionization of the ionosphere by lightning radiation [Fukunishi et al., 1996], are believed to occur when  $I_{\text{peak}}$  of the causative CG discharge is above a certain threshold [Barrington-Leigh and Inan, 1999; Newsome, 2010]. Thus, more accurate  $I_{\text{peak}}$  estimates could help better predict elve occurrences, compute ionospheric heating,

and improve our understanding of elves.

In summary, we find that the selected four waveform features, 20%–90% rise-time, waveform peak, 90%–20% fall-time, and 20% to 20% integral could provide enough information to drive an inverse model that predicts the source parameters  $t_{\text{rise}}$ ,  $v_{\text{rs}}$ , and  $I_{\text{peak}}$ . However, the relationships between the input and output variables are complex, nonlinear, and multivariate (multiple waveform features could be used to better infer one source parameter), which justifies the use of a statistical inverse model, such as ANN, that can capture nonlinear and multivariate relationships. Once trained, the system could be applied to testing data to assess the accuracy of predictions (Section 4.4.2) and to experimental waveforms to compare source parameters of land and ocean lightning (Section 4.5).

We use the MATLAB implementation of neural networks to train the single hidden layer system. The system uses four nodes in the input layer ( $M = 4$ ; four waveform features), three nodes in the output layer ( $N = 3$ ; three current parameters), and 10 nodes in the hidden layer ( $K = 10$ ). The optimal number of hidden nodes is determined using the training data and  $K$ –fold cross validation (refer to [Hastie et al. \[2009, p.241-247\]](#) for a description of cross-validation). We use the *tansig* activation function in the hidden layer and the linear function in the output layer. In the training process, we apply mean squared error to calculate the mismatch and LM to iteratively solve for the optimal weights, with early stopping to prevent from overfitting. By design, the small network size leads to an efficient training, lasting  $< 45$  minutes for all 14 ANNs, and to instantaneous prediction speed ( $< 5$  seconds for the  $14 \times 2,000$  test runs).

#### 4.4.2 Inverse Model Accuracy Testing

The accuracy of the system of neural network is determined using the testing dataset, which consists of  $14 \times 2,000$  simulations, where true values of  $t_{\text{rise}}$ ,  $I_{\text{peak}}$ , and  $v_{\text{rs}}$  are known. The inversion is applied to 28,000 test simulations by running them down the ANNs. Each predicted current parameter value is then compared to its true value and the percentage mismatch (or testing error) is calculated. Figure 4.8 shows

the testing error between the predicted and true return stroke parameters, namely  $t_{\text{rise}}$  (top panel),  $I_{\text{peak}}$  (middle panel), and  $v_{\text{rs}}$  (bottom panel). Each panel shows the median mismatch between predicted and true source values, and the 25<sup>th</sup>, and 75<sup>th</sup> percentiles are shown as error bars.

The top panel shows the median error of  $t_{\text{rise}}$  predictions for a range of channel base current rise-time, varying between 1  $\mu\text{s}$  and 10  $\mu\text{s}$ , with the 25<sup>th</sup> and 75<sup>th</sup> percentile error bars. The prediction errors are mostly centered around zero, indicating that there is minimal systematic bias/error. For current rise-times greater than 4  $\mu\text{s}$ , error bars are within  $-3\%$  and  $+3\%$ . For shorter current rise-times, the magnitude of the error bars increases. This is expected as the prediction performance drops when  $t_{\text{rise}}$  approaches the sampling limit of 2  $\mu\text{s}$ , below which the LF system cannot resolve features ( $2T_s^{\text{LF}} = 2 \mu\text{s}$ ).

The middle panel displays the median error of  $I_{\text{peak}}$  predictions for a range of peak currents varying between 15 kA and 300 kA, with the 25<sup>th</sup> and 75<sup>th</sup> percentile error bars. The prediction errors are mostly centered around zero with error bars between  $-10\%$  and  $+10\%$ . The error bar sizes decrease for higher peak currents to be between  $-5\%$  and  $+5\%$ , suggesting better performance in predicting higher peak currents. The improved accuracy is likely due to the higher signal to noise ratio of larger peak current events. In our application, our goal is to study and compare high peak current events over the ocean and land, which fall in the range in which the proposed inverse model performs best.

The bottom panel corresponds to the median error of  $v_{\text{rs}}$  predictions for a range of return stroke speeds varying between  $c/6$  and  $c$ , with the corresponding 25<sup>th</sup> and 75<sup>th</sup> percentile error bars. The errors are mostly centered around zero for return stroke speed greater than  $c/3$ , with error bars between  $-10\%$  and  $+10\%$ . The performance drops for slower speeds with the median error reaching  $\sim 15\%$  at  $c/6$ . The return stroke speed remains poorly understood and the subject of many current studies. Various field experiments that measure the optical speed of the return stroke along the lightning channel find that the speed of propagation varies between  $c/3$  and  $c/2$  [*Hubert and Mouget, 1981; Idone and Orville, 1984; Willet et al., 1988; Mach and Rust, 1989*]. Recent return stroke modeling results have shown that the return

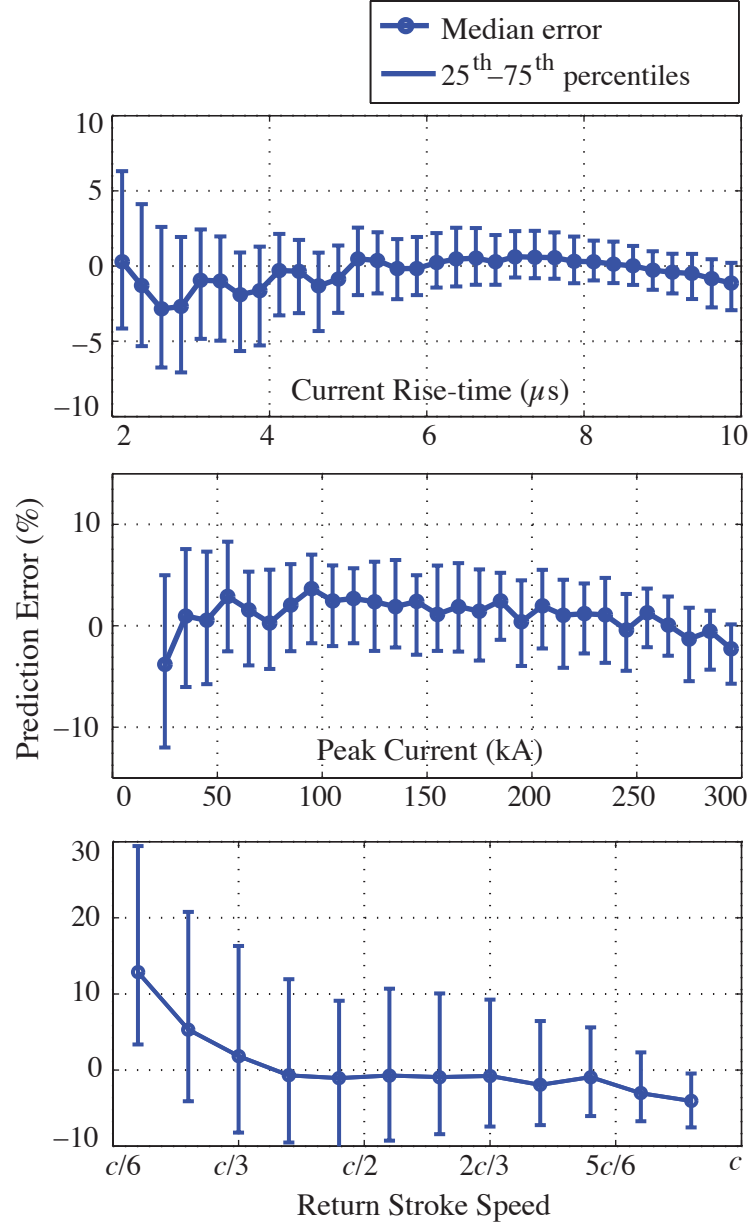


Figure 4.8: Median testing error with 25<sup>th</sup> and 75<sup>th</sup> percentile error bars for predictions of current rise-time (top panel), peak current (middle panel), and return stroke speed (bottom panel).

stroke speed could in fact be propagating faster than the optical speed, as the optical emissions due to heating lag the propagation of electric currents [Liang *et al.*, 2014] (discussed in Section 4.5.2). In all, it is however agreed that the speed of the return stroke speed is higher than  $c/3$ , falling in the range in which the inverse model performs well.

Throughout this section, we have shown that the system of neural networks that is designed and constructed in this dissertation is the first accurate inverse model that solves for the inverse relationship between LF waveforms captured several hundred kilometers from the source and lightning current parameters. The inverse model inherently accounts for all the physics of propagation and radiation that are captured by gwFWM and MTLE (refer to Section 3.3) and estimates the channel base rise-time, the speed, and the peak current of the return stroke.

## 4.5 Application to Experimental Waveforms

The proposed system of neural networks is the first model that infers lightning current parameters using the LF signature. The system is designed to work with experimental LF waveforms, which are inherently noisy and band limited. In this section, we apply the statistical inverse model to thousands of waveforms captured during our novel oceanic experiment, in which we collect highly sensitive LF data around GLD360 events with high network-reported peak currents. The inverse model provides accurate estimates of current rise-time, peak current, and return stroke speed for each discharge, which could be analyzed to compare physics of land and oceanic lightning. We are interested in comparing land and oceanic initial negative CG strokes. As discussed in the introductory chapter (Section 1.4), various studies show that estimated peak currents of initial negative CGs are higher over saltwater than over land. Due to the limitation of GLD360 in specifying the source type (CG versus IC), we only select strokes with GLD360-reported peak currents  $>30$  kA. This is done to filter out the weaker IC discharges and to mitigate the effect of sensor threshold (discussed later in Section 6.2). We apply the inversion to 1,098 oceanic and 1,649 land waveforms, all occurring at distances between 225 km and 275 km from the ship. The events

are close enough to the receiver to capture the high resolution LF features, which attenuate with propagation.

#### 4.5.1 Current Rise-Time

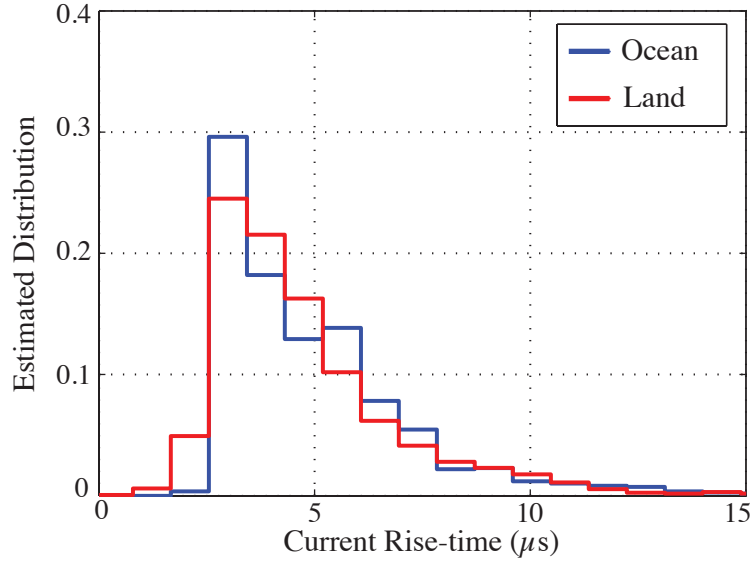


Figure 4.9: Estimated distribution of current rise-time for 1,098 oceanic (in blue) and 1,649 land (in red) discharges.

Figure 4.9 shows the estimated occurrence distribution (probability mass function) of current rise-time for ocean (in blue) and land (in red) lightning. The ocean and land distributions are found by computing the histograms of the 1,098 oceanic rise-time and the 1,649 land rise-time, respectively. The histograms are then normalized to unit area, with  $0.88 \mu s$  bin size. The two distributions are found to be visually similar. The median of the current rise-time of oceanic and land lightning are  $3.41 \mu s$  and  $3.32 \mu s$ , respectively. Our results suggest that the channel base current rise-time in oceanic and land lightning are similar and within 3%. Interestingly, the predicted rise-times are close to ones measured in the field. [Berger et al. \[1975\]](#) computed the 2 kA to peak rise-time for 101 negative CGs, using oscillograms of current measurements from the



top of two towers in Lugano, Switzerland. The median rise-time were found to be  $5.5 \mu\text{s}$  and  $1.1 \mu\text{s}$  for initial and subsequent negative CGs, respectively. The bulk of the data used in the inversion corresponds to initial discharges. The mismatch between the medians of the predicted and experimentally measured rise-times,  $3.32\text{--}3.41 \mu\text{s}$  and  $5.5 \mu\text{s}$ , could be due to the incorrect inclusion of subsequent strokes (due to GLD360 location and detection inaccuracies) which have a faster rise-time [Berger *et al.*, 1975], limited time resolution of oscillographic data, or due to inaccuracies in our forward and inverse models. Still, our inverse model is able to successfully estimate reasonable current rise-times using LF electromagnetic measurements from over 200 km distances from the source lightning. In addition, our results suggest that lightning current rise-times in the attachment processes of land and oceanic  $-CGs$  are similar, and thus do not explain the observed increase in peak currents over the ocean.

#### 4.5.2 Return Stroke Speed

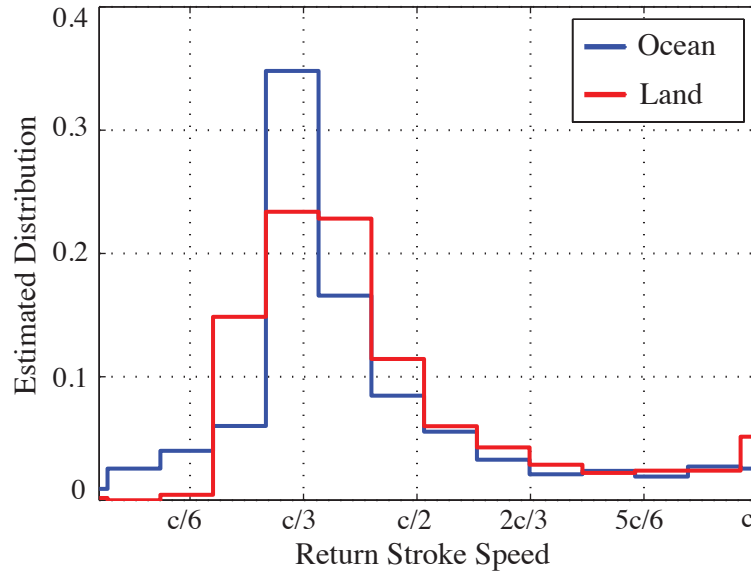


Figure 4.10: Estimated distribution of return stroke speed for 1,098 oceanic (in blue) and 1,649 land (in red) discharges.

Figure 4.10 shows the estimated occurrence distribution of return stroke speed for ocean (in blue) and land (in red) lightning. Similarly to the previous figure, the ocean and land distributions are found after normalizing the occurrence histograms of the 1,098 oceanic and the 1,649 land return stroke speeds to unity. The two distributions are found to be visually similar. The median of the return stroke speed of oceanic and land lightning are  $\sim 0.28c$  and  $\sim 0.31c$ , respectively. Interestingly, the predicted stroke speeds are close to ones measured in field experiments. To date, the return stroke speed remains poorly understood and an active field of research. Various studies have measured the speed of propagation of the optical lightning pulse, using techniques similar to ones presented by *Idone and Orville* [1982]. The optical speed at the channel base is typically around  $c/3$  and decreases to  $\sim c/6$  as it propagates upward along the channel [*Schonland et al.*, 1935]. *Idone and Orville* [1982] captured the optical speed of 63 natural negative CGs, using streak cameras, and calculated a mean speed of  $\sim 0.43c$ , and in several cases reaching up to  $0.8c$ . Other studies of natural and triggered lightning found that the optical speed varies between  $c/3$  and  $c/2$  [*Hubert and Mouget*, 1981; *Idone and Orville*, 1984; *Willet et al.*, 1988; *Mach and Rust*, 1989]. shows that it is possible that the corona sheath affect how the currents map to fields, which skews measurements to lower return stroke speed, consistent with recent return stroke modeling efforts [*De Conti et al.*, 2014]. Our inverse model uses a novel radio remote sensing technique to estimate return stroke speeds, from distant electromagnetic LF measurements, providing results consistent with the optically-inferred field experiment values. In addition, our results show that return stroke speeds in land and oceanic  $-CGs$  are comparable and do not explain the observed increase in peak currents in oceanic discharges, consistent with our conclusions from Section 3.4.3.

### 4.5.3 Peak Current

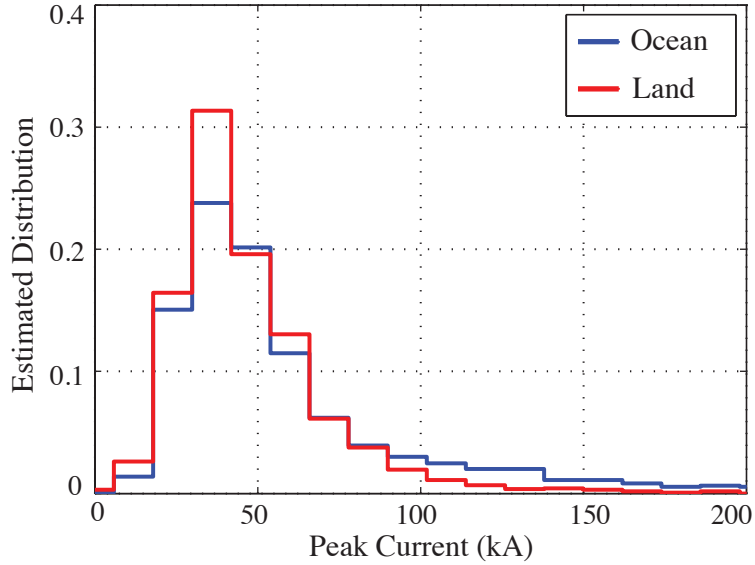


Figure 4.11: Estimated distribution of peak current for 1,098 oceanic (in blue) and 1,649 land (in red) discharges.

Figure 4.11 shows the estimated distribution of peak currents for ocean (in blue) and land (in red) lightning. Similarly to the previous figures, the ocean and land distributions are found after normalizing to unity the occurrence distribution of the 1,098 oceanic and the 1,649 land peak current predictions, with 12 kA bin size. The left portion of the two distributions are somewhat (visually) similar. However, the oceanic distribution (in blue) has a longer right tail, suggesting that oceanic discharges tend to have higher peak currents than land discharges. The median (geometric mean) of oceanic and land lightning peak currents are 35 kA (37.6 kA) and 29 kA (29.5 kA), respectively. We observe a 20–27% increase in the median and geometric mean of oceanic peak currents over their land counterparts, suggesting that the amplitude of the currents is higher in oceanic discharges. The increase is consistent with the network-reported peak current enhancements along coastlines using both GLD360 [Said *et al.*, 2013] and NLDN data [Orville and Huffines, 2001]. However, the magnitude of the enhancement in median peak currents remains much smaller than values

reported by GLD360 for deep-oceanic discharges. [Said et al. \[2013\]](#) reported that the geometric mean of peak current estimates for ocean –CG flashes are 22%-88% higher compared to land –CGs, with values reaching 100%-150% deep inside the ocean. The authors however admitted that network coverage and inaccuracies could be contributing to the observed results. In the next chapter, we analyze years of GLD360 and NLDN data to support our field experiment with large statistical studies. The statistical patterns enable us to draw a more complete picture of the peak currents climatology of land, coastal, and deep-ocean return strokes.

## Chapter 5

# Statistical Patterns in Lightning Location

Lightning discharges neutralize a fraction of the charge separation within the cloud leading to sharp variations in the ambient electric field. These variations and the subsequent electric field recovery curves were first measured on the ground below thunderstorms in the 1910s and were used to study the electrification processes and the charge structure inside thunderclouds [[Wilson, 1916](#)]. Following a lightning discharge, the electric field at the Earth’s surface takes tens of seconds to get back to its pre-flash value, with fast initial recovery rates that decrease over time [[Wilson, 1916](#); [Wormell, 1939](#); [Gopalakrishnan et al., 2011](#)]. The recovery rates are also proportional to the cloud convective activity, with faster recoveries for stronger activity levels [[Jacobson and Krider, 1976](#)].

To further study the electrification processes of thunderstorms, balloon-borne instruments with electric field sensors have been used to monitor the impact of a lightning discharge on the electric field inside the storm and to extract the true shape of the subsequent recovery. [Winn and Byerley \[1975\]](#) monitored the variations of the magnitude of the electric field inside storms following forty-four lightning flashes and found that on average the magnitude of the electric field dropped by 40% (20% standard deviation). Such field experiments have quantified the impact of lightning and have modeled the amount of charge lowered to ground, but these studies collected

limited amount of data around the experimental setup. The aim of our work is to expand these studies to a global scale and to monitor the charging and discharging processes of storms using lightning geo-location data and statistical techniques that bring out these recovery patterns. This could potentially allow the analysis of the electric field recoveries for various lightning parameters, storm types, geographical regions, and seasons. Much of the material in this chapter has been published in [Zoghzoghy et al. \[2013\]](#) and [Zoghzoghy et al. \[2014\]](#).

## 5.1 Remote Sensing of Electric Field Recovery

[Dennis \[1970\]](#) presented the delay times between flashes in a storm and found a very small autocorrelation, indicating a nearly independent process. This early study of the temporal pattern of lightning flashes in a given storm cell suggested that the flashing process of a storm is essentially time-independent. This hypothesis implied either that the spatial scales within a thundercloud are very small [[Kuettnner, 1950](#)], or that the charge neutralized in a flash is relatively small compared to the amount that was present before the flash, leaving the basic cloud charge structure largely unchanged. But recent studies have agreed that the successive flashes in a storm cannot be modeled as an independent process, implying that a typical lightning flash destroys a substantial fraction of the cloud charges. [Finke \[1998\]](#) used lightning geo-location data as a proxy measure of thunderstorm characteristics, analyzing the autocorrelation function of the lightning distribution and empirically deriving events such as velocity, lifetime, and size of lightning storms.

In this section, we introduce a statistical technique similar to that of [Finke \[1998\]](#), but applied to the shorter lengths and timescales characteristic of the charging and discharging processes in a storm, and using more lightning data with more accurate geo-location accuracy and detection efficiency, reported by NLDN and GLD360. We propose that this technique could be used as a proxy measure of the characteristic charging time and a measure of the percentage of charge transferred in a lightning flash. This statistical tool can be used as a surrogate measure of electric field recovery

and can be applied on a global level for thousands of storms to study the effects of seasonality, geographic location, and weather parameters on lightning. These advances may contribute both to lightning protection and prediction systems.

### 5.1.1 Statistical Approach

We use statistics to extract the average impact of a lightning event on the activity within its lightning cluster. We assume that this impact is stationary and does not vary with time, geographic location, and other characteristics of the corresponding lightning cluster (such as activity level, duration, and phase). The analysis of the impact of lightning for different storm types and phases requires complementing the global lightning geo-location data with global radar data, but such global radar data are unavailable. This could be the subject of a future study where the analyzed lightning data correspond to a specific region with available radar data.

The aim is to quantify the impact of a lightning event on the surrounding lightning activity, assuming that this effect is stationary. To do so, we use the following approach: around each lightning event in space and time, other neighboring events will occur based on the probability of an event occurring at that location and time delay. For instance, for a lightning event  $i$  occurring at time  $t_i$ , latitude  $\lambda_i$ , and longitude  $\phi_i$  (hence position  $(\lambda_i, \phi_i)$ ), another event  $j$  can be written by its relative time of occurrence and position compared to  $i$ , as shown in Equation 5.1. Throughout the text, we refer to event  $i$  as the reference event.

$$(\Delta t_{ij} = t_j - t_i, \Delta d_{ij} = \text{distance from event } i \text{ to } j) \quad (5.1)$$

This process is repeated for all the events in a thunderstorm and the results are summed up to construct a two-dimensional lightning-occurrence histogram of the relative distances and time delays. This histogram can be thought of as the autocorrelation function of the lightning activity [Finke, 1998]. The occurrence histogram could be used to either monitor the lightning activity in a given region as a function of time delay (fix the value of  $\Delta d$  and vary  $\Delta t$ ) or to monitor the spatial distribution of the lightning activity for a given time delay (fix the value of  $\Delta t$  and vary  $\Delta d$ ).

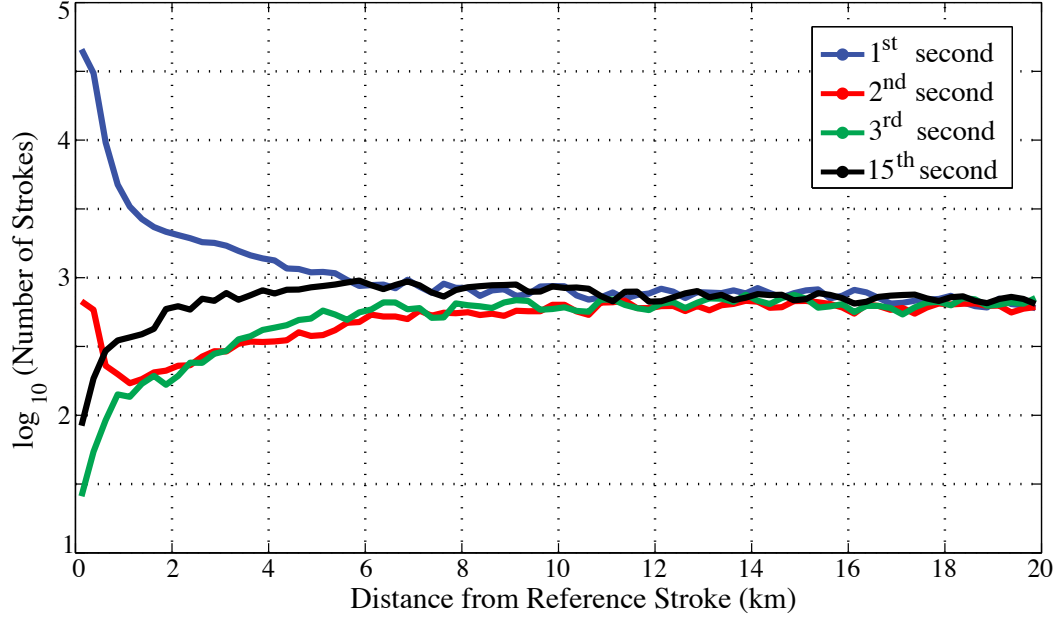


Figure 5.1: The number of strokes as a function of distance from the source stroke, plotted separately for the 1<sup>st</sup> second (blue), second 2<sup>nd</sup> (green), 3<sup>rd</sup> second (red), and 15<sup>th</sup> second (black). NLDN Data from the largest lightning cluster of 23–August–2007 are used.

It can be presumed that for large time delays,  $\Delta t > \Delta t_{\max}$ , or for large distances,  $\Delta d > \Delta d_{\max}$ , the number of events in the histogram should be roughly constant (as a function of time  $\Delta t$ ). Such a result is expected because a lightning discharge's impact on its surroundings extends only over a finite distance and time, which we define respectively as  $\Delta d_{\max}$  and  $\Delta t_{\max}$ . Throughout our analyses, we extract the values of  $\Delta d_{\max}$  and  $\Delta t_{\max}$  (roughly the spatial and temporal extent of the impact of lightning) for different lightning parameters.

### 5.1.2 Application to Negative CG Stroke Data

First, we apply this technique to NLDN –CG data, for the largest lightning storm from August 23, 2007, consisting of a total of 64,058 –CG strokes. The histogram is constructed after summing up all the relative distances and time delays of all 64,058



strokes in this cluster. Figure 5.1 shows the relative occurrences of strokes as a function of distance and parameterized by time delay, showing the spatial distribution of  $-CG$ s around other  $-CG$ s for four different time delays (one second long windows). The curves are parameterized by the first, second, third, and fifteenth seconds after the reference stroke. The relative locations of the strokes,  $\Delta d$ , are binned in 250 m distances. The four time windows are selected (a) because the first three display the lightning activity during and right after the CG discharge and (b) because the last time window (15<sup>th</sup> second) corresponds to ground-truth timescales for electric field recovery inside storms [Wilson, 1916]. In this example, only  $-CG$ s with NLDN-recorded peak current  $>5$  kA and reasonably small location uncertainty (with chi-squared value of the geo-location optimization algorithm smaller or equal to 2) are included.

$-CG$  strokes occurring within one second are dominated by those that are part of the same flash, since a flash typically lasts on the order of a second. For this particular lightning cluster, within the first second there are three notable components (blue curve): (1) A very large number of strokes values in each bin occurring in nearly the same location (i.e., within the  $\sim 400$  m NLDN geo-location uncertainty). (2) An elevated number of strokes within  $\sim 5$  km, corresponding to strokes that are part of the same flash but with different channels or ground contact points. (3) An independent region  $>10$  km where the number of strokes occurring roughly reflects the expected number given the overall stroke rate of the cluster.

The curve for the 2<sup>nd</sup> second following a stroke (red curve) is dominated by subsequent strokes that are near the end of the flashes, with an elevated number of strokes along the original channel (0–1 km) and a substantially smaller number of subsequent strokes in new contact points ( $\sim 1$ –5 km). This result is consistent with past observations that strokes in the later portion of the flash are more likely to occur along an existing channel than along a newly formed dart-stepped leader channel [Stall et al., 2009]. Thus the first two seconds following a  $-CG$  return strokes are dominated by subsequent strokes that occur part of the same flash.

In the 3<sup>rd</sup> second following a stroke (green curve), the number of strokes occurring near the reference stroke (0–1 km) drops to a minimum, the original flash is nearly

always extinguished, and a new dynamic is apparent. For distances below 10 km, there is a clear suppression in the number of strokes, as new flashes are less likely to be established near the previous flash. The effect is more prominent closer to the location of the discharge and recovers with time. This can be seen in the curve corresponding to the 15<sup>th</sup> second which shows an increase in the number of strokes occurring around the location of the initial stroke, compared to the number of strokes in these location occurring during the 3<sup>rd</sup> second (recovery). This implies that the flash partially destroys the background conditions for a flash. The lack of any suppression (or enhancement) in the number of strokes beyond 12–20 km (the number of strokes in these regions is the same for the four different time windows) would seem to be an indication that (1) the impact of lightning only extends over a finite distance for this particular cluster  $\Delta d_{\max} = 12$  km.

The background conditions that lead to lightning are still not fully understood; however, it is generally agreed that an intensification in the local electric field causes the dielectric breakdown of air, eventually leading to a lightning flash [[Hagenguth, 1951](#)]. Thus, the probability of a dielectric breakdown, hence lightning, is correlated with the electrical conditions inside the cloud: the probability of a lightning flash in a given region of the cloud is proportional to the strength of the local electric field in that region. Consequently, monitoring the recovery of the probability of a lightning flash at a given distance  $\Delta d$  from the previous flash and as a function of time delay ( $\Delta t$ ) can provide a surrogate measure of the recovery of the electric field at that relative distance  $\Delta d$  from the discharge. We follow this approach to characterize the suppression effect that is shown in Figure 5.1, extracting the timescales and the shapes of the recoveries, and quantifying the temporal and spatial extents of a lightning event (respectively  $\Delta t_{\max}$  and  $\Delta d_{\max}$ ; defined previously). In the following section, we use lightning flash data to quantify this flash suppression effect.

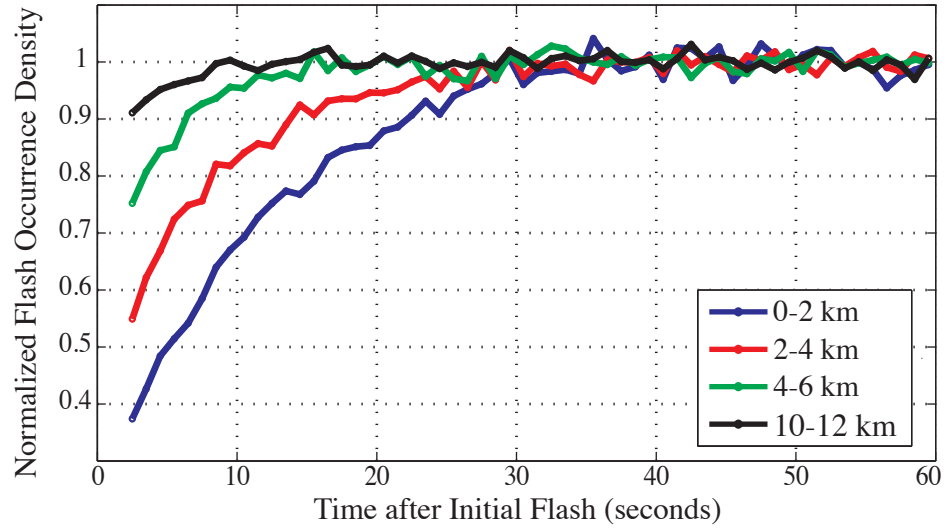


Figure 5.2: The flash occurrence density as a function of time, separately parameterized by concentric rings from the source. The recoveries in the 0–2 km region (blue), 2–4 km region (red), 4–6 km region (green), and 10–12 km region (black) are shown. All –CG flash data from August 23, 2007 are used.

## 5.2 Application to Lightning Flash Data

### 5.2.1 Negative CG Flash Data

We follow the same statistical approach to analyze NLDN  $-CG$  flash data (instead of stroke data). The NLDN  $-CG$  stroke data are clustered into  $-CG$  flash data using a clustering algorithm similar to the one described by *Cummins et al.* [1998b]. The clustering tolerances used here for a  $-CG$  flash are  $\sim 1$  second and a maximum radius of 10 kilometers. We produce the two-dimensional occurrence histogram using all the  $-CG$  flashes from August 23, 2007 (274,860  $-CG$  flashes). Figure 5.2 shows the recovery in different regions around the  $-CG$  lightning flash as a function of time. The area around the discharge is divided into concentric rings of 2 km thickness and the time delays are sorted into 1 second bins. For visualization purposes, each curve is normalized to plateau at a value of 1. This is done by dividing the entries of a given row of the occurrence histogram (for a given  $\Delta d$ ) by the steady-state (post-recovery) flashing rate, which is reached at  $\Delta t_{\max}$ . For instance, the blue curve shows the recovery of the occurrence of a subsequent  $-CG$  flash in the 0–2 km region as a function of time. We note that the first two seconds in the plot are ignored because they correspond to the original flash duration, as discussed in the paragraph following Figure 5.1.

The recovery time  $\Delta t_{\max}$  varies for different regions around the discharge, with the region nearest to the flash (blue curve) recovering in  $\sim 30$  seconds, the region around 5 km (green curve) recovering in only  $\sim 15$  seconds, and the farthest region of the cloud (black curve) recovering in only 5 seconds. Similarly, the magnitude of the suppression is more pronounced near the source discharge (blue curve), with a  $\sim 62\%$  drop in post-flash lightning activity, while at more distant regions (red, green, black curves) the activity only drops by 45%, 25%, and 9% respectively. Hence, the probability suppression effect closes in as time advances, eventually lingering only very close to the original  $-CG$  flash. This indicates that, at least for this particular summer day, the charge redistribution at 5–10 km horizontal distance from the flash is present, but the effect is much stronger at the flash location. The results depicted in Figure 5.2 are consistent with timescales of electric field recovery measured in field

experiments [[Wilson, 1916](#); [Wormell, 1939](#); [Jacobson and Krider, 1976](#)]. The magnitude of the drop in the flashing occurrence around the flash vary between 25% and 62% which is also consistent with balloon-borne experiments that measure the drop in the magnitude of the electric field inside the storm following a lightning discharge [[Winn and Byerley, 1975](#)]. These similarities between the electric field recovery and the flash probability recovery indicate that the proposed method is a valid surrogate measure of electric field buildups in thunderclouds and could be applied to study thunderstorm electrification processes for different lightning parameters for years of available lightning data, for which consistent and global electric field measurements or radar data are unavailable.

### 5.2.2 Negative CG Multiplicity

Next, we study the probability suppression effect as a function of the multiplicity of the  $-CG$  discharge (the number of return strokes per flash). Figure 5.3 illustrates the suppression effect for NLDN  $-CG$  lightning as a function of  $-CG$  flash multiplicity. The same statistical technique is applied to all the lightning activity from August 22–24, 2007 but is done separately to study the lightning activity around multiplicity 1 reference flashes (single return stroke flashes) (blue curve), multiplicity 2 reference flashes (red curve), and for the remaining flashes with higher multiplicities (green curve), with the following  $-CG$  flashes being of any multiplicity. The multiplicities are computed after clustering NLDN  $-CG$  strokes into flashes. We monitor the recovery in two different regions: one near the source discharge (0–5 km; top panel), and one in a more distant region (5–10 km; bottom panel), as a function of time delay (1 second bins). The first two seconds in the plot are ignored because they correspond to the original flash duration (discussed earlier).

We find that, on average, higher multiplicity flashes lead to longer and stronger probability suppression effects. In the region near the flash (top), the occurrence of a subsequent flash recovers in only  $\sim 25$  seconds following a single-stroke flash (blue curve) while the recovery takes  $\sim 30$  seconds after multiple-stroke flashes (red, green curves). In addition, the flash occurrence rate only drops by  $\sim 32\%$  in the

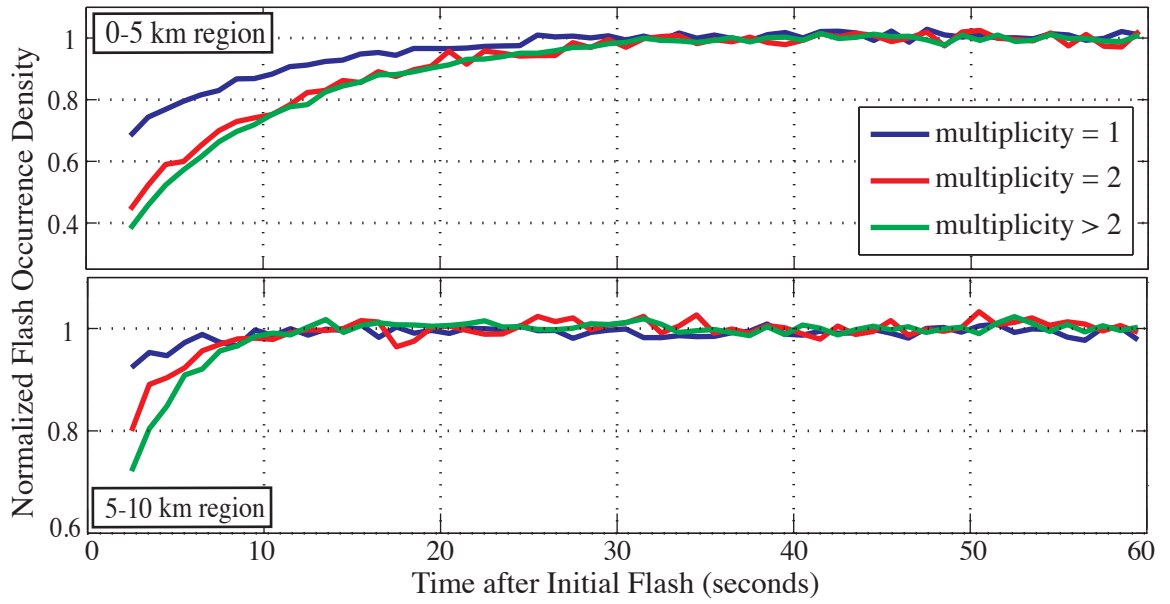


Figure 5.3: (Top) The flash occurrence density in the 0–5 km region around the source discharge as a function of time, parameterized by flash multiplicity. The flash occurrence densities are plotted separately for flashes with multiplicity 1 (blue), multiplicity 2 (red), and higher multiplicities (green). (Bottom) Similar analysis but for the 5–10 km region. –CG flash data from August 22–24, 2007 are used.

single-stroke case (blue curve), while the other two curves (red, green curves) drop  $\sim 56\%$  and  $\sim 62\%$  respectively. The more pronounced suppression effect following higher multiplicity flashes indicates that the total charge neutralized by a flash is proportional to the flash multiplicity; additional return strokes transport more charge to ground, requiring a longer re-charging process before another flash can take place.

The curves corresponding to the single-stroke flashes (blue curves, top and bottom panels) show a substantial suppression effect in the region nearest to the flash (top) but a substantially smaller effect in the 5–10 km region (bottom), where the flash occurrence is uniform over time. This indicates that a single return stroke mostly neutralizes charge from the region around the stroke and does not destroy an appreciable portion of the charge from more distant regions (5–10 km), which seems to suggest that a single return stroke, on average, has a maximum spatial extent of 5 kilometers.

On the other hand, the two curves corresponding to multiple-stroke flashes (red and green curves) still show a significant suppression effect at 5–10 km. This suggests that multiple-stroke flashes can neutralize charge from distant locations of the cloud, more than 5 kilometers away from the initial channel. These conclusions are consistent with results from field experiments where an electric field measuring system is used to study the charge source locations for subsequent return strokes and continuing currents in multiple-stroke flashes in New Mexico [*Krehbiel et al.*, 1979].

### 5.2.3 Negative CG Peak Current

Figure 5.4 shows the flash suppression effect as a function of NLDN-reported return stroke peak current, determined separately for intense ( $>30$  kA, in red) or less intense (5–20 kA, in blue) –CG strokes from August 22–24, 2007. Only single-stroke flashes are analyzed to eliminate the effect of multiplicity. More intense lightning return strokes seem to more strongly suppress the occurrence of subsequent flashes, but the recovery characteristics appear to be similar. This may indicate that more intense –CG strokes are redistributing a larger amount of charge, but are also occurring in systems where the charging currents are particularly strong and counteract the

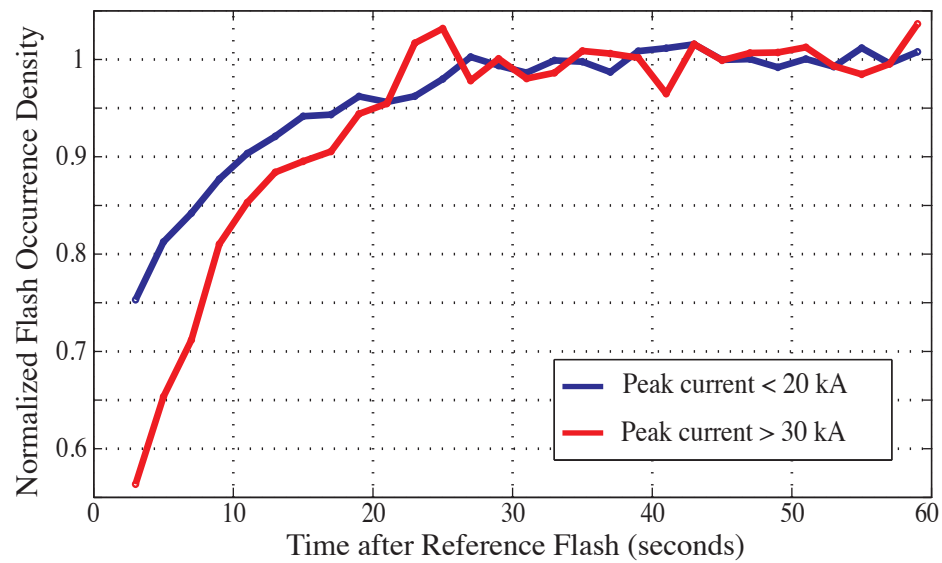


Figure 5.4: The flash suppression effect as a function of time and parameterized by distance, separately for  $-CG$  single-stroke flashes with peak currents in the 5–20 kA range (blue) and for  $-CG$  single-stroke flashes with peak currents greater than 30 kA. Single-stroke  $-CG$  flash data from August 22–24, 2007 are used.



larger charge neutralization. Although the peak current of the return stroke is not necessarily proportional to the total charge transfer in the lightning stroke (since the duration of the stroke and the intensity of continuing currents is important), since we are considering here only a specific type of stroke ( $-CG$ ) they are at least correlated.

So far, we have focused on  $-CG$  land lightning flashes. Figures 5.1–5.4 show that  $-CG$  flashes affect the occurrence of subsequent  $-CG$  flashes, implying destruction of charge in the negative cloud charge layer, and that the extent of suppression varies with multiplicity and peak current. For the remaining of this section, we extend our analysis to include different flash types and oceanic and land lightning.

### 5.2.4 Ground and Cloud Flash Type

We explore the flash occurrence suppression effect for different flash types (IC and CG) using NLDN IC and CG data. We note that our statistical technique is not affected by the low IC NLDN detection efficiency (10–20%). The low detection efficiency is equivalent to a random sampling of the entries of the occurrence histogram to remove data points, deleting equal number of samples from each histogram entry (assuming that the detection efficiency is uniform for IC flashes). This process does not affect the underlying probability distribution but requires a larger sample size (more lightning data) to better estimate the recovery patterns. We use all lightning data collected by NLDN from 02–25 July, 2011.

Figure 5.5 presents the interactions between different flash types. The first panel (top–left) shows the distribution of CG flashes (both polarities) following a CG flash, the second (top–right) shows the distribution of IC flashes (both polarities) following an IC flash, the third (bottom–left) is for IC activity following a CG discharge, and the fourth (bottom–right) is for CG activity following an IC discharge. Each panel is produced using the aforementioned statistical method applied to all NLDN IC and CG data from 02–25 July, 2011 (6,271,850 IC flashes, 13,653,455 CG flashes). For instance, the occurrence histogram of the third panel is generated by counting the relative distances and time delays of IC flashes that follow CG flashes. The plots include a logarithmic time axis and are normalized by the steady-state flashing rate

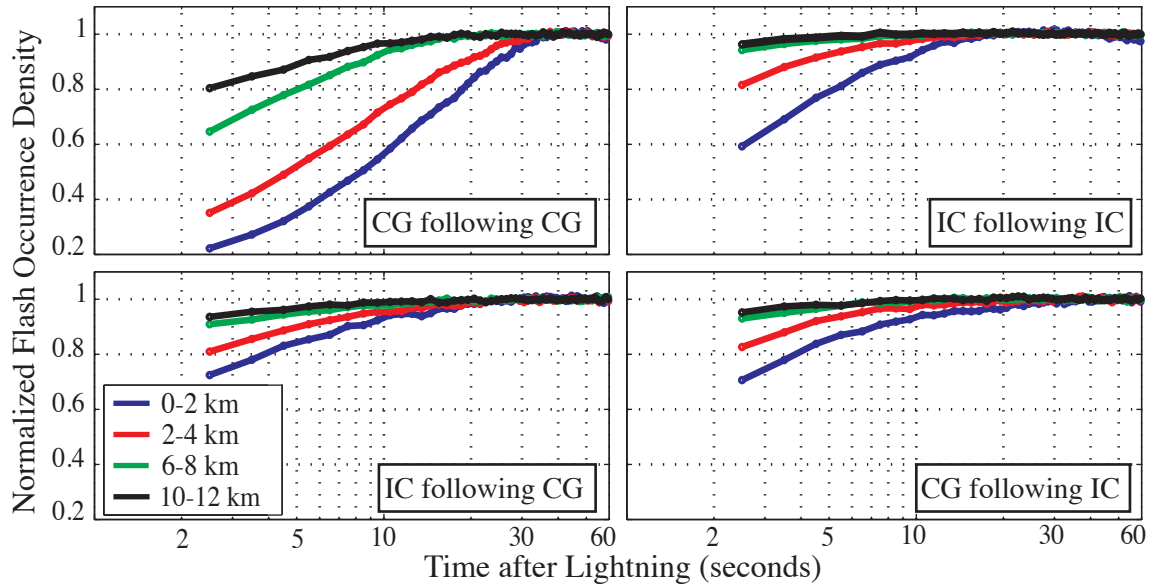


Figure 5.5: The flash suppression effect as a function of time parameterized by flash type and distance from the discharge: (top-left) recovery of CG lightning occurrence following a CG discharge, (top-right) recovery of IC lightning occurrence following an IC discharge, (bottom-left) recovery of IC lightning occurrence following a CG discharge, (bottom-right) recovery of CG lightning occurrence following an IC discharge. The recoveries in the 0–2 km region (blue), 2–4 km region (red), 6–8 km region (green), and 10–12 km region (black) are shown. NLDN CG and IC flash data from July 02–25, 2011 are used.

(plateau at a value of 1). The first two seconds are ignored (discussed earlier).

All four panels suggest that the flash occurrence suppression effect exists for the four scenarios, but the strength of the suppression, in terms of duration and magnitude, varies. In the 0–4 km region (blue and red curves; panel 1–4), the lightning activity following any type of discharge is affected for tens of seconds, implying that a lightning flash partially destroys the local conditions for a subsequent discharge. The more prominent suppression effect corresponds to the CG following CG case (Panel 1). The impact of a CG flash (on other CG lightning) extends to  $\sim 12$  kilometers ( $\Delta d_{\text{max}}$ ) and affects the region around the CG flash for  $\sim 30$  seconds (top–left panel, in blue). The magnitude of the drop in lightning occurrence varies from  $\sim 76\%$  (0–2 km, in blue) to  $\sim 37\%$  in more distant regions (6–8 km, in green).

The second panel of Figure 5.5 shows the occurrence distribution of IC flashes following IC flashes. The recovery pattern is still present but is much weaker than the CG following CG case (Panel 1). The impact here extends to only  $\sim 5$  kilometers ( $\Delta d_{\text{max}}$ ) and affects the region around the CG flash for  $\sim 15$  seconds (top–right panel, in blue). The magnitude of the drop in lightning occurrence varies from  $\sim 40\%$  (0–2 km, in blue) to  $< 5\%$  in more distant regions (6–8 km, in green). Although IC lightning may or may not neutralize more charge than CG lightning, our results suggest that an IC discharge does not affect the conditions for another IC discharge as efficiently as a CG discharge affects the conditions for another CG discharge. This suggests a fundamental difference between IC and CG flashes and their corresponding initiation and electrification processes.

The third panel (bottom–left) and the fourth panel (bottom–right) present the occurrence distribution for IC following CG and for CG following IC respectively. Both recoveries only extend to  $\sim 4$  km and are similar in magnitude and in duration. The resemblance between the cross-type results indicate that both lightning types equally affect the flashing conditions of the other. The cross suppression effects are present but not as strong as the same-type suppression effects (Panel 1 and 2), suggesting that a lightning flash is more effective in destroying the conditions of a subsequent flash of the same type. Figure 5.5 provides more insight into the different types of discharges and the physics behind lightning, but the theoretical

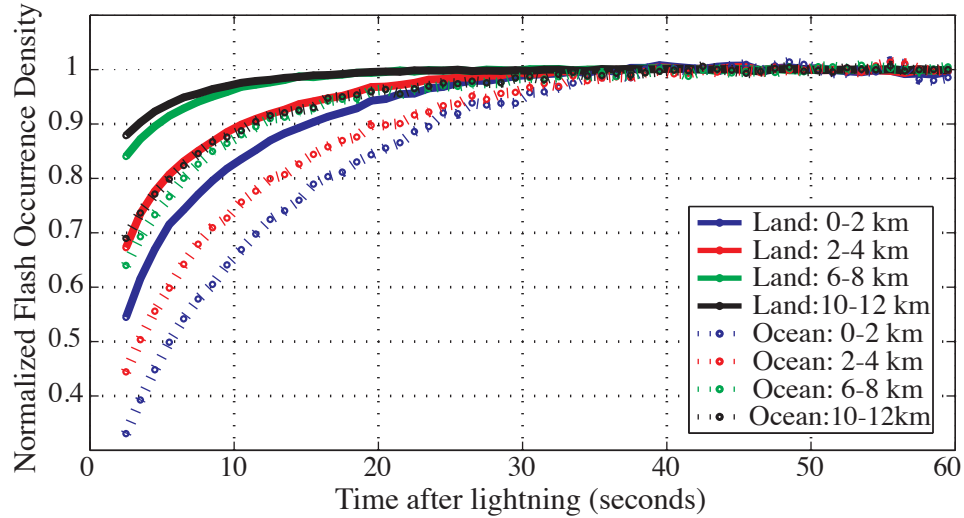


Figure 5.6: The flash suppression effect as a function of time for land and for oceanic lightning parameterized by distance. The solid lines correspond to the recovery following land lightning and the dashed lines for the oceanic lightning recovery. All GLD360 flashes from July 02–25, 2011 are used.

interpretations of these results are beyond the scope of this paper.

### 5.2.5 Land and Ocean Lightning

Although lightning occurs about an order of magnitude more frequently over land [Christian *et al.*, 2003], recent studies have shown that the average measured fields radiated from oceanic lightning is much higher [Orville *et al.*, 2001; Lyons *et al.*, 1998; Cummins *et al.*, 2005]. This indicates that either the meteorology of oceanic lightning, or the effect of highly conductive seawater, has a remarkable tendency to intensify the observed field when an oceanic flash does occur.

Figure 5.6 compares the recoveries following land lightning (solid—line curves) to the recoveries following oceanic lightning (dashed—line curves), parameterized by distance from the lightning discharge. The plots are produced using the same statistical approach applied to all GLD360 flash data from 02–25 July, 2011. GLD360 lightning data is global and includes both IC and CG discharges (GLD360 does not classify

the flashes). The first two seconds are ignored (discussed earlier) and the curves are normalized by their steady-state flashing rate (plateau at a value of 1).

The results suggest that more prominent suppression effects follow oceanic lightning. The lightning activity in the region around the oceanic discharge (0–2 km, dashed–blue) drops by  $\sim 65\%$  and recovers in  $\sim 40$  seconds. The activity in the same region around a land discharge (0–2 km, solid–blue) only drops by  $\sim 45\%$  and recovers in only  $\sim 30$  seconds. The remaining curves show that the average impact of land lightning on the 2–4 km region (solid–red) is comparable to the impact of an oceanic discharge on the more distant 10–12 km region (dashed–black), providing a rough idea about the greater extent and duration of the post-oceanic lightning suppression effect. These results suggest either that (1) oceanic lightning destroys more charge due to its high intensity, leading to longer recoveries, or that (2) oceanic storms have slower charging processes (weaker convective activity) than land storms, hence take more time to recover.

### 5.3 Patterns between Negative CG Strokes

As presented in Section 1.1, a typical  $-CG$  flash consists of multiple individual return strokes. The spatial and temporal distributions of various lightning events throughout the discharge provide a surrogate look inside the  $CG$  flash and offer insight into the underlying physical processes. In this section, we conduct a multi-year analysis to study spatio-temporal patterns of inter-stroke lightning activity and the statistical linkage between strokes in  $-CG$  flashes. We combine eight years of lightning data from the North Alabama LMA and NLDN to investigate leader propagation inside the cloud throughout  $CG$  flashes and to determine the spatial and temporal linkage between  $-CG$  strokes. We propose that a  $-CG$  stroke can produce a new cloud leader that propagates at lower cloud altitudes, potentially turning into a stepped leader forming a new channel for subsequent strokes. These advances contribute to the understanding of in-cloud activity in  $-CG$  flashes, the development of lightning grouping algorithms, and a more accurate interpretation of LMA readings around ground lightning.

### 5.3.1 Previous LMA Studies

Many studies have used LMA and other systems to study the formation of the IC lightning channel. *Krider et al.* [1975] was the first to investigate the regular pulse bursts in IC flashes. ICs are composed of two stages, an early/active stage and a late/final stage [*Rakov and Uman*, 2007, p.322]. The early stage lasts tens to hundreds of milliseconds, consisting of a channel that extends in an intermittent manner with a speed in the order of 100 km/s [*Shao and Krehbiel*, 1996; *Akita et al.*, 2010]. In general, this active stage is similar to the breakdown process and the propagating stepped leader in  $-CG$  flashes [*Rakov and Uman*, 2007, p.340]. The early stage transitions to the late stage as the connection between the positive and negative charge layers weakens. The late stage transports negative charge from more distant pockets of charge to the location of the discharge. The late stage is also known as the J-type stage due to similarities of the associated physical process to the J-process in  $-CG$  flashes. The various transient processes throughout the late stage are referred to as K-processes and are sometimes called “recoil streamers” [*Rakov and Uman*, 2007, p.322]. These K-processes can retrace the same path several times and lead to step-like field changes known as K-changes [*Akita et al.*, 2010]. The observation that many K-changes occur with regular pulse bursts suggests that a process similar to the stepped leader in  $-CG$  flashes is involved [*Rakov et al.*, 1996].

The stepped leader currents in cloud flashes and the stepped leader currents in ground flashes are comparable [*Proctor*, 1997]. *Shao and Krehbiel* [1996] reported that the propagation speeds of initial leaders in ICs and CGs are similar at approximately 100 to 300 km/s. *Proctor* [1981, 1991, 1997] further studied the initial breakdown processes in cloud and ground flashes and showed that CGs only initiate at lower altitudes while ICs originate at both higher and lower altitudes. *Proctor* [1997] also showed that CGs and lower-origin ICs have indistinguishable VHF and UHF signatures and that the characteristics of the breakdown in CGs and lower-origin ICs “differ in no way that can be detected at Medium Frequency (MF; 0.3–3 MHz), at High Frequency (HF; 3–30 MHz), or at VHF.” Further, these studies argued that there is no way of determining a priori whether subsequent stages of a flash following a low-origin breakdown will or will not involve a path to ground.

### 5.3.2 Application to NLDN Negative CG Stroke Data

We apply the statistical approach presented in section 5.1.1 to all of the NLDN-detected  $-CG$ s from July 2010 to study the spatial and temporal distribution of first subsequent  $-CG$  strokes following initial  $-CG$  strokes. The occurrence histogram is constructed using 2,033,935 pairs of events (initial, 1<sup>st</sup> subsequent) that occur within 10 km and 1 second time delay. We only use  $-CG$  data with peak currents stronger than  $-15$  kA to minimize the population of mislabeled IC strokes [Cummins and Murphy, 2009].

The histogram values are scaled by their corresponding bin area (more distant bins map to larger physical regions) and the summation over all the histogram values is normalized to unity. The resulting (normalized) space-time occurrence distribution of 1<sup>st</sup> subsequent strokes around initial  $-CG$ s is shown in Figure 5.7. The radial distances  $\Delta d$  are binned into 100 m bins and the relative time delays  $\Delta t$  are in 4 ms bins.

Two features stand out in Figure 5.7: (i) a horizontal feature that extends in time (up to 800 ms) dominated by events that are within  $\sim 750$  m and (ii) a vertical (but slightly tilted) feature that extends from 1 km to 7 km. The horizontal feature corresponds to subsequent strokes that recur along the existing ground channel (EGC). EGCs occur in the same location as the initial return stroke but the EGC feature appears to have a non-zero spatial width due to the NLDN geo-location error (two strokes that occur in the same channel should have the same location). We note that the EGC feature also includes subsequent events that follow dart-stepped leaders that deviate from the initial channel at lower altitudes (0.7–3.5 km) and contact the ground within 750 m. These events affect the spatial and temporal shape of the EGC feature.

The second feature corresponds to subsequent strokes that create a new ground channel (NGCs) 1–7 km away from the location of the initial channel. The ability to visually separate between the two features is consistent with ground-truth video recordings that show that NLDN can identify the different channel locations in a  $-CG$  discharge with a separation criterion of about 1 km throughout most of the United States [Stall et al., 2009]. Stall et al. [2009] found the mean horizontal separation

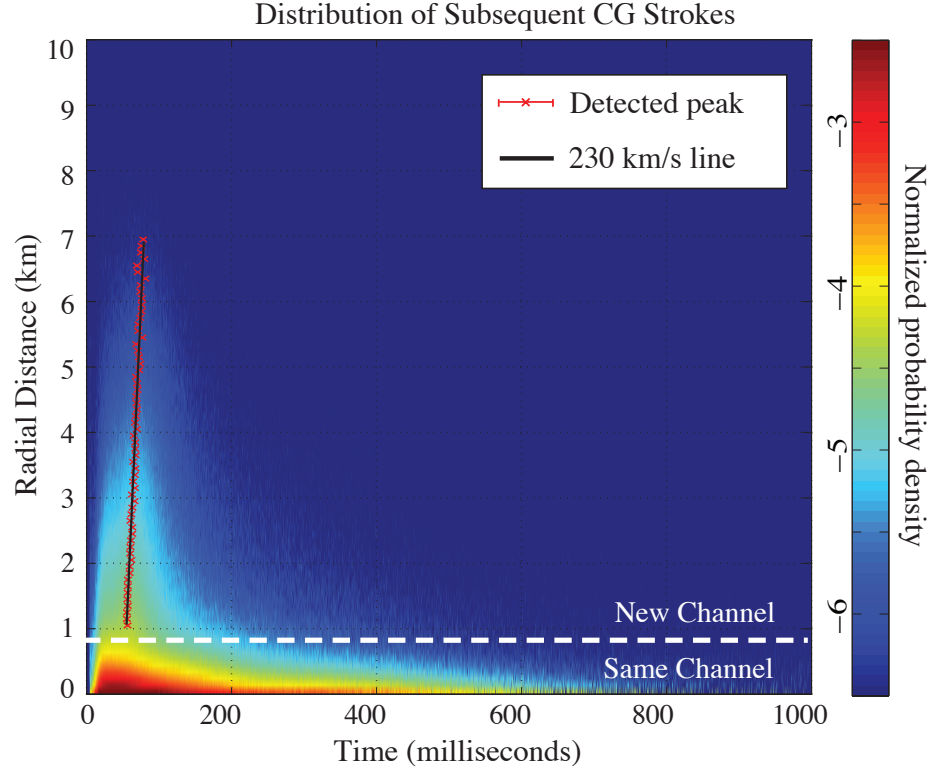


Figure 5.7: Spatial and temporal distribution of subsequent strokes around initial strokes, using 2,033,935 NLDN –CG strokes from July 2010. The histogram is scaled by the differential area of each radius bin, the summation over the histogram entries is normalized to unity, and the resulting values are displayed (color coded) on a logarithmic scale. The white dashed line corresponds to a radial distance of 750 meters, marking an approximate separation between the NLDN-reported location of subsequent strokes that recur along the same channel and the reported location of strokes that form a new one [Stall *et al.*, 2009]. The solid black line corresponds to the least squares linear fit through the peaks of lightning occurrence as a function of radial distance.



between first and other subsequent strokes to be  $2.3 \pm 1.7$  km (59 observations). [Thottappillil et al. \[1992\]](#) used video recordings to compute the distances between all different pairs of contacts in 22 CG flashes and found an average spatial separation of 1.7 km and a maximum separation of 7.3 km. [Ishii et al. \[1998\]](#) observed that the spatial extent of a flash can be as large as 10 km for  $-$ CG flashes with an average of 2.1 km (40 observations). Multi-stroke flashes and the spatial separation between subsequent strokes are important for both lightning attachment physics and for lightning safety and protection systems and are an active field of research [[Fleenor et al., 2009](#); [Saba et al., 2010](#); [Ballarotti et al., 2012](#)].

We note that NLDN could miss the 1<sup>st</sup> subsequent stroke and detect a later one due to the 60–80% network stroke detection efficiency [[Cummins and Murphy, 2009](#); [Nag et al., 2011](#)]. Thus, these invalid pairs of events occur later in time and add a right tail (in time) to the EGC and NGC features. However, their effect should be minimal for a large sample size.

Roughly half of the first subsequent strokes are NGCs and the other half EGCs. We follow the same approach to study (i) the spatial and temporal distribution of 2<sup>nd</sup> subsequent strokes around 1<sup>st</sup> subsequent strokes and (ii) the distribution of 3<sup>rd</sup> subsequents around 2<sup>nd</sup> subsequents. The resulting distributions are (visually) similar to the one presented in Figure 5.7 showing both the NGC and the EGC features. However, the proportion of EGC events increases from 49% for first subsequent strokes (following initials) to 60% for second subsequents (following first subsequents) and to 68% for third subsequents (following second subsequents), suggesting that higher order subsequent strokes are more likely to recur along the preceding channel. These proportions are consistent with various ground-truth studies [[Stall et al., 2009](#); [Rakov and Uman, 1990b](#); [Rakov et al., 1994](#); [Valine and Krider, 2002](#); [Saba et al., 2006](#)]. Although the proportion of NGCs and EGCs vary between these field studies (mostly due to the limited sample size) the authors suggest that the first subsequent stroke in a  $-$ CG flash is more likely to produce a new ground termination.

The vertical NGC feature is characterized with a (roughly) constant speed. We determine this speed as follows: In each distance bin of the histogram, we first apply a 20 ms moving average window to smooth the histogram entries as a function of

time, resulting in  $H_{A,B}^*(\Delta d, \Delta t)$  and then find the peak time delay  $\Delta t_i^{\text{peak}}$  as shown in Equation (5.2). We repeat the process for all distances between 1 and 7 km. The peaks are shown in red. The black line is a least squares regression best-fit line:  $\Theta_{ls} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}$ , where  $x_i = \Delta t_i^{\text{peak}}$  and  $y_i = r_i$ , and  $\Delta t_i^{\text{peak}}$  is given by:

$$\Delta t_i^{\text{peak}} = \arg \max_{\Delta t} H_{A,B}^*(\Delta d = r_i, \Delta t) \quad (5.2)$$

Our results suggest that the peak –CG activity around the initial stroke is moving outward at a speed of  $\sim 230$  km/s. For instance, a subsequent stroke occurring 5 km away from the initial stroke is most likely to occur after  $\sim 77$  ms delay. We repeat this analysis for three separate months of NLDN data (August 2010, July 2011, August 2011) and find the slope of the NGC feature to vary between 200 and 280 km/s, so it is possible that there is some variation by storm type and season. These speeds are comparable to the speed of the stepped leader reported by [Rakov and Uman \[2007, p.123\]](#) and to the 100–200 km/s speeds measured by [Mazur et al. \[1995\]](#); [Montanya et al. \[2014\]](#) using high-speed video recordings.

It is believed that the formation of a new ground contact point follows a dart-stepped leader. [Davis \[1999\]](#) showed that dart-stepped leaders deviate from the existing channel at heights of 0.7 to 3.4 km then travel to ground at typical speed of 100–300 km/s. This mechanism is consistent with the observed speed of the NGC feature but is unlikely to explain the formation of new ground terminations with 4–10 km separation distances (supported by the aforementioned field studies) as the dart-stepped leader would have to propagate horizontally for several kilometers after it branches out at 0.7–3.4 km heights. We propose another mechanism in Section 5.3.3 that could potentially explain NGCs that form with separation distances as large as the ones observed.

### 5.3.3 Proposed Mechanism

[Proctor \[1981, 1997\]](#) used VHF and UHF imaging systems to study cloud discharges, and suggested that horizontal ICs follow a horizontally-propagating cloud stepped leader. The two studies described four lower-origin horizontal ICs following horizontal

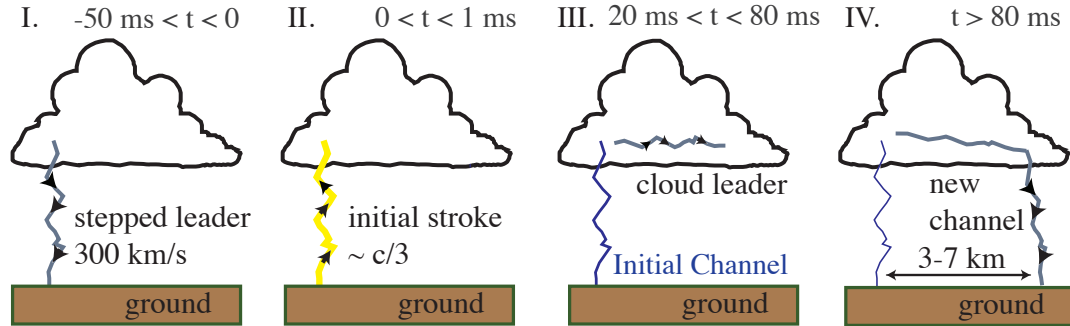


Figure 5.8: Cartoon of the suggested physical process that governs the formation of distant subsequent channels in  $-CG$  flashes. (Panel I) A stepped leader propagates downward at a speed of 100–300 km/s to form an initial channel. (Panel II) The return stroke propagates upward along the ionized channel at  $\sim c/3$ . (Panel III) A leader follows the return stroke propagating outward in the cloud. (Panel IV) The cloud leader turns into a downward stepped leader and forms a new lightning channel 3–7 kilometers from the location of the initial one.

leaders and claimed that stepped leaders in lower-origin ICs are indistinguishable from the stepped leader in CGs. Additionally, the author argued that there is no way of determining a priori whether the flash following a low-origin breakdown will or will not involve a path to ground.

The cartoons shown in Figure 5.8 present a lightning mechanism that could potentially explain our observations. Panel I and II illustrate the initial stepped leader which precedes the initial CG return stroke, producing a conductive channel in virgin air. Following the CG stroke, we propose that in-cloud leaders could form at the tip of the lightning channel, propagating horizontally in the cloud, as shown in Panel III. The cloud leader could then grow into a second cloud-to-ground stepped leader, producing a new ground termination several kilometers away from the previously existing one. The proposed mechanism could explain the deterministic linkage between distant return strokes within the same  $-CG$  discharge.

Previous observations and studies support this mechanism and offer similar physical explanations. [Montanya et al. \[2014\]](#) observed a ground-to-cloud-to-ground flash

which initiate as an upward negative leader with a 100 km/s propagation speed. The leader then branched out in the cloud into negative and positive leaders (at 5–6 km altitude), and turned into a stepped leader producing a –CG several kilometers from the initiation point. [Mazur et al. \[1995\]](#) studied a negative CG flash consisting of six return strokes using high-speed video recordings, a VHF interferometer, and electric, magnetic, and optical sensors, and observed that the first subsequent return stroke occurs in a new channel following a stepped leader that initiated from the location of the initial channel. [Krehbiel \[1981\]](#) offered a physical explanation for the formation of new channels using electric field data from multiple-channel –CGs, and found that new channels are preceded by strokes that have a cut-off along the lower extent of the channel. The resulting deposition of negative charge along the channel (above the cut-off) could lead the subsequent leader to follow a different path to ground. The author notes that the cut-off field changes are most pronounced after strokes initiated by stepped leaders, which is consistent with our results from Section 5.3.2.

Next, we use NALMA 3-D maps of VHF sources to image the intermittent breakdown activity inside the cloud throughout –CG flashes to observe the signature of the proposed mechanism.

### 5.3.4 LMA Case Studies

Figure 5.9 consists of three NLDN –CG flashes from July 12, 2011 with overlaid NALMA data. The altitude of the LMA sources is plotted as a function of time delay relative to the initial NLDN return stroke. All three –CG flashes consist of an initial and a subsequent stroke. For visual reference, we draw two vertical lines, one in blue and another in red, to highlight the time of the initial and 1<sup>st</sup> subsequent stroke, respectively. The NLDN reported separation between the strokes is 3.5 km (top panel), 4.7 km (middle panel), and 5.8 km (bottom panel), suggesting that these subsequent strokes occur in new channels. The separation distances are well above the median geo-location error and are consistent with the LMA locations. We only include LMA events that occurred within a 10 km horizontal range around the location of the initial stroke, ensuring the LMA-detected sources are predominantly

related to the intermittent inter-stroke activity (corresponding to that flash).

All three initial strokes are followed by VHF activity inside the cloud at altitudes near 5 km. The intermittent activity then grows into a stepped leader preceding the NGC. The presence of VHF radiation as the leader propagates to ground suggests that the leader descends in a stepped-like fashion, forming a new channel in virgin air (dart leaders are nearly invisible in VHF). Figure 5.9 shows that all three NGCs do not follow dart-stepped leaders and are likely due to a mechanism consistent with the one proposed in section 5.3.3. In addition, Figure 5.9 suggests that more distant NGCs have longer inter-stroke time delays. The separation distances and inter-stroke time delays are (3.5 km, 95 ms), (4.7 km, 107 ms), and (5.8 km, 117 ms) for the top, middle, and bottom panels, respectively. This suggests the presence of a feature that propagates away from the location of the initial stroke at an average speed  $\sim 110$  km/s, taking more time to reach greater separation distances. To better observe this effect, we investigate the 3-D distribution of LMA sources as a function of time throughout the flash.

Figure 5.10 provides a top-down view (top panel) and a side-view (bottom panel) of the  $-CG$  flash from the bottom panel of Figure 5.9. The top-down view shows the 2-D location (latitude and longitude) of the LMA sources color coded in time. The side-view shows the altitude of the sources throughout the duration of the flash. Both panels use the same color coding scale, with a reference time corresponding to the onset of the initial return stroke (shown in blue). The locations of the initial and subsequent strokes are marked in both panels in blue and in red, respectively.

The top-down view reveals that inter-stroke LMA sources initially occur around the location of the existing channel and then propagate away to the location of the new one. The majority of the horizontal propagation occurs within the cloud (green to orange sources  $\sim 40$ -80 ms) at an altitude of  $\sim 5$  km. The leader propagates downward to form an NGC 5.8 km away. The LMA sources form a typical stepped leader VHF signature; the dart-stepped leader is unlikely to produce such observations.

Both Figure 5.9 and Figure 5.10 provide evidence that the initial return stroke could potentially lead to a new stepped leader, initially at the tip of the preexisting channel, which propagates horizontally in the cloud and then downward to create an

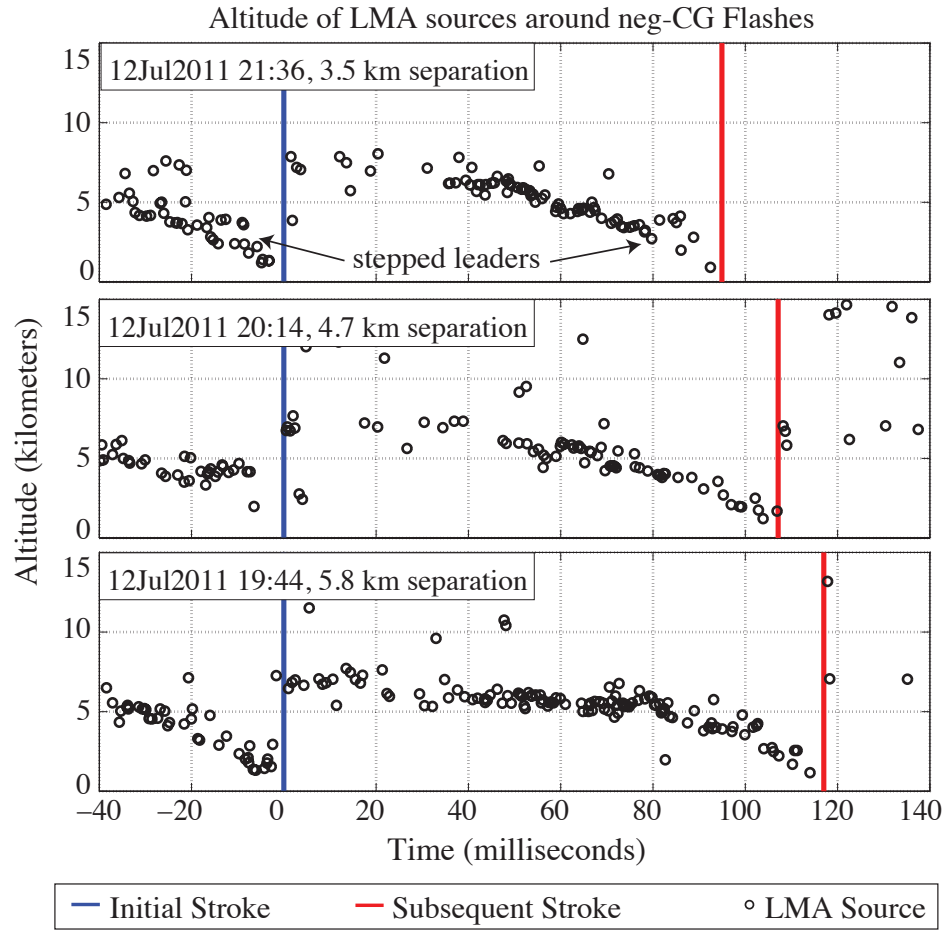


Figure 5.9: Case studies of the occurrence of North Alabama LMA sources coincident with NLDN  $-$ CG return strokes. The panels correspond to three  $-$ CG flashes, showing the altitude of LMA sources as a function of time relative to the initial NLDN stroke. The times of the initial and subsequent strokes are displayed in blue and red, respectively.

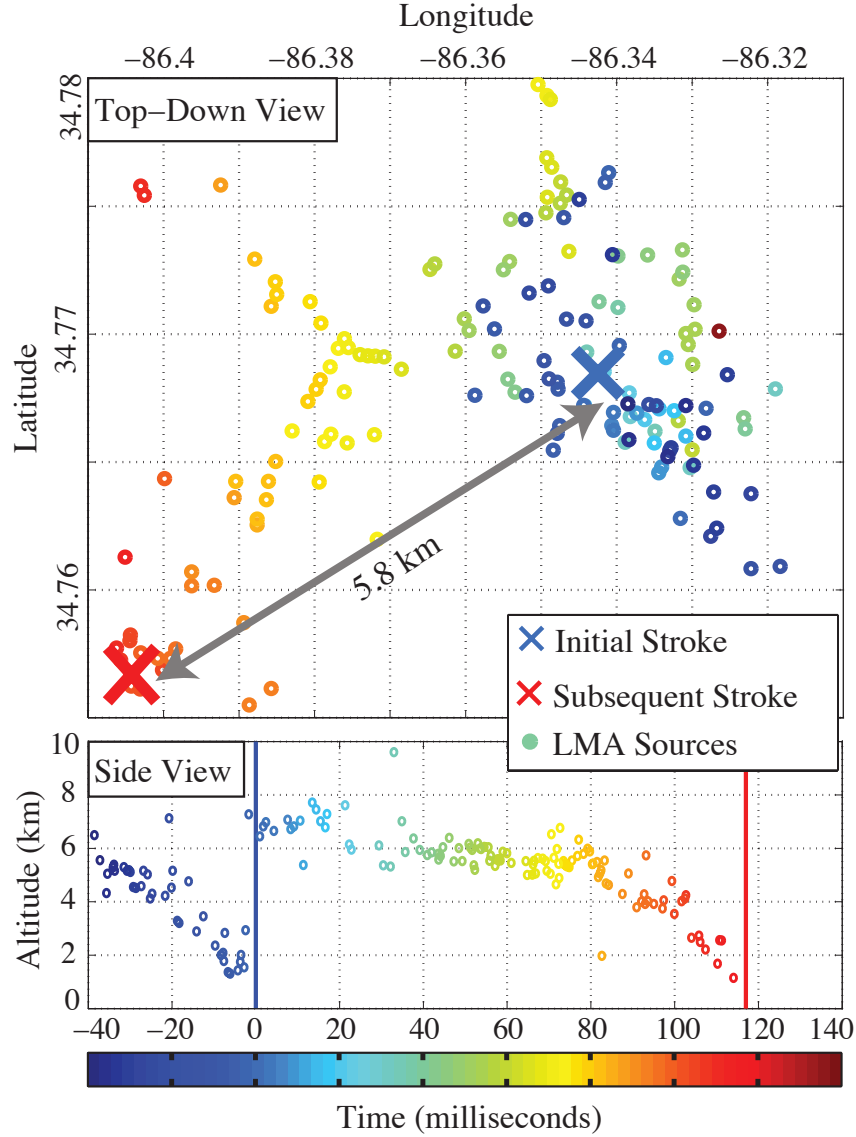


Figure 5.10: (Top) Top-down view of an NLDN –CG flash with overlaid LMA sources, color coded in time relative to the initial stroke (blue). (Bottom) Side-view of the same –CG flash, showing the altitude of sources as a function of time color-coded in time consistent with the top panel. The event occurred on July 12, 2011 at 19:44 UT.

NGC. We inspect 980 multi-channel NLDN –CG flashes (NGCs) that occur within 150 km from the center of NALMA from July 8–14, 2011. Roughly 230 flashes have LMA sources but only 34 are close enough to the sensors and have VHF sources (that propagate to ground) corresponding to the initial stepped leader. We find that half of these flashes (17 out of 34) exhibit a behavior similar to the one shown in Figure 5.10. The remaining cases do not show a clear (or any) connection between the two lightning channels.

In the following section, we explore this with much larger statistics with multiple years of LMA and NLDN data, to characterize the (horizontal) space-time distribution of sources around initial NLDN –CG strokes.

### 5.3.5 LMA Data around NLDN Negative CG Data

We apply the same technique to NLDN –CG stroke data and NALMA data from 2004 to 2010. We only select initial –CG events that occur within 200 km from the center of the NALMA array and that lead to subsequent NGCs. We compute the space-time distribution of in-cloud LMA sources following initial strokes, using LMA sources with altitudes above 4 km. Our aim is to extract the average behavior of sources inside the cloud throughout a multi-channel –CG. The occurrence histogram is computed using pairs of events (initial strokes, LMA sources) that occur within 10 km (horizontal distance) and 1 second time delay. We combine all initial –CG strokes reported by NLDN between 2004 and 2010 with the corresponding LMA sources above 4 km. The NLDN events occur within 200 km from the LMA center and have peak currents stronger than  $-15$  kA (previously discussed).

Figure 5.11 shows the occurrence rate of in-cloud LMA sources around initial –CGs as a function of time, parameterized by separation distance. We divide the region around the initial stroke into concentric regions with 1 km increments and compute the occurrence of LMA sources in each region as a function of time delay with a  $100 \mu\text{s}$  resolution, and display the corresponding 5-point moving average. We note that the majority of the LMA sources correspond to negative leaders because (i) negative leaders radiate stronger VHF than positive ones, and (ii) we focus our



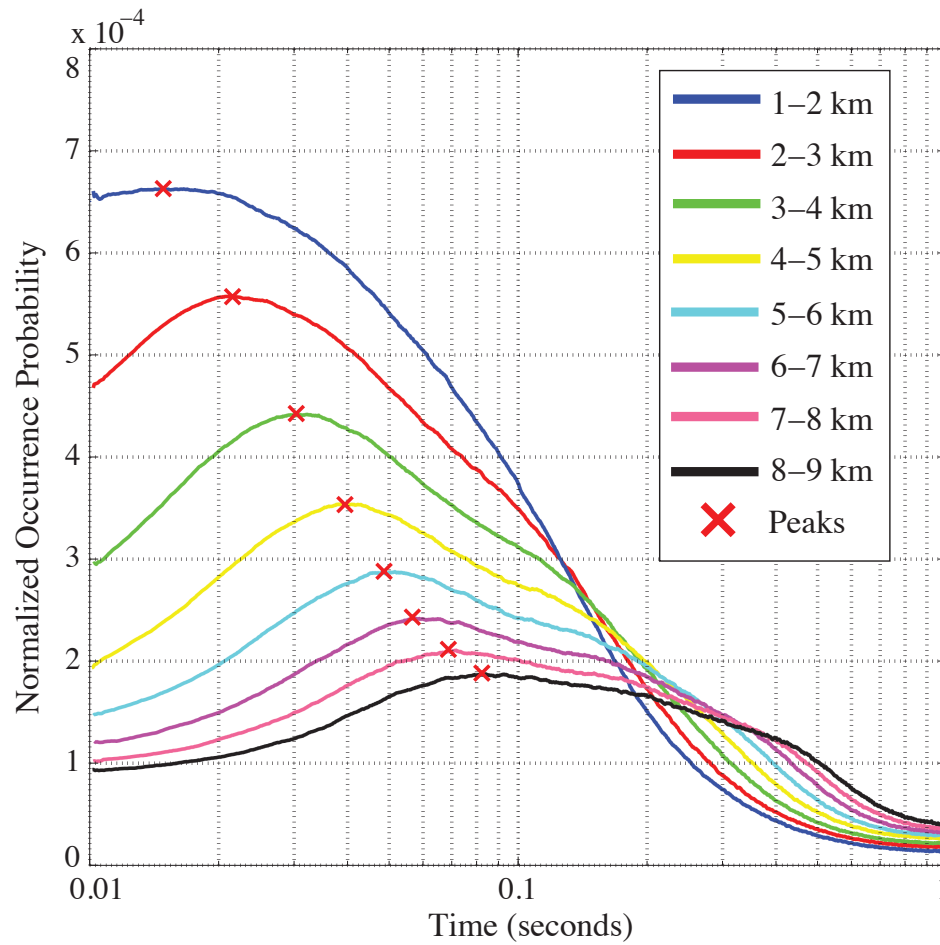


Figure 5.11: Temporal distribution of LMA sources around NLDN strokes, parameterized by distance, using NLDN and LMA data from 2004 to 2010. The curves are normalized to have unit area and are displayed on a logarithmic time scale.

analysis on  $-CGs$ . Since we are only interested in comparing the shapes of the curves, we normalize them to unit area (that is, we are not interested in the absolute values of the peaks). The time is displayed on a logarithmic scale to highlight features with a wide range of timescales. We use a method similar to the one presented in Equation (5.2) to find the peak occurrence rate for each curve and display it with a red X symbol.

The results in Figure 5.11 suggest that the distribution of the relative location of cloud sources propagates outward, with an expanding radial envelope. The rate of occurrence peaks in the 1–2 km region (in blue) after 14.9 ms, in the 4–5 km region (in yellow) after 39.4 ms, and in the 7–8 km region (in magenta) after 68.8 ms. The horizontal propagation of the cloud LMA occurrence distribution is in agreement with our case study observations and further support the mechanism proposed in section 5.3.3. The average extracted leader speed is  $\sim 110$  km/s, consistent with ground-truth measurements of cloud leader activity [Shao and Krehbiel, 1996; Akita et al., 2010]. The speed of the cloud leader is on the lower end of the 100–300 km/s range of stepped leader speeds [Rakov and Uman, 2007, p.123], and is comparable to the speed of the intermittent breakdown process in the early stage of the horizontal cloud discharge [Shao and Krehbiel, 1996; Akita et al., 2010]. We note that the LMA-extracted 110 km/s speed for cloud leader propagation is slower than the previously observed 200–280 km/s using NLDN return stroke data. There are several reasons that could explain the difference between the resulting speeds. We suggest two: (i) the faster dart-stepped leaders contribute to the NLDN spatio-temporal statistics but do not contribute to the LMA statistics as dart-stepped leaders do not generate cloud VHF sources, and (ii) the speed of the stepped leader increases as the leader approaches ground. Recent high-speed camera (1000 frames per second) measurements by Kong et al. [2007] showed that the speed of a single positive leader increase from 10 km/s to 380 km/s during its descent.

The LMA activity in the 1–2 km region (in blue) appears to have a normal distribution on a logarithmic timescale, which translates into a log-normal distribution (for linear time) of leader time of arrival. In the more distant 4–5 km region (in yellow), the time of arrival distribution and its peak shift (to the right) due to leader

propagation delay and the spread of the distribution increases. The 90–10% fall time roughly doubles from 279 ms at 1–2 km to 575 ms at 4–5 km. The larger spread at larger distances results from the randomness in leader growth and propagation, similar to the behavior of a random walk where the variance increases linearly with time. In a recent study, *Campos et al.* [2014] used video-recordings to compute the leader speeds and found that the speed of dart leaders follows a log-normal distribution. Our method could be extended (i) to empirically study the stepped leader speed distribution and (ii) to compute the change in the speed of the stepped leader as it approaches ground. These are important to stepped leader growth models but are beyond the scope of this paper.

The results of the multi-year statistical analysis support the proposed mechanism for the deterministic distribution evolution of new ground terminations. We repeat the same study for single stroke –CG flashes and find that cloud leaders also follow single stroke flashes, propagating horizontally in the cloud without approaching ground. Our findings provide new insight into inter-stroke lightning physics and are relevant to various lightning applications. The quantified space-time evolution and spatial reach of the –CG flash is of particular interest to those interested in lightning prediction and protection systems. The observations are also important for the correct interpretation of lightning breakdown data around –CG discharges.

# Chapter 6

## Analysis of Oceanic Peak Currents

As presented in Section 1.4, various statistical studies using NLDN and GLD360 geo-location data show that lightning peak current estimates are larger over the ocean than over land [*Tyahla and Lopez*, 1994; *Lyons et al.*, 1998; *Orville and Huffines*, 2001; *Bardo et al.*, 2004; *Cummins et al.*, 2005; *Orville et al.*, 2011; *Cooray and Rakov*, 2011; *Said et al.*, 2013; *Hutchins et al.*, 2013; *Cooray et al.*, 2013]. In this chapter, we conduct large-scale statistical analyses of NLDN and GLD360 data, focusing on peak currents in coastal regions for all stroke types (CG and IC), stroke numbers (initial and subsequent), and polarities (negative and positive). We combine NLDN and GLD360 data to quantify systematic biases in network-estimated peak currents over the ocean and present a novel approach to correct for the impact of range filtering on peak currents in deep oceanic regions, using non-linear least-squares optimization. In Section 6.3, we summarize and discuss our dissertation findings that relate to differences between ocean and land lightning.

### 6.1 Large-scale Peak Current Statistics

*Orville et al.* [2011] aggregated nine years of NLDN data between 2001 and 2009 and presented thematic maps of median peak currents of first and subsequent positive and negative CGs in the continental U.S. and along the eastern coastlines. In total, 267 million  $-$ CGs and 17 million  $+$ CGs were analyzed and the average peak currents were

found to be  $-16.1$  kA for  $-$ CGs (all stroke orders combined) and  $25.7$  kA for  $+$ CGs (after deletion of all events below  $15$  kA). The study also suggested that peak currents of initial  $-$ CGs are higher over the ocean than over land, with sharp transitions at the land-ocean boundaries. In contrast to initial  $-$ CGs, the sharp increase in peak currents at the land-ocean boundaries was absent for subsequent  $-$ CGs that followed a dart-leader (occur in the same channel as the previous stroke) and for  $+$ CGs.

In a previous and similar study, *Orville and Huffines* [2001] aggregated ten years of NLDN data (1989–1998) that suggested the same patterns. The fact that the sharp increase was limited to initial  $-$ CGs was taken as evidence that the jump in peak currents at the land-ocean boundary cannot be due to electromagnetic attenuation over lower ground conductivities (otherwise the increase would appear for all stroke types and polarities). Using NLDN data, *Cummins et al.* [2005] also showed that the land-ocean boundary produced a  $\sim 25\%$  increase in average peak currents over the ocean for initial  $-$ CGs. The magnitude of the increase did not correlate with ocean depth and the abrupt jump in peak currents was limited to initial  $-$ CGs, as noted by numerous other studies [*Tyahla and Lopez*, 1994; *Lyons et al.*, 1998; *Bardo et al.*, 2004; *Cooray and Rakov*, 2011; *Said et al.*, 2013; *Hutchins et al.*, 2013; *Cooray et al.*, 2013]. These studies agreed that the sharp transition in peak currents is only observed for initial  $-$ CGs suggesting that it is likely due to physical differences in the attachment process or in the initiation of  $-$ CGs over the ocean.

Using one year of GLD360 data with 353 million flashes, *Said et al.* [2013] showed that the geometric mean of peak currents for oceanic initial  $-$ CGs are 22%-88% higher compared to land  $-$ CGs, with sharp transitions in peak currents along the coastlines. In addition, *Said et al.* [2013] analyzed three coastal regions, marked using three boxes that are shown on the map of Figure 1.5, in which the peak current enhancement was extremely sharp at the land-ocean boundaries. Similar to the NLDN studies, the sharp transition in peak currents was missing for subsequent (same channel)  $-$ CGs and for  $+$ CGs.

The thematic maps presented by *Said et al.* [2013] (Figure 1.6) and others provided a great way to visualize the presence or absence of a sharp increase in peak currents along the coastlines. In this section, we take a closer look at the probability density

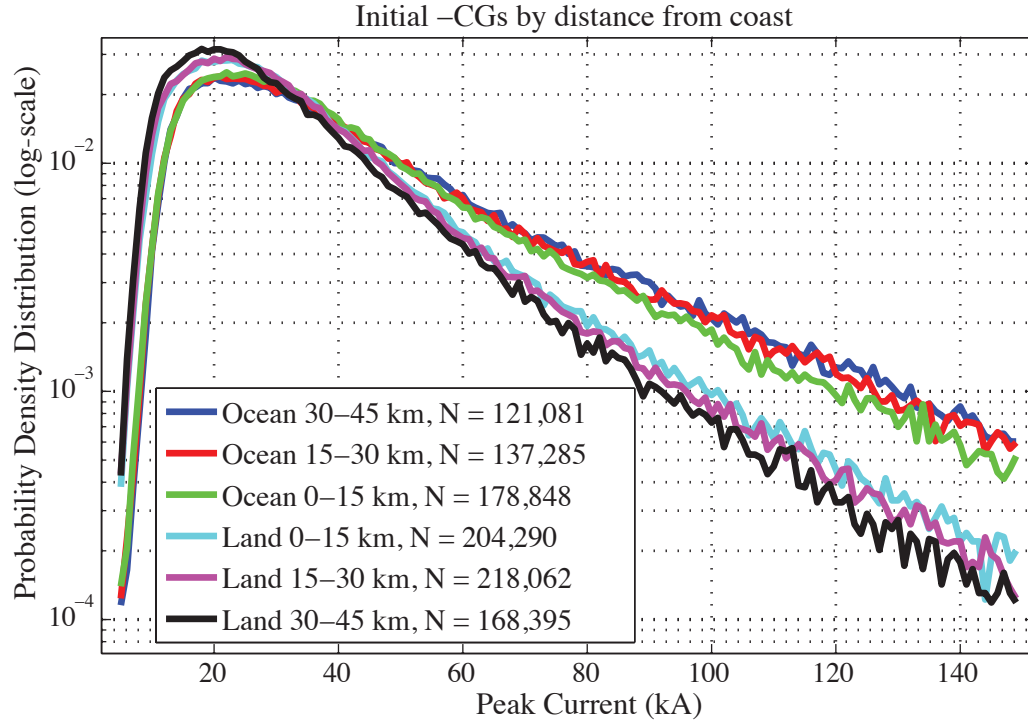


Figure 6.1: Distribution of NLDN peak currents of initial  $-CGs$  from July 2013 in the following regions: 30–45 km ocean (in blue), 15–30 km ocean (in red), 0–15 km ocean (in green), 0–15 km land (in cyan), 15–30 km land (in magenta), and 30–45 km land (in black)

distribution of peak currents in coastal regions for different stroke types, orders, and polarities. Unlike basic sample statistics, such as the mean/ median/ geometric mean, which represent the distribution with one value, we compute and compare probability density distributions of peak currents over ocean to those over land.

### 6.1.1 Application to NLDN CG Stroke Data

Figure 6.1 shows the distribution of peak currents of initial  $-CGs$  collected by NLDN in July 2013. The area along the eastern coast of the continental U.S. is divided into six swaths, according to distance to the coast. The first three regions are over the ocean and correspond to oceanic areas located 0–15 km, 15–30 km, and 30–45

km from the coastal boundary. The other three swaths correspond to land areas in similar distance to coast ranges. The NLDN coverage in these regions is mostly uniform [Cummins and Murphy, 2009]. We select all initial  $-CG$ s that occur in each region, compute the corresponding probability density distribution of peak currents, and display them as follows: 30–45 km ocean (in blue), 15–30 km ocean (in red), 0–15 km ocean (in green), 0–15 km land (in cyan), 15–30 km land (in magenta), and 30–45 km land (in black). Each probability density distributions is found by computing the peak current occurrence histogram in each region and normalizing it to unit area. The number of data points in each region are shown in the figure legend. We plot the probability distributions on a logarithmic scale to display a wide range of peak currents.

The distributions show a clear difference between ocean and land peak currents. The three distributions over the ocean (in blue, red, and green) are visually identical to each other, but different than the three distributions over land (in cyan, magenta, and black), which are visually identical to each other. The right tails of the oceanic distributions show a clear increase in the probability of higher peak current events over the ocean than over land. The geometric mean is  $-36.09$  kA in the 0–15 km oceanic region (in green; closest to the coast), roughly 20% higher than the geometric mean of  $-30.09$  kA in the 0–15 km land region (in cyan). Our results are consistent with the previously-mentioned studies, showing a sharp  $\sim 20\%$  increase in NLDN peak currents at the land-ocean boundary [Tyahla and Lopez, 1994; Lyons *et al.*, 1998; Orville and Huffines, 2001; Bardo *et al.*, 2004; Cummins *et al.*, 2005; Orville *et al.*, 2011; Cooray and Rakov, 2011; Said *et al.*, 2013; Hutchins *et al.*, 2013; Cooray *et al.*, 2013].

Figure 6.2 shows the statistical distribution of peak currents of subsequent  $-CG$  strokes that follow a dart leader, recurring in an existing channel. We follow the same approach to compute the probability distributions of NLDN peak currents in each region. Interestingly, the distributions show a consistent difference between ocean and land peak currents, though smaller than in the case of initial  $-CG$ s (Figure 6.1). The three distributions over the ocean (in blue, red, and green) are visually identical to each other, but slightly different than the three distributions over land (in cyan,

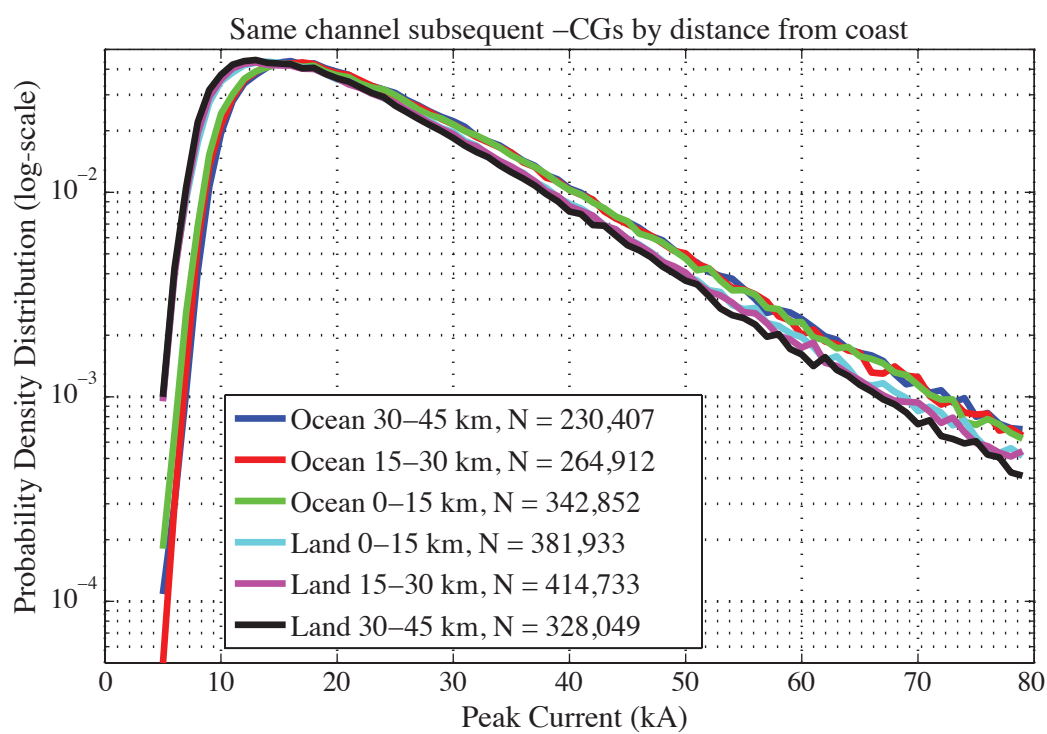


Figure 6.2: Same as Figure 6.1 but for same-channel subsequent –CGs.



magenta, and black). The geometric mean is  $-23.1$  kA in the 0–15 km oceanic region (in green; closest to the coast), roughly 10% higher than the geometric mean of  $-21.1$  kA in the 0–15 km land region (in cyan). Although the magnitude of the increase in the geometric mean is smaller (roughly half) than the increase in the case of initial  $-CG$ s, it does appear to be abrupt at the land-ocean boundary. Visually, this can be seen as the shapes of the distributions vary abruptly (and not gradually) as we transition from ocean to land regions. The abrupt increase is reflected in the geometric means of the six distributions:  $-23.7$  kA (30–45 km ocean; in blue),  $-23.4$  kA (15–30 km ocean; in red),  $-23.1$  kA (0–15 km ocean; in green),  $-21.1$  kA (0–15 km land; in cyan),  $-20.9$  kA (15–30 km land; in magenta), and  $-20.4$  kA (30–45 km land; in black).

Next, we apply a bootstrapping approach to quantify the statistical significance of the differences in the calculated geometric means (of same-channel  $-CG$  peak currents) in the 0–15 km land and 0–15 km oceanic regions. The bootstrapping technique, initially introduced by [Efron \[1979\]](#), provides a way to construct confidence intervals for an estimated parameter. In our case, we aim to compute confidence intervals around the estimated geometric mean (of peak currents) in each swath. In the  $i^{th}$  swath,  $N_i$  lightning discharges occur. The estimated geometric mean of peak currents in the  $i^{th}$  swath is computed using the  $N_i$  peak current observations. Then,  $N_i$  random data points are chosen with replacement from the  $N_i$  peak current measurements. Sampling with replacement means that a particular data point could appear multiple times. The geometric mean of the  $N_i$  randomly chosen samples is then computed. The procedure (sampling of the original set with replacement) is repeated  $M$  times resulting in  $M$  estimates of the geometric mean. The distribution of the  $M$  geometric means is used to obtain 95% confidence intervals and a measure of uncertainty of the geometric mean in each swath. The main advantage of bootstrapping is that it provides confidence intervals of the estimated quantity without prior knowledge of the shape of the underlying probability distribution [[Efron, 1979](#)].

We apply the bootstrapping technique to find the confidence interval corresponding to the geometric means in the 0–15 km oceanic region ( $N = 343,000$  and  $M = 100$ ) and 0–15 km land region ( $N = 382,000$  and  $M = 100$ ). We find that the 95%

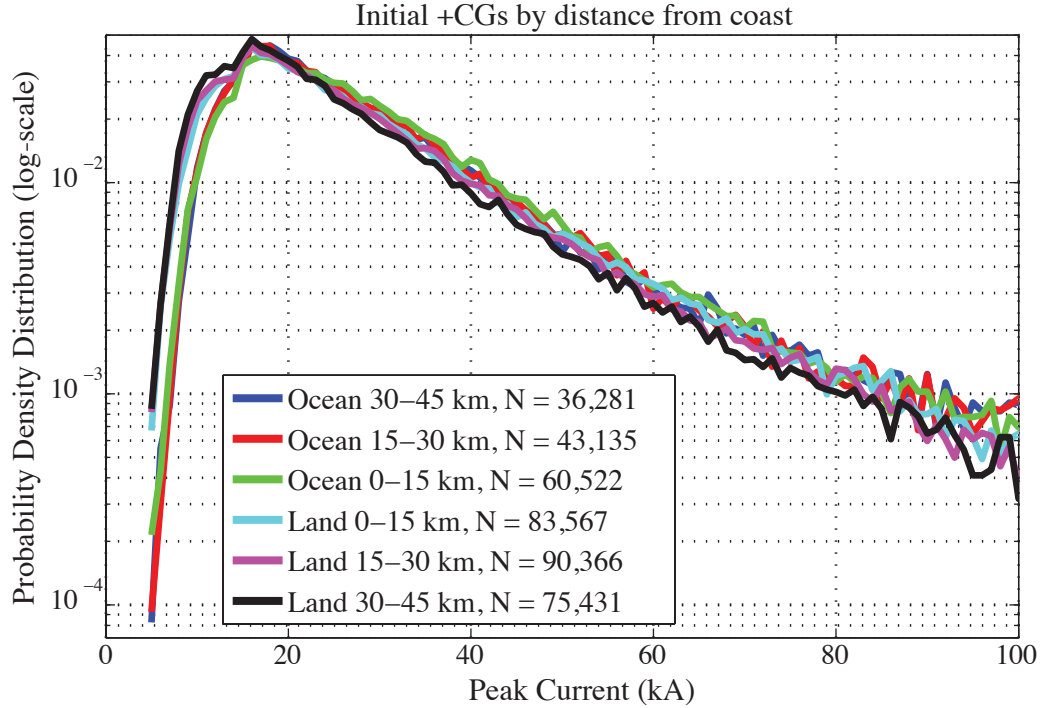


Figure 6.3: Same as Figure 6.1 but for initial +CGs.

confidence intervals are  $[-23.0, -23.3]$  kA and  $[-21.0, -21.3]$  kA, respectively. The confidence intervals show that the difference between the ocean and land geometric means is statistically significant, reinforcing the visual evidence that the increase in peak currents of (dart-leader) subsequent  $-$ CGs is abrupt along coastlines. The small (but abrupt) increase in same-channel subsequent  $-$ CGs peak current was not brought out in previous studies [*Tyahla and Lopez, 1994; Lyons et al., 1998; Orville and Huffines, 2001; Bardo et al., 2004; Cummins et al., 2005; Orville et al., 2011; Cooray and Rakov, 2011; Said et al., 2013; Hutchins et al., 2013; Cooray et al., 2013*]. Similarly to (same-channel) subsequent  $-$ CGs, these studies do not find an abrupt increase in peak currents for initial +CGs along land-ocean boundaries, which we inspect next.

Figure 6.3 shows the statistical distribution of peak currents of initial +CG strokes.

Again, the probability distributions are found after computing the peak current occurrence histogram and normalizing it to unit area. We find that the geometric mean is 27.1 kA in the 0–15 km oceanic region (in green; closest to the coast), roughly 9% higher than the geometric mean of 24.8 kA in the 0–15 km land region (in cyan). Interestingly, our results also show a consistent difference, though smaller than in the case of initial –CGs, between the ocean and land distributions. The three peak current distributions over the ocean (in blue, red, and green) are visually identical to each other, but slightly different (skewed to the right towards higher peak currents) than the three land distributions of peak currents (in cyan, magenta, and black). We find the 95% confidence intervals, using bootstrapping, to be [26.9, 27.1] kA and [24.7, 24.9] kA in the 0–15 km oceanic region ( $N = 61,000$  and  $M = 100$ ) and 0–15 km land region ( $N = 84,000$  and  $M = 100$ ), respectively. This suggests that the difference between the distribution of +CG peak currents is statistically significant over land than over ocean, with abrupt (but smaller) jump at the land-ocean boundaries.

We note that in previous studies, positive CGs with peak currents below 15 kA were not included in the comparison of oceanic and land +CGs. This filtering of weak events, which, as shown here, are more frequent over land, leads to more similar ocean and land peak current distributions and washes out the jump in peak currents at the land-ocean boundary. For instance, we repeat our analysis after removing all +CGs with peak currents smaller than 15 kA. We find that the increase in peak currents at the boundary drops to 3%, which explains the absence of positive CG peak current increase in previous studies. However, the analysis of the complete peak current distribution and the inclusion of weak events in neighboring land and oceanic regions (0–15 km land and ocean) allows us to compare the proportion of weak and strong events over land and over the ocean and provides a more complete picture of the increase in peak currents at the land-ocean boundary.

The two results from Figure 6.2–6.3 draw an intriguing statistical picture of oceanic lightning, showing that the increase in peak currents is abrupt (though to different extent) across stroke orders and polarities. This suggests that the peak current increase at the land-ocean boundary is not limited to initial –CGs and might not result from differences in the initiation or attachment processes of initial

–CGs, which are the leading physical explanations in the lightning community.

So far, our analysis and previous NLDN studies have focused on statistics of CGs, which, unlike ICs, are efficiently detected by the network (refer to Section 1.3.2). In this dissertation, we take advantage of a recent (mid 2013) software upgrade, that substantially improved the NLDN IC detection rate, to compute the first statistical distributions of the effective peak currents of negative ICs (–IC) and positive ICs (+IC) at the land-ocean boundary. We use the term effective as intra-cloud lightning does not attach to ground. Therefore the reported values do not estimate the peak current on the ground but provide an estimate of the reference radiated fields to those from CG lightning.

### 6.1.2 Application to NLDN IC Stroke Data

Figure 6.4 consists of two panels that show the statistical distribution of the effective peak currents of –ICs (top) and +ICs (bottom). Again, the probability distributions are found after computing the peak current occurrence histogram and normalizing it to unit area. In the top panel, the distributions correspond to –ICs. The geometric mean is –20.4 kA in the 0–15 km oceanic region (in green; closest to the coast), roughly 17% higher than the geometric mean of –17.4 kA in the 0–15 km land region (in cyan). In the bottom panel, the peak current distributions correspond to +ICs. The geometric mean is 16.9 kA in the 0–15 km oceanic region (in green; closest to the coast), roughly 22% higher than the geometric mean of 13.9 kA in the 0–15 km land region (in cyan)

The distributions in Figure 6.4 show a consistent and abrupt  $\sim 20\%$  increase in effective peak currents of ICs over ocean than over land. These results provide the first statistics of effective peak currents of ICs at the land-ocean boundary, suggesting that the increase in radiated fields is not limited to CGs, but also extends to intra-cloud activity. These findings are of great practical interest to those concerned with the safety of aviation and, more importantly, provide the first evidence that the increase in oceanic peak currents is not necessarily related to differences in the attachment process or specific to the cloud-to-ground channel.

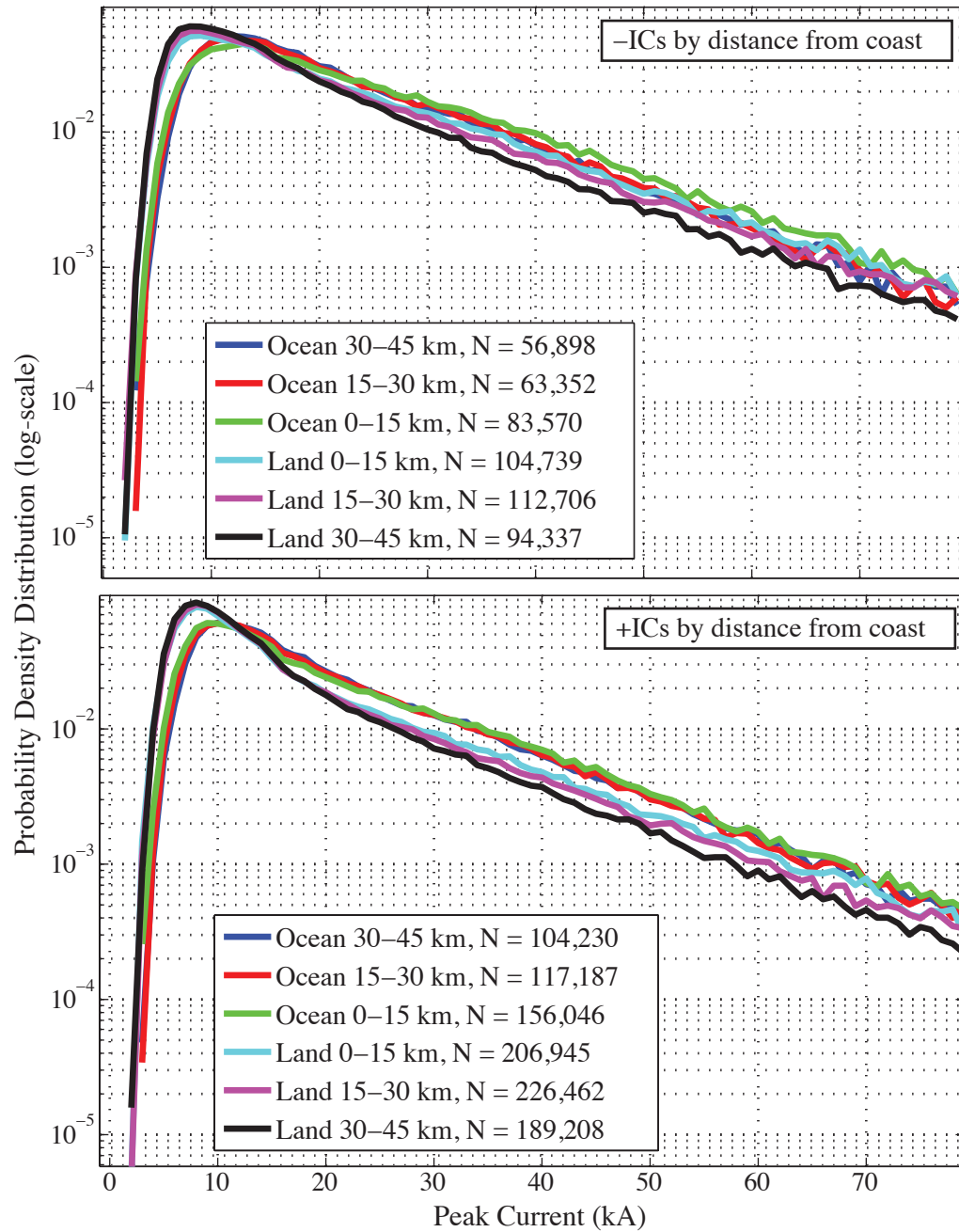


Figure 6.4: Same as Figure 6.1 but for -ICs (top) and +ICs (bottom).

	Land (kA)		Ocean (kA)		Ocean/Land
	15–30 km	0–15 km	0–15 km	15–30 km	increase
–CG init.	–29.3	–30.1	–36.1	–37.4	+19.9%
–CG subs.	–20.9	–21.1	–23.1	–23.4	+9.6%
+CG init.	+24.0	+24.8	+27.1	+26.4	+9.3%
–IC	–16.4	–17.4	–20.4	–19.5	+17.1%
+IC	+13.4	+13.9	+16.9	+17.1	+21.9%

Table 6.1: Summary of geometric mean statistics from Figure 6.1–6.4. The last column shows the percentage increase between the 0–15 km oceanic region and the 0–15 km land region.

Table 6.1.2 summarizes the results from Figure 6.1–6.4. The geometric mean of peak currents is computed in each of the four regions closest to the land-ocean boundaries. The percentage increase in peak current at the coastlines (0–15 km ocean to 0–15 km land) is reported in the last column. The results show a consistent 9–22% jump in peak currents at the coastlines for all categories. The magnitude of the increase varies with stroke type, order, and polarity. The consistent increase in peak currents across ICs and CGs suggest that the increase at the land-ocean boundaries could be due to meteorological differences in land and ocean storm dynamics and/or convective activity, leading to stronger lightning discharges over the ocean (discussed further in Section 6.3).

## 6.2 Systematic Network Errors

Typically, electromagnetic geo-location networks, such as NLDN and GLD360, suffer from range filtering, due to the fact that weak events far from sensors are likely to be missed. This artifact biases the distribution of peak currents and leads to overestimates of the corresponding mean, median, or geometric mean. The impact of range filtering is highest deep in oceanic regions. As discussed in Section 1.3, the range of the NLDN LF sensors is limited to few hundred kilometers. Thus, NLDN only provides coverage in oceanic regions that are in the vicinity of the eastern coast of

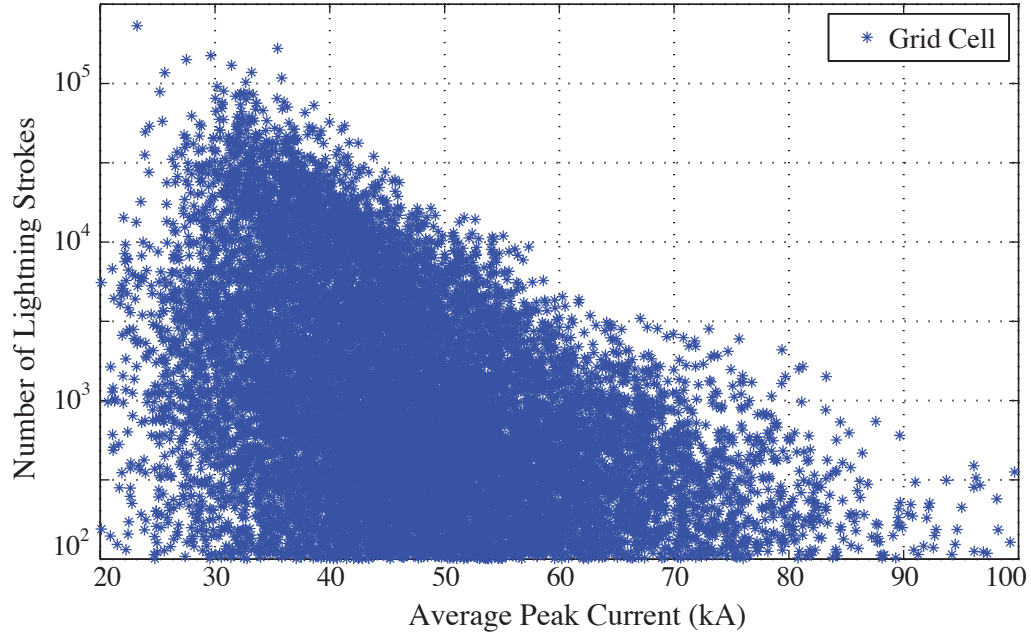


Figure 6.5: Occurrence and average peak current of GLD360 initial  $-CGs$  in oceanic grid cells (2013).

the continental United States. The GLD360 VLF sensors have global reach, providing coverage in deep oceanic regions. However, as seen in Figure 1.6 (adapted from [Said et al. \[2013\]](#)), the geometric mean of GLD360 peak currents is highest in remote oceanic regions. [Said et al. \[2013\]](#) admitted that the fairly recent network could be overestimating peak currents in these regions due to range filtering and/or due to errors in modeling VLF propagation.

### 6.2.1 Range Filtering and Correction

We conduct a statistical study using 48 million oceanic initial  $-CGs$  captured by GLD360 in 2013 to bring out the effect of range filtering. We divide the oceanic regions into  $1^\circ \times 1^\circ$  latitude-longitude cells. In each cell, we compute the total number of initial  $-CGs$  that occurred and the average of their peak currents. Figure 6.5 consists of a scatter plot showing the occurrence of initial  $-CGs$  as a function of the

average peak currents in each cell. Each scatter point corresponds to one grid cell. We only include cells that have more than 100 events and plot the occurrence on a logarithmic scale. The scatter plot shows a clear range filtering pattern. The grid cells that have larger peak current averages tend to have fewer events. In these cells, only the strong events are captured by GLD360 and weaker ones are missed, leading to inaccurate high peak currents and low occurrence values.

Using the same GLD360 data set, we compute the geometric mean of initial  $-CG$ s in two regions, corresponding to 500–515 km from the coast, for ocean and for land, respectively. We find that the geometric mean is  $-43.0$  kA in the first region (ocean; 500–515 km) and  $-25.3$  kA in the second (land; 500–515 km). The 70% increase in the geometric mean of peak currents in deep oceanic regions is much higher than the 20% jump that is found using coastal NLDN data (Table 6.1.2). The difference in magnitude could be either due to network inaccuracies, such as range filtering, or due to stronger lightning in deep oceanic regions.

Typically, the distribution of peak current of  $-CG$  lightning follows a log-normal distribution. The log-normal distribution is a probability distribution of a random variable, whose logarithm follows a normal distribution. The log-normal is often denoted by  $\ln N(\mu, \sigma^2)$ , where  $\mu$  and  $\sigma^2$  are the mean and variance of the normal distribution, respectively. The expression of its probability density function is given by Equation 6.1. Generally, a variable follows a log-normal distribution if it could be thought of as the multiplicative product of independent positive random variables. Experimental evidence show that the distribution of initial  $-CG$  peak currents follow a log-normal distribution [Berger et al., 1975].

$$f(x; \mu, \sigma^2) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}} \quad (6.1)$$

In this dissertation, we develop a novel approach that uses weighted non-linear least-squares optimization to correct for the peak current distribution in remote regions which suffer from range filtering. We assume that the true peak current distribution follows a log-normal shape. We aim to best-fit the observed peak current distribution with a log-normal shape. The optimization problem is formulated in Equation



6.2, where  $f$  is the probability density function of the log-normal distribution,  $B$  is the number of histogram bins,  $x_i$  denotes the peak current value corresponding to the  $i^{th}$  bin,  $H_i$  denotes the frequency of occurrence of  $x_i$  (value of the  $i^{th}$  histogram bin),  $\mathbf{w}$  is a vector of non-negative weights, and  $\mu$ ,  $\sigma$ ,  $s$  are the optimization parameters.

$$\underset{\mu, \sigma, s}{\operatorname{argmin}} \xi(\mu, \sigma, s) = \sum_{i=1}^M w_i \|H_i - s f(x_i; \mu, \sigma)\|^2 \quad (6.2)$$

The first two optimization parameters ( $\mu$ ,  $\sigma$ ) characterize the best-fit log-normal distribution and the scalar  $s$  accounts for linear scaling of its shape to fit the data. We apply the Levenberg-Marquardt method (Section 4.2.2) to solve for the optimal values of  $\mu$ ,  $\sigma$ , and  $s$ , by minimizing the mean squared error between the original distribution and the best-fit log-normal shape. The magnitude of the weight  $w_i$  associated with each data point indicates the precision of the information contained in the  $i^{th}$  bin. The weights determine the contribution of each bin to the best-fit log-normal shape. In a generic least-squares (LS) optimization, all the weights are chosen to equal 1. In a weighted least-square (WLS) problem, the weights are chosen in a way that the right side of the peak current distribution (which corresponds to high peak currents) contributes more to the optimal log-normal shape, minimizing the contributions of the left portion of the distribution which suffers from range filtering. Once the optimization problem is solved, we can find the corrected geometric mean of the (peak current) distribution. The geometric mean and the median of a log-normal random variable both equal to  $e^\mu$ .

Figure 6.6 shows the observed probability distribution of GLD360 –CG peak currents (in blue) in the 500–515 km oceanic region. The distribution is calculated by normalizing the histogram of peak currents in that region to unit area. The peak currents are shown on a logarithmic scale to display a wide range of peak currents and, more importantly, to visually bring out the range filtering effect. As mentioned previously, the logarithm of a log-normal random variable follows a normal distribution. Therefore, in the absence of range filtering, the distribution of –CG peak currents (in blue) should have a symmetric shape around its median (which is also the geometric mean). In our case, the asymmetry is visually clear as the left portion

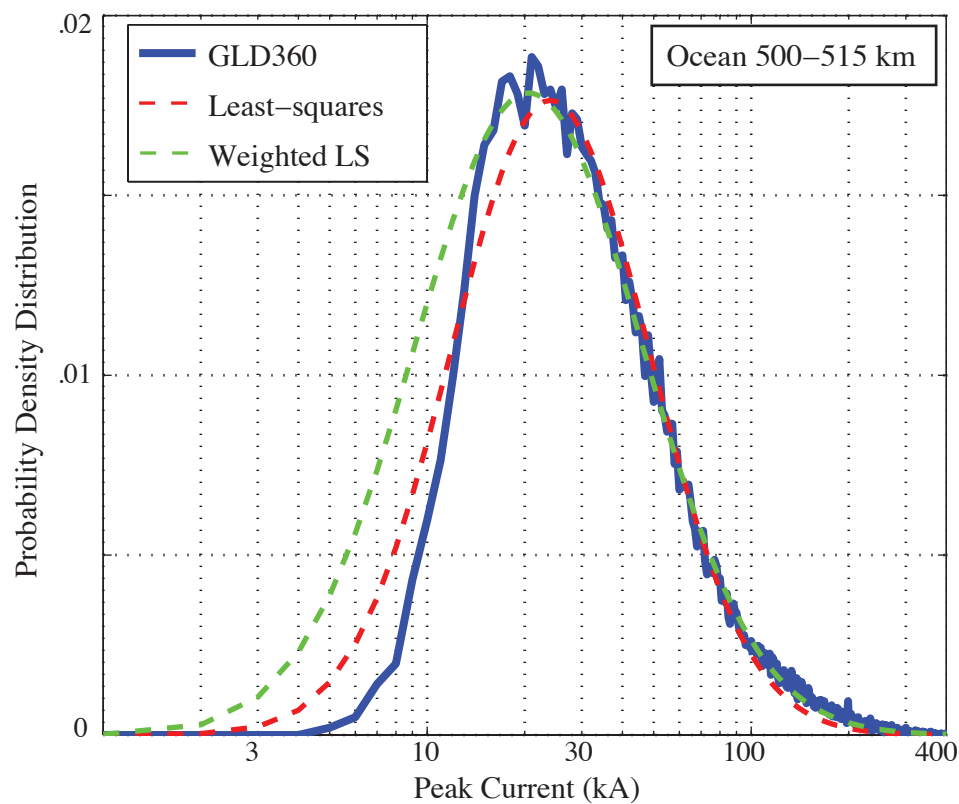


Figure 6.6: Original (blue) and corrected (dashed green, dashed red) peak current distributions of initial –CGs reported by GLD360 in the 500–515 km oceanic region (2013).

of the distribution (weak events) drops abruptly, suffering from range filtering. We apply both the LS and WLS (weights are 0 for peak currents below 20 kA, 1 for 20–60 kA, and 5 for >60 kA) correction techniques to the original distribution. We apply both as together they can provide a range of corrected values and a measure of sensitivity to the choice of weights. The best-fit distributions corresponding to the LS solution (in dashed red) and to the WLS solution (in dashed green) are displayed in Figure 6.6. We find that the corrected geometric means drop to  $-23.97$  kA in the WLS case and to  $-20.54$  kA in the LS case. The geometric mean of the original distribution is  $-43.0$  kA.

We repeat the same procedure in the 500–515 km land region. We find that the original geometric mean of  $-25.24$  kA drops to  $-17.40$  kA and  $-15.96$  kA following the WLS and LS corrections, respectively. The drop in the mean is expected as GLD360 is also likely to miss a portion of weaker events over land. In all, the original increase in the geometric mean of peak currents in the 500–515 km oceanic region compared to the 500–515 km land region is approximately 70–75%. The magnitude of the increase drops to 29% using the WLS correction and to 38% using the LS correction. Thus, range filtering is significantly biasing the observed oceanic peak currents, leading to the overestimation of the mean of peak currents and to the overestimation of the magnitude of peak current increase (compared to land) in deep oceanic regions. Nevertheless, ocean peak currents in these regions remain 30–40% higher than their land counterparts, even after accounting for range filtering.

### 6.2.2 Peak Current Estimation

As presented in Section 1.3.2, NLDN is a well established network and its peak currents have been tested against ground-truth current measurements for  $-CGs$  in triggered lightning experiments at Camp Blanding [Cummins and Murphy, 2009; Nag et al., 2011; Turman et al., 2014]. In turn, peak currents of the more recent GLD360 network have been calibrated using ones reported by NLDN [Said et al., 2010, 2013]. In this section, we match one year (2013) of GLD360 and NLDN events that occur in the continental United States and in neighboring coastal regions and compare the

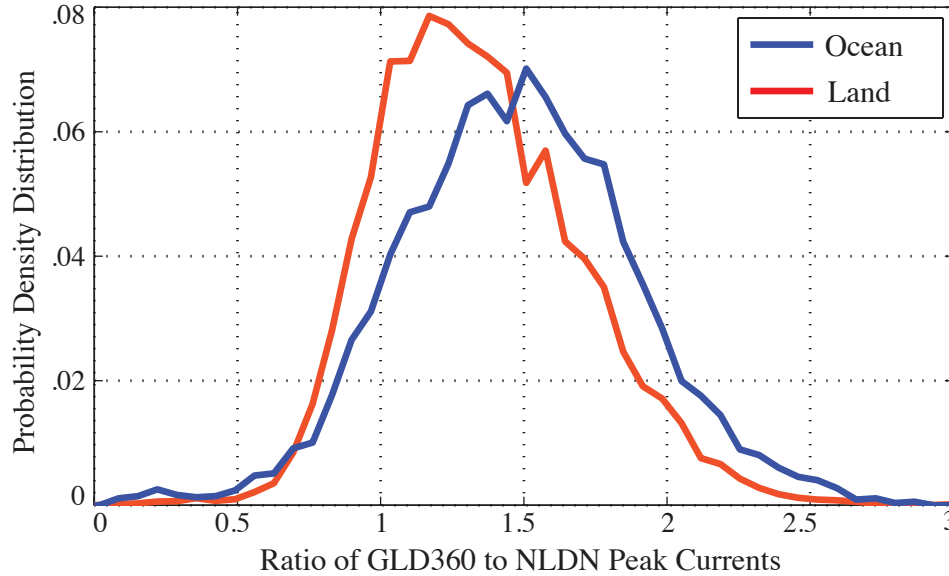


Figure 6.7: Distribution of the ratio of GLD360 to NLDN peak currents over ocean (in blue) and over land (in red).

estimates of the two networks.

Figure 6.7 shows the distribution of the ratio of GLD360 to NLDN peak currents over ocean (in blue) and over land (in red). Ideally, in case the GLD360 and NLDN are aligned, the two distributions would be centered around one with a small spread. However, we find that GLD360 consistently over-estimates peak currents both over land and over ocean. The distributions are centered around 137% for land –CGs and around 153% for ocean –CGs. More importantly, the GLD360 over-estimation is 10–15% more pronounced over the ocean than over land (obtained by taking the ratio of 153% and 137%). The additional over-estimation of peak currents over the ocean could thus lead to an artificial 10–15% increase in GLD360 peak currents in oceanic lightning (compared to land lightning).

The systematic inaccuracy in the over-estimation of oceanic peak currents could be contributing to the results from [Said et al. \[2013\]](#) and to our results from Section 6.2.1. In fact, at the time of this writing, the operators of GLD360 validated these systematic biases. In addition, testing of new GLD360 algorithms produced peak currents that

<i>Compared to Land</i>	Original	Post-Range Filtering	Post-Calibration
Ocean 500–515 km	70%	30–40%	15–25%

Table 6.2: Original and post-correction increase in geometric mean in GLD360 peak currents in the 500–515 km oceanic region, compared to land.

better match those of NLDN, both over land and ocean. Table 6.2.2 summarizes the impact of network inaccuracies on the magnitude of the geometric mean increase in deep oceanic regions (500–515 km) compared to land regions (500–515 km). The original distributions show a 70% increase in peak currents. The increase drops to 30–40% following the range-filtering correction and to 15–25% after accounting for the additional GLD360 over-estimation. Interestingly, the corrected 15–25% increase of GLD360-reported peak currents in (500–515 km) oceanic region is comparable to the 20% increase in coastal regions that we observed in Section 6.1 using NLDN data.

### 6.3 Discussion of Ocean Lightning Findings

In this dissertation, we conduct various studies that offer new insight into the physics of lightning. In this section, we discuss and summarize our findings that relate to differences and similarities between ocean and land lightning.

As introduced in Section 1.4, multiple geo-location networks estimate lightning peak currents using the range-normalized electromagnetic radiation. Earlier statistical studies of hundreds of millions of discharges suggested that peak currents of initial  $-CGs$  are higher over the ocean than land, with sharp transitions at the land-ocean boundaries. The sharp transition, however, was absent for initial  $+CGs$  and same-channel subsequent  $-CGs$ , suggesting that the increase in peak currents came from physical differences in the attachment process to salt water or in the initiation of initial  $-CGs$  over the ocean. We do not find this to be true.

The conductivity of salt water (5 S/m) is three orders of magnitudes higher than the conductivity of land ( $\sim 2$  mS/m). On average, electromagnetic fields from oceanic lightning propagate to sensors over higher conductivity ground than their land counterparts. If not properly accounted for, differences in electromagnetic attenuation due

to the lower land conductivity could lead to a systematic under-estimation of land peak currents. In Section 3.3.2, we show that the increase in ocean peak currents is not due to errors in modeling electromagnetic propagation, radiation, or reflections. In addition, our statistical studies from Section 6.1 find that the increase in peak currents is sharp at the coastlines, which in tandem with our modeling results, strongly suggest that the observed increase in peak currents cannot be explained by differences in the electric properties of the propagation path.

There is also a possibility that radiation fields are enhanced over the ocean without increasing the magnitude of the lightning currents inside the channel. At the lightning source, faster return stroke speeds or shorter channel-base current rise-time could affect the radiation patterns, resulting from differences in the attachment process to salt water. In Section 3.4, we use our gwFWM forward model to visualize the impact of changes in the return stroke speed or the current rise-time on the lightning waveforms. In Section 3.4.1, we combine the modeling efforts with experimental ocean and land LF waveforms to show that land and ocean  $-CGs$  have similar current rise-times. We follow the same approach in Section 3.4.3 to show that they also have similar return stroke speeds. In Section 4.5, we apply our inverse model to estimate the distribution of rise-times and return stroke speeds for ocean and land lightning. The distributions additionally suggest that current rise-times and return stroke speeds in ocean and land lightning are similar.

In Section 6.1, we conduct large statistical studies of NLDN peak currents at the land-ocean boundaries. We show a consistent 9-22% jump in peak currents at the coastlines for initial  $-CGs$ , same-channel subsequent  $-CGs$ , initial  $+CGs$ ,  $-ICs$ , and  $+ICs$ . This is the first analysis that shows that the increase in peak currents extends to  $ICs$  and all types of  $CGs$ . Although the magnitude of the peak current increase varies with stroke polarities, orders, and types, our results suggest that the increase is not just limited to initial  $-CGs$  and is not limited to differences in physical process of initial  $-CGs$ . Additionally, the results suggest that the stronger radiation cannot be due to physical differences in the attachment process to higher conductivity salt water ( $ICs$  do not contact ground) and thus cannot be due to differences in current rise-time or return stroke speed.

In Section 4.5, we use our inverse model to estimate peak currents of experimental ocean LF waveforms and compare them to their land counterparts. We find that the median and geometric mean of initial  $-CGs$  are 20–27% higher over the ocean. Interestingly, the magnitude of increase is consistent with the one measured in Section 6.1 at the coastlines using NLDN. However, it is much lower than the one reported by GLD360 in deep oceanic regions [Said *et al.*, 2013]. In Section 6.2, we show that the 70% increase in GLD360 peak currents in regions deeper in ocean (500–515 km) is mostly due to range filtering and to inaccuracies. We apply a novel statistical technique to correct for these inaccuracies. We find that the corrected GLD360 peak current increase is about 15–25%, closer to the increase in coastal regions and consistent with our LF inversion results. This shows that the magnitude of the increase of lightning currents in deep oceanic regions is similar to the increase in coastal regions. Although the increase is not as high as previously reported, oceanic lightning peak currents remains 10–25% higher than land peak currents.

In Section 5.1 we develop a novel statistical technique to remote sense the charging and discharging processes of thunderstorms and apply it in Section 5.2.5 to show that the charging and discharging of land and ocean thunderstorms appear to be different.

In all, our findings suggest that electric currents are on average (geometric mean) 10%–30% stronger in oceanic than in land lightning. This increase is consistent in coastal and in deep oceanic regions and is across all stroke types, polarities, and orders. Our analyses suggest that the stronger oceanic currents could either result from lightning initiation differences (across all strokes types and polarities) or from meteorological differences (charging and discharging processes) between oceanic and land thunderstorms, leading to more frequent lightning over land but stronger lightning over the ocean. In Section 7.1 we propose future experiments that can further our understanding of the physics behind these observations.

## Chapter 7

# Summary and Suggestions for Future Work

Natural lightning is fascinating and frequent, with on average 3–4 million discharges occurring around the globe every day. To date, the physical processes behind this phenomenon are not fully understood, primarily due to the difficulty of obtaining direct measurements inside thunderstorms or inside the lightning channel. In this dissertation, we used radio remote sensing in tandem with statistical techniques, that leverage the frequent occurrence of lightning, to investigate the physics of the lightning discharge. We deployed a sensitive LF radio receiver system aboard the NOAA Ronald W. Brown research vessel to collect magnetic field emissions close to deep-ocean lightning. We processed thousands of ocean and land waveforms to study poorly-understood differences and similarities between ocean and land lightning. We analyzed millions of VLF, LF, and VHF geo-location data points to find statistical patterns that offer new insight into the physics of lightning and to propose new discharge mechanisms.

In Chapter 1, we introduced the lightning discharge and the lightning radiation spectrum. We reviewed lightning geo-location networks, listing advantages and limitations of VHF, LF, and VLF sensors. We summarized previous studies that draw an intriguing (but poorly understood) picture of stronger lightning over the ocean than over land and presented potential explanations of these observations.



In Chapter 2, we introduced the electric properties of Earth, the ionosphere, and the modeling of electromagnetic propagation of lightning radiation. We presented a recently-developed LF receiver system and a novel ship-borne experiment, which started in February, 2013 in San Juan, Puerto Rico, and ended in May, 2014 in San Francisco, California. Throughout the experiment, the system locally stored permanent broadband LF data around tens of thousands of deep-oceanic return strokes within a 750 kilometer radius from the ship.

In Chapter 3, we reviewed the modeling of return stroke currents and constructed an efficient ground wave electromagnetic model (gwFWM), which accounts for radiation and propagation physics. We validated our model against analytical approximations of ground wave propagation from [Cooray \[1987\]](#) and against a time-domain numerical model introduced by [Marshall \[2012\]](#). We conducted sensitivity analyses to study the impact of lightning current parameters on the waveform at the receiver. We showed that channel-base current rise-time and return stroke speed have measurable effects on the waveform and found that current rise-times and return stroke speeds of ocean and land lightning are similar, using statistical studies of the collected LF waveforms.

In Chapter 4, we constructed an inverse model of the gwFWM model, in which we used the lightning waveform to infer current parameters at the lightning source. We developed a system of neural networks that successfully solved for the inverse problem and quantified the accuracy of the model, using simulation waveforms. We used the inverse model to estimate the distribution of channel-base current rise-times, return stroke speeds, and peak currents of land and ocean lightning. We found that land and ocean discharges have similar rise-times and return stroke speeds, in agreement with speeds measured optically. We also found that negative cloud-to-ground lightning peak currents were on average 20%-27% higher over the ocean than land.

In Chapter 5, we studied statistical patterns in the location of lightning, by analyzing and aggregating millions of NLDN, GLD360, and NALMA data. We developed a novel statistical technique to measure the charging and discharging of thunderstorms and the impact of a lightning discharge on the surrounding electric field, using geo-location data. We showed that higher peak current discharges lead to longer electric

field recoveries and that charging and discharging processes were different in ocean and land thunderstorms. We studied patterns between initial and subsequent strokes and proposed a new discharge mechanism related to subsequent cloud-to-ground discharges. Following the CG stroke, we showed that in-cloud leaders could form at the tip of the lightning channel, propagating horizontally in the cloud. The cloud leader could then grow into a second cloud-to-ground stepped leader, producing a new ground termination several kilometers away from the previously existing one.

In Chapter 6, we analyzed NLDN and GLD360 peak currents in regions along the coastlines and others deep in the ocean. We proposed and applied a statistical technique based on weighted non-linear squares optimization to correct for the impact of range filtering. We presented results that suggested that lightning strokes across all types, orders, and polarities tend to be stronger over the ocean, with sharp transitions in peak currents along coastlines. We found that the increase varied between 10% and 30% depending on stroke order, polarity, and type. We summarized our dissertation findings that related to differences and similarities between land and ocean lightning. We concluded that the increase in oceanic peak currents could potentially result from differences in the meteorology or the lightning initiation process in land and ocean thunderstorms, and that the increase is not limited to initial negative CG processes.

## 7.1 Future Work

In this section, we propose future work that could help advance our general understanding of lightning, in particular of oceanic lightning.

### 7.1.1 Coastal VHF Array

Recent years have seen the installation of many land VHF Lightning Mapping Array (LMA) systems, similar to the North Alabama LMA that is used in this dissertation. Although limited to line of sight coverage, LMA networks provide accurate (10s of meters) three-dimensional maps of leader breakdown and intra-cloud activity. We propose installing an LMA system in a coastal region to monitor and investigate

breakdown processes in ICs and CGs in land and ocean storms.

### 7.1.2 Inverse Modeling of LF and VHF Data

In Chapter 4, we developed an inverse model that estimated lightning currents using the LF waveform. The leader breakdown activity, which could be captured by VHF LMA systems, precedes the return stroke, depositing electric charge along the lightning channel. The charge is then neutralized by the return stroke, potentially affecting the current profile along the channel. We propose using both VHF and LF radiation data to improve the remote estimation of lightning return stroke currents. This would involve adding inputs to the inverse model that characterize VHF activity preceding and following the return stroke. The approach would however be limited to short distances ( $<100$  km) due to line of sight coverage of LMA.

### 7.1.3 Coastal Thunderstorms

In Chapter 6, we conducted statistical studies that showed a sharp transition in peak currents across stroke types, orders and polarities at the land-ocean boundaries. The analyses included all storms that occurred along the eastern coast of the United States. We propose conducting case studies of individual thunderstorm systems and monitoring how peak currents vary as the thunderstorm crosses the land-ocean boundary.

# Bibliography

- Akita, M., Y. Nakamura, S. Yoshida, T. Morimoto, T. Ushio, Z. Kawasaki, and D. Wang (2010), What occurs in K process of cloud flashes?, *J. Geophys. Res.*, *115*(D7), doi:10.1029/2009JD012016.
- Ballarotti, M. G., C. Medeiros, M. F. Saba, W. Schulz, and O. J. Pinto (2012), Frequency distributions of some parameters of negative downward lightning flashes based on accurate-stroke-count studies, *J. Geophys. Res.*, *117*, doi:10.1029/2011JD017135.
- Bardo, E. A., K. L. Cummins, and W. A. Brooks (2004), Lightning current parameters derived from lightning location systems, in *International Conference on Lightning Detection, Helsinki, Finland*.
- Barr, R. (1971), The propagation of ELF and VLF radio waves beneath an inhomogeneous anisotropic ionosphere, *Journal of Atmospheric and Terrestrial Physics*, *33*, 343–353.
- Barrington-Leigh, C. P., and U. S. Inan (1999), Elves triggered by positive and negative lightning discharges, *Geophys. Res. Lett.*, *26* (6), 683686.
- Berger, K., R. Anderson, and H. Kroninger (1975), Parameters of lightning flashes, *Electra*, *80*, 223–237.
- Broyden, C. G. (1970), The convergence of a class of double-rank minimization algorithms, *Journal of the Institute of Mathematics and Its Applications*, *6*, 7690, doi:10.1093/imamat/6.1.76.

- Budden, K. (1961), *The Wave-guide Mode Theory of Wave Propagation*, Logos Press.
- Budden, K. G. (1985), *The propagation of radio waves: the theory of radio waves of low power in the ionosphere and magnetosphere*, Cambridge University Press.
- Campos, L. S., M. F. Saba, T. A. Warner, O. Pinto Jr., E. P. Krider, and R. E. Orville (2014), High-speed video observations of natural cloud-to-ground lightning leaders - A statistical analysis, *Atmospheric Research*, 135-136, 285–305, doi:10.1016/j.atmosres.2012.12.011.
- Cauchy, A. (1847), Mthode gnrale pour la rsolution des systmes d'equations simultanes, p. 536538.
- Christian, H. J., R. J. Blakeslee, S. J. Goodman, D. A. Mach, M. F. Stewart, D. E. Buechler, W. J. Koshak, J. M. Hall, W. L. Boeck, K. T. Driscoll, and D. J. Boccippio (1999), The lightning imaging sensor, *NASA CONFERENCE PUBLICATION*, pp. 746–749.
- Christian, H. J., R. J. Blakeslee, D. J. Boccippio, W. L. Boeck, D. E. Buechler, K. T. Driscoll, S. J. Goodman, J. M. Hall, W. J. Koshak, D. M. Mach, and M. F. Stewart (2003), Global frequency and distribution of lightning as observed from space by the Optical Transient Detector, *J. Geophys. Res.*, 108(D1), 4005.
- Cohen, M. B., U. S. Inan, and E. P. Paschal (2010), Sensitive broadband ELF/VLF radio reception with the AWESOME instrument, *IEEE Trans. Geosc. Remote Sensing*, 48(1), 3–17, 10.1109/TGRS.2009.2028,334.
- Cohen, M. B., U. S. Inan, D. Piddychiy, N. G. Lehtinen, and M. Gokowski (2011), Magnetospheric injection of elf/vlf waves with modulated or steered hf heating of the lower ionosphere, *J. Geophys. Res.*, 116, doi:10.1029/2010JA016194.
- Cohen, M. B., N. G. Lehtinen, and U. S. Inan (2012), Models of ionospheric vlf absorption of powerful ground based transmitters, *Geophys. Res. Lett.*, 39, doi:10.1029/2012GL054437.

- Cooray, V. (1987), Effects of propagation on the return stroke radiation fields, *Radio Science*, 22:5, 757–768.
- Cooray, V., and V. A. Rakov (2011), Engineering lightning return stroke models incorporating current reflection from ground and finitely conducting ground effects, *IEEE TRANSACTIONS ON ELECTROMAGNETIC COMPATIBILITY*, 53, doi:10.1109/TEM.2011.2113350.
- Cooray, V., R. Jayaratne, and K. L. Cummins (2013), On the peak amplitude of lightning return stroke currents striking the sea, *Atmospheric Research*, 149, 372376, doi:doi:10.1016/j.atmosres.2013.07.012.
- Cummer, S. A., and U. S. Inan (1997), Measurement of charge transfer in sprite-producing lightning using ELF radio atmospherics, *Geophys. Res. Lett.*, 24(14), 1731–1734.
- Cummer, S. A., W. A. Lyons, and M. A. Stanley (2013), Three years of lightning impulse charge moment change measurements in the united states, *J. Geophys. Res.*, 118, 1–14, doi:10.1002/jgrd.50442.
- Cummins, K. L., and M. J. Murphy (2009), An overview of lightning locating systems: History, techniques, and data uses, with an in-depth look at the U.S. NLDN, *IEEE Trans. Electromag. Compatibility*, 51(3), 499–518.
- Cummins, K. L., M. J. Murphy, E. A. Bardo, W. L. Hiscox, R. B. Pyle, and A. E. Pifer (1998a), A combined TOA/MDF technology upgrade of the U.S. National Lightning Detection Network, *J. Geophys. Res.*, 103, 9035–9044.
- Cummins, K. L., E. P. Krider, and M. D. Malone (1998b), The U.S. national lightning detection network and applications of cloud-to-ground lightning data by electric power utilities, *IEEE Trans. on Electromag. Compat.*, 40, 465–480.
- Cummins, K. L., J. A. Cramer, W. A. Brooks, and E. P. Krider (2005), On the effect of land:sea and other earth surface discontinuities on lls-inferred lightning parameters, in *VIII International Symposium on Lightning Protection*.

- Davis, S. M. (1999), Properties of lightning discharges from multiple-station wideband electric field measurements, Ph.D. thesis, University of Florida, Gainesville.
- De Conti, A., F. H. Silveira, and S. Visacro (2014), A study on the influence of corona on currents and electromagnetic fields predicted by a nonlinear lightning return-stroke model, *J. Geophys. Res. Atmos.*, *119*, 51425156, doi:10.1002/2013JD020915.
- Dennis, A. S. (1970), The flashing behavior of thunderstorms, *J. Atmos. Sci.*, *27*, 171–172.
- Dennis, J. E., and R. B. Schnabel (1983), *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, Englewood Cliffs, NJ: Prentice-Hall.
- Efron, B. (1979), Bootstrap methods: Another look at jackknife, *Ann. Stat.*, *7*, 1–26.
- Fine, H. (1953), An effective ground conductivity map for continental United States, *Federal Communs. Comm., T. R. R. Rept.*
- Finke, U. (1998), Space-time correlations of lightning distributions, *Monthly Weather Review*, *127*, 1850–1861.
- Fleenor, S. A., C. J. Biagi, K. L. Cummins, E. P. Krider, and X. Shao (2009), Characteristics of cloud-to-ground lightning in warm-season thunderstorms in the Central Great Plains, *Atmospheric Research*.
- Fletcher, R. (1970), A new approach to variable metric algorithms, *Computer Journal*, *13*(3), 317322, doi:10.1093/comjnl/13.3.317.
- Fletcher, R., and C. M. Reeves (1964), Function minimization by conjugate gradients, *Computer Journal*, *7*, 149–154.
- Fukunishi, H., Y. Takahashi, M. Kubota, K. Sakanoi, U. S. Inan, and W. A. Lyons (1996), Elves: Lightning-induced transient luminous events in the lower ionosphere, *Geophys. Res. Lett.*, *23*(16), 21572160, doi:10.1029/96GL01979.

- Goldfarb, D. (1970), A family of variable metric updates derived by variational means, *Mathematics of Computation*, *24*(109), 2326, doi:10.1090/S0025-5718-1970-0258249-6.
- Goodman, S. J., T. R. Blakeslee, H. Christian, W. Koshak, J. Bailey, J. Hall, E. McCaul, D. Buechler, C. Darden, J. Burks, T. Bradshaw, and P. Gatlin (2005), The North Alabama Lightning Mapping array: recent severe storm observations and future prospects, *Atmospheric Research*, *76*, 423–437, doi:10.1016/j.atmosres.2004.11.035.
- Gopalakrishnan, V., S. D. Pawar, P. Murugavel, and K. P. Johare (2011), Electrical characteristics of thunderstorms in the Eastern part of India, *J. Atmos. and Solar-Terres. Phys.*, *73*, 1876–1882.
- Graf, K. L., N. G. Lehtinen, M. Spasojevic, M. B. Cohen, R. A. Marshall, and U. S. Inan (2013), Analysis of experimentally validated trans-ionospheric attenuation estimates of vlf signals, *J. Geophys. Res. Space Physics*, *118*, doi:doi:10.1002/jgra.50228.
- Hagan, M. T., H. B. Demuth, and M. H. Beale (1996), *Neural Network Design*.
- Hagenguth, J. H. (1951), The lightning discharge, *Compendium of Meteorology*, pp. 136–143.
- Hargreaves, J. K. (1992), *The Solar-Terrestrial Environment*, Cambridge University Press, New York.
- Harriman, S. K., E. W. Paschal, and U. S. Inan (2009), Magnetic sensor design for femto-tesla low-frequency signals, *IEEE Trans. Geosc. Remote Sensing*, p. Paper in press.
- Hastie, T., R. Tibshirani, and J. Friedman (2009), *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, second ed., Springer Science and Business Media.



- Hubert, P., and G. Mouget (1981), Return stroke velocity measurements in two triggered lightning flashes, *J. Geophys. Res.*, *86*, 5253–5261.
- Hutchins, M. L., R. H. Holzworth, K. S. Virts, J. M. Wallace, and S. Heckman (2013), Radiated vlf energy differences of land and oceanic lightning, *Geophys. Res. Lett.*, *40*, doi:doi:10.1002/grl.50406.
- Idone, V. P., and R. E. Orville (1982), Lightning return stroke velocities in the Thunderstorm Research International Program, *J. Geophys. Res.*, *87*, 4903–4915.
- Idone, V. P., and R. E. Orville (1984), Three unusual strokes in a triggered lightning flash, *J. Geophys. Res.*, *89*, 7311–7316.
- Inan, U. S., and A. S. Inan (2000), *Electromagnetic waves*, Prentice Hall.
- Ishii, M., K. Shimizu, J. Hojo, and K. Shinjo (1998), Termination of multiple-stroke flashes observed by electromagnetic field, *J. Geophys. Res., Proc. 24th Int. Conf. on Lightning Protection, Birmingham, United Kingdom*, 11–16.
- Jacobson, E. A., and E. P. Krider (1976), Electrostatic field changes produced by Florida lightning, *J. Atmos. Sci.*, *33*, 113–117.
- Kong, X., X. Qie, and Y. Zhao (2007), Characteristics of a positive cloud-to-ground lightning flash observed by high-speed video camera, *13th International Conference on Atmospheric Electricity, ICAE, Beijing, China*.
- Krehbiel, P. R. (1981), An analysis of the electric field change produced by lightning, Ph.D. thesis, Univ. of Manchester Inst. of Sci. and Technol., Manchester, England.
- Krehbiel, P. R., M. Brook, and R. A. McCrory (1979), An analysis of the charge structure of lightning discharges to the ground, *J. Geophys. Res.*, *84*, 2432–2456.
- Krider, E. P., G. J. Radda, and R. C. Noggle (1975), Regular radiation field pulses produced by intracloud lightning discharges, *J. Geophys. Res.*, *80*(27), 3801–3804, doi:10.1029/JC080i027p03801.

- Kuettner, J. (1950), The electrical and meteorological conditions inside thunderclouds, *Journal of Meteorology*, 7, 322–332.
- Lehtinen, N. G., and U. S. Inan (2008), Radiation of ELF/VLF waves by harmonically varying currents into a stratified ionosphere with application to radiation by a modulated electrojet, *J. Geophys. Res.*, 113(A06301), doi:10.1029/2007JA012,911.
- Lehtinen, N. G., and U. S. Inan (2009), Full-wave modeling of transionospheric propagation of VLF waves, *Geophys. Res. Lett.*, 36(L03104), doi:10.1029/2008GL036,535.
- Lehtinen, N. G., T. F. Bell, L. Qiu, M. B. Cohen, and U. S. Inan (2012), Full wave modeling of vlf wave scattering and propagation in curvilinear stratified ionosphere, in *2012 International Conference on Electromagnetics in Advanced Applications (ICEAA)*, pp. 195–8, doi:doi:10.1109/ICEAA.2012.6328619.
- Liang, C., B. Carlson, N. Lehtinen, M. Cohen, R. A. Marshall, and U. Inan (2014), Differing current and optical return stroke speeds in lightning, *Geophys. Res. Lett.*, 41, doi:10.1002/2014GL059703.
- Lyons, W. A., M. Uliasz, and T. E. Nelson (1998), Large Peak Current Cloud-to-Ground Lightning Flashes during the Summer Months in the Contiguous United States, *Monthly Weather Review*, 126, 2217, doi:10.1175/1520-0493(1998)126(2217:LPCCTG)2.0.CO;2.
- MacGorman, D. R., and W. D. Rust (1998), *The Electrical Nature of Storms*, Oxford University Press.
- Mach, D. M., and W. D. Rust (1989), Photoelectric return stroke velocity and peak current estimates in natural and triggered lightning, *J. Geophys. Res.*, 94, 13,237–13,247.
- Marshall, R. A. (2012), An improved model of the lightning electromagnetic field interaction with the d-region ionosphere, *J. Geophys. Res.*, 117, doi:10.1029/2011JA017408.

- Mazur, V., P. R. Krehbiel, and X. M. Shao (1995), Correlated high-speed video and radio interferometric observations of a cloud-to-ground lightning flash, *J. Geophys. Res.*, *100*, 25,731–25,753.
- Montanya, J., O. Van Der Velde, and E. R. Williams (2014), Lightning discharges produced by wind turbines, *J. Geophys. Res.*, *119*, doi:10.1002/2013JD020225.
- Nag, A., S. Mallick, V. A. Rakov, J. Howard, C. Biagi, D. Hill, M. A. Uman, D. M. Jordan, K. Rambo, J. Jerauld, B. A. DeCarlo, K. L. Cummins, and C. J. A. (2011), Evaluation of U.S. National Lightning Detection Network performance characteristics using rocket-triggered lightning data acquired in 2004-2009, *J. Geophys. Res.*, *116*(D02123), doi:10.1029/2010JD014,929.
- Nag, A., V. A. Rakov, and K. L. Cummins (2012), Analysis of NLDN-estimated peak currents for positive cloud-to-ground lightning, in *2012 International Conference on Lightning Protection (ICLP), Vienna, Austria*.
- Newsome, R. T. (2010), Ground-based photometric imaging of lightning EMP-induced transient luminous events, Ph.D. thesis, Stanford University.
- Norton, K. A. (1936), The propagation of radio waves over the surface of the Earth and in the upper atmosphere, *Radio Science*, *24*:10.
- Nucci, C. A., C. Mazzetti, F. Rachidi, and M. Ianoz (1988), On lightning return stroke models for the lemp calculations, *Proc. 19th Int. Conf. on Lightning Protection*, pp. 463–469.
- Orville, R. E., and G. R. Huffines (2001), Cloud-to-ground lightning in the United States: NLDN results in the first decade 1989-1998, *Monthly Weather Review*, *129*, 1179-1193.
- Orville, R. E., G. R. Huffines, W. R. Burrows, R. L. Holle, and K. L. Cummins (2001), North American Lightning Detection Network (NALDN) – First Results: 1998-2000, *AGU Fall Meeting Abstracts*, p. A59.

- Orville, R. E., G. R. Huffines, W. R. Burrows, R. L. Holle, and K. L. Cummins (2011), The North American Lightning Detection Network (NALDN) analysis of flash data: 200109, *Monthly Weather Review*, doi:10.1175/2010MWR3452.1.
- Prentice, S. A., and D. Mackerras (1977), The Ratio of Cloud to Cloud-Ground Lightning Flashes in Thunderstorms, *J. of Applied Meteorology*, 16, 545–550, doi:10.1175/1520-0450(1977)016<0545:TROCTC>2.0.CO;2.
- Proctor, D. E. (1981), VHF radio pictures of cloud flashes, *J. Geophys. Res.*, 86(C5), 4041–4071, doi:10.1029/JC086iC05p04041.
- Proctor, D. E. (1991), Regions where lightning flashes began, *J. Geophys. Res.*, 96(D3), 5099–5112, doi:10.1029/90JD02120.
- Proctor, D. E. (1997), Lightning flashes with high origins, *J. Geophys. Res.*, 102(D2), 1693–1706, doi:10.1029/96JD02635.
- Rakov, V. A., and A. A. Dulzon (1987), Calculated electromagnetic fields of lightning return stroke, *Tekh. Elektrodinam*, 1, 87–89.
- Rakov, V. A., and M. A. Uman (1990a), Waveforms of first and subsequent leaders in negative lightning flashes, *J. Geophys. Res.*, 95(D10), 16,561–16,577, doi:10.1029/JD095iD10p16561.
- Rakov, V. A., and M. A. Uman (1990b), Some properties of negative cloud-to-ground lightning flashes versus stroke order, *J. Geophys. Res.*, 95(D5), 5447–5453, doi:10.1029/JD095iD05p05447.
- Rakov, V. A., and M. A. Uman (2007), *Lightning: Physics and Effects*, Cambridge University Press.
- Rakov, V. A., R. Thottappillil, and M. A. Uman (1992), Electric field pulses in K and M changes of lightning ground flashes, *J. Geophys. Res.*, 97, 9935–9950.
- Rakov, V. A., M. A. Uman, and R. Thottappillil (1994), Review of lightning properties from electric field and TV observations, *J. Geophys. Res.*, 99(D5), 10,745–10,750, doi:10.1029/93JD01205.

- Rakov, V. A., M. A. Uman, G. R. Hoffman, M. W. Masters, and M. Brook (1996), Bursts of pulses in lightning electromagnetic radiation: Observations and implications for lightning test standards, *IEEE Trans. on Electromagn. Compat.*, *38*, 156–164, doi:10.1109/15.494618.
- Ramo, S., J. R. Whinnery, and T. Van Duzer (1994), *Fields and Waves in Communications Electronics*, John Wiley and Sons, New York.
- Rasmussen, J. E., P. A. Kossey, and E. A. Lewis (1980), Evidence of an ionospheric reflecting layer below the classical D region, *Journal of Geophysical Research*, *85*, 3037–3044.
- Ratcliffe, J. A. (1959), *The Magneto-ionic Theory and its Applications to the Ionosphere*, Cambridge University Press, Cambridge.
- Rhodes, C., and P. R. Krehbiel (1989), Interferometric observations of a single stroke cloud-to-ground flash, *Geophys. Res. Lett.*, *16*, 1169–1172.
- Saba, M. F., M. G. Ballarotti, and O. Pinto Jr. (2006), Negative cloud-to-ground lightning properties from high-speed video observations, *J. Geophys. Res.*, *111*, doi:10.1029/2005JD006415.
- Saba, M. F., W. Schulz, T. A. Warner, L. S. Campos, C. Schumann, E. P. Krider, K. L. Cummins, and R. E. Orville (2010), High-speed video observations of positive lightning flashes to ground, *J. Geophys. Res.*, *115*, doi:10.1029/2010JD014330.
- Said, R. K. (2009), Accurate and efficient long-range lightning geo-location using a VLF radio atmospheric waveform bank, Ph.D. thesis, Stanford University.
- Said, R. K., U. S. Inan, and K. L. Cummins (2010), Long-range lightning geo-location using a VLF radio atmospheric waveform bank, *J. Geophys. Res.*, *115*(doi:10.1029/2010JD013863), D23,108.
- Said, R. K., M. B. Cohen, and U. S. Inan (2013), Highly intense lightning over the oceans: Estimated peak currents from global gld360 observations, *J. Geophys. Res.*, *118*(doi:10.1002/jgrd.50508).

- Schonland, B. F. G., D. J. Malan, and H. Collens (1935), Progressive lightning ii, *Proc. Roy. Soc. A*, *152*, 595625.
- Shanno, D. F. (1970), Conditioning of quasi-newton methods for function minimization, *Math. Comput.*, *24(111)*, 647656, doi:10.1090/S0025-5718-1970-0274029-X, MR42:8905.
- Shao, X. M., and P. R. Krehbiel (1996), The spatial and temporal development of intracloud lightning, *J. Geophys. Res.*, *101(D21)*, 26,641–26,668, doi:10.1029/96JD01803.
- Stall, C. A., K. L. Cummins, E. P. Krider, and J. A. Cramer (2009), Detecting multiple ground contacts in cloud-to-ground lightning flashes, *J. Atmos. Oceanic Tech.*, *26*, 2392–2402.
- Stratton, J. A. (1941), *Electromagnetic Theory*, New York, McGraw Hill.
- Thomson, N. R. (1993), Experimental daytime VLF ionospheric parameters, *J. Atmos. Terr. Phys.*, *55*, 831–834.
- Thottappillil, R., V. A. Rakov, and M. A. Uman (1990), K and M changes in close lightning ground flashes in Florida, *J. Geophys. Res.*, *95*, 18,631–18,640.
- Thottappillil, R., V. A. Rakov, M. A. Uman, W. H. Beasley, M. J. Master, and D. V. Shelukhin (1992), Lightning subsequent-stroke electric field peak greater than the first stroke peak and multiple ground terminations, *J. Geophys. Res.*, *97(D7)*, 7503–7509, doi:10.1029/92JD00557.
- Turman, B. N. (1977), Detection of lightning superbolts, *J. Geophys. Res.*, *82(18)*, 25662568, doi:10.1029/JC082i018p02566.
- Turman, B. N., V. A. Rakov, J. D. Hill, T. Ngin, W. R. Gamerota, J. T. Pilkey, C. J. Biagi, D. M. Jordan, M. A. Uman, J. A. Cramer, and A. Nag (2014), Performance characteristics of the nldn for return strokes and pulses superimposed on steady currents, based on rocket-triggered lightning data acquired in florida in 20042012, *J. Geophys. Res. Atmos.*, *119*, 38253856, doi:10.1002/2013JD021401.

- Tyahla, L. J., and R. E. Lopez (1994), Effect of surface conductivity on the peak magnetic field radiated by first return strokes in cloud-to-ground lightning, *J. Geophys. Res.*, *99*, 10,517-10,525.
- Uman, M. A. (1987), *The Lightning Discharge*, Dover Publications.
- Ushio, T., S. Heckman, K. Driscoll, D. Boccippio, H. Christian, and Z. I. Kawasaki (2002), Cross-sensor comparison of the Lightning Imaging Sensor (LIS), *Int. J. Remote Sens.*, *23*, 2703-2712, doi:10.1080/01431160110107789.
- Valine, W. C., and E. P. Krider (2002), Statistics and characteristics of cloud-to-ground lightning with multiple ground contacts, *J. Geophys. Res.*, *107*(D20), doi:10.1029/2001JD001360.
- Watt, A. D. (1967), *VLF Radio Engineering*, Pergamon Press.
- White, K. D., G. T. Stano, and B. Carcione (2013), An investigation of North Alabama Lightning Mapping array data and usage in the real-time operational warning environment during the March 2, 2012 severe weather outbreak in northern Alabama, *Proceedings of the 93rd American Meteorological Society Annual Meeting*, Austin, Texas.
- Willet, J. C., V. P. Idone, R. E. Orville, C. Leteinturier, A. Eybert-Berard, L. Barret, and E. P. Krider (1988), An experimental test of the "transmission-line model" of electromagnetic radiation from triggered lightning return strokes, *J. Geophys. Res.*, *93*, 3867-3878.
- Wilson, C. T. R. (1916), On some determinations of the sign and magnitude of electric discharges in lightning flashes, *Proc. Roy. Soc.*, *92*, 555-574.
- Winn, W. P., and L. G. Byerley (1975), Electric field growth in thunderclouds, *Q. J. R. Met. Soc.*, *101*, 979-994.
- Wood, T. G. (2004), Geo-location of individual lightning discharges using impulsive VLF electromagnetic waveforms, Ph.D. thesis, Stanford University.

- Wormell, T. W. (1939), The effect of thunderstorms and lightning discharges on the Earth's electric field, *Phil. Trans. Roy. Soc.*, *328*, 249–303.
- Zoghzoghy, F. G., M. B. Cohen, R. K. Said, and U. S. Inan (2013), Statistical patterns in the location of natural lightning, *J. Geophys. Res. Atmos.*, *118*, 787–796, doi:10.1002/jgrd.50107.
- Zoghzoghy, F. G., M. B. Cohen, R. K. Said, S. S. Basilico, R. J. Blakeslee, and U. S. Inan (2014), Lightning activity following the return stroke, *J. Geophys. Res. Atmos.*, *119*, doi:10.1002/2014JD021738.